54th INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

(ICEST 2019)

Ohrid, North Macedonia, June 27-29, 2019

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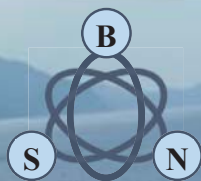Faculty of Technical Sciences,
Bitola, Macedonia

University of Niš,
Faculty of Electronic Engineering,
Serbia

Technical University of Sofia,
Faculty of Telecommunications,
Bulgaria

Proceedings of Papers

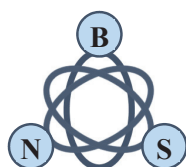**54ᵗʰ INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES**

iCEST 2019

Proceedings of Papers

ICEST 2019 – 54th INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES, Ohrid, North Macedonia, June 27 - 29, 2019

Proceedings of Papers

The 54th International Scientific Conference on Information, Communication and Energy Systems and Technology ICEST 2019 was held from 27 to 29 June in Ohrid, North Macedonia. The Conference is for eighteenth time jointly organized by the Faculty of Technical Sciences, Bitola, N. Macedonia, the Faculty of Electronic Engineering, Niš, Serbia and by the Faculty of Telecommunications, Sofia, Bulgaria. As to the earlier ICEST conferences, authors from institutions all over the world submitted their papers. This year 252 authors from 14 countries presented their 103 scientific and application oriented papers.

A plenary lecture was given by Pavle Boškoski, Department of Systems and Control, Jožef Stefan Institute, Ljubljana, Slovenia. 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.

Prof. Dr Mitko Kostov,

ICEST 2019 Conference Chairman

ICEST 2019 conference is supported by

LIST OF ICEST 2019 REVIEWERS

Nikola Acevski
St. Kliment Ohridski University, Bitola, N.Macedonia

Blagoja Arapinoski
St. Kliment Ohridski University, Bitola, N.Macedonia

Aleksandar Atanasković
University of Niš, Serbia

Ivaylo Atanasov
Technical University of Sofia, Bulgaria

Metodija Atanasovski
St. Kliment Ohridski University, Bitola, N.Macedonia

Vesna Ceselkoska
St. Kliment Ohridski University, Bitola, N.Macedonia

Dragan Denić
University of Niš, Serbia

Tijana Dimitrijević
University of Niš, Serbia

Ivo Draganov
Technical University of Sofia, Bulgaria

Goran Đorđević
University of Niš, Serbia

Sandra Đošić
University of Niš, Serbia

Ivo Dochev
Technical University of Sofia, Bulgaria

Nebojša Dončov
University of Niš, Serbia

Gordana Janevska
St. Kliment Ohridski University, Bitola, N.Macedonia

Jugoslav Joković
University of Niš, Serbia

Goran Jovanović
University of Niš, Serbia

Zoran Jovanović
University of Niš, Serbia

Kiril Kasev
Technical University of Sofia, Bulgaria

Mitko Kostov
St. Kliment Ohridski University, Bitola, N.Macedonia

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

Zlatica Marinković
University of Niš, Serbia

Aleksandar Markoski
St. Kliment Ohridski University, Bitola, N.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

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

Natasa Mojsoska
St. Kliment Ohridski University, Bitola, N.Macedonia

Marin Nedelchev
Technical University of Sofia, Bulgaria

Goran Nikolić
University of Niš, Serbia

Tatjana Nikolić
University of Niš, Serbia

Kamelia Nikolova
Technical University of Sofia, Bulgaria

Oleg Panagiev
Technical University of Sofia, Bulgaria

Aleksandra Panajotović
University of Niš, Serbia

Bratislav Predić
University of Niš, Serbia

Olivera Pronić-Rančić
University of Niš, Serbia

Nikola Rendevski
St. Kliment Ohridski University, Bitola, N.Macedonia

Blagoja Ristevski
St. Kliment Ohridski University, Bitola, N.Macedonia

Mare Srbinovska
Ss. Cyril and Methodius University, N.Macedonia

Zoran Stanković
University of Niš, Serbia

Blagoja Stevanovski
St. Kliment Ohridski University, Bitola, N.Macedonia

Dragan Stojanović
University of Niš, Serbia

Biljana Stošić
University of Niš, Serbia

Ljupco Trpezanovski
St. Kliment Ohridski University, Bitola, N.Macedonia

Zlatka Valkova-Jarvis
Technical University of Sofia, Bulgaria

Boban Veselić
University of Niš, Serbia

TECHNICAL PROGRAM COMMITTEE

Chairman:

Mitko Kostov   St. Kliment Ohridski University, Bitola, N.Macedonia

Vice Chairmen:

Nebojša Dončov   University of Niš, Serbia
Kalin Dimitrov   Technical University of Sofia, Bulgaria

Honorary Chairmen
Cvetko Mitrovski   *St. Kliment Ohridski University, Bitola, N.Macedonia*
Bratislav Milovanović   *Singidunum University, Niš, Serbia*
Rumen Arnaudov   *Technical University of Sofia, Bulgaria*

Local Coordinator:

Stojanče Nusev   Dean of the Faculty of Technical Sciences,
                                                    St. Kliment Ohridski University, Bitola, N.Macedonia

Members:

Agata Manolova   Technical University of Sofia, Bulgaria
Biljana Stošić   University of Niš, Serbia
Boyanka Nikolova   Technical University of Sofia, Bulgaria
Brian Barsky   University of California, Berkeley, USA
Chrysostomos Stylios   Technological Education Institute of Epiros, Greece
Đani Juričić   Jožef Stefan Institute, Slovenia
Désiré Dauphin Rasolomampionona   Warsaw University of Technology, Poland
Dimitar Taskovski   Ss. Cyril and Methodius University Skopje, N. Macedonia
Dimiter Alexandrov   Lakehead University, Canada
Dragan Denić   University of Niš, Serbia
Dragan Stojanović   University of Niš, Serbia
Dušanka Bošković   University of Sarajevo, Bosnia and Herzegovina
Eduard Ivanjko   University of Zagreb, Croatia
Evelin Krmac   University of Ljubljana, Slovenia
Evelina Pencheva   Technical University of Sofia, Bulgaria
Fernando Alvarez   University of Extremadura, Spain
Filiz Güneş   Yildiz Technical University, Turkey
Ismail Altas   Karadeniz Technical University, Turkey
Ivan Corbev   Ss. Cyril and Methodius University Skopje, N.Macedonia
J.A.R.P. de Carvalho   University Beira Interior, Portugal
Jair Minoro Abe   Paulista University, Brazil
Jan Machac   Czech Technical University, Prague, Czech Republic
Jugoslav Joković   University of Niš, Serbia
Karel Raz   University of West Bohemia, Czech Republic
Kazumi Nakamatsu   University of Hyogo, Japan
Ljupčo Trpezanovski   St. Kliment Ohridski University Bitola, N.Macedonia
Mariofanna Milanova   University of Arkansas at Little Rock, USA
Mikolaj Bartłomiejczyk   Gdańsk University of Technology, Poland
Mikolaj Leszczuk   AGH University of Science and Technology, Poland
Nikola Donato   University of Messina, Italy
Pece Mitrevski   St. Kliment Ohridski University Bitola, N.Macedonia
Petar Popovski   Aalborg University, Denmark
Peter Planinšič   University of Maribor, Slovenia
Slavko Rupčić   University of Osijek, Croatia

Stanimir Valtchev   New University of Lisbon, Portugal
Stanimir Valtchev   New University of Lisbon, Portugal
Stefan Stancescu   Polytechnic University of Bucharest, Romania
Stevo Bozinovski   South Carolina State University, USA
Szilvia Nagy   University of Gyor, Hungary
Urban Burnik   University of Ljubljana, Slovenia
Uwe Siart   Technical University of Munich, Germany
Vincenzo Piuri   University of Milan, Italy
Wojciech Krzysztofik   Wroclaw University of Technology, Poland
Wojtek Bock   University of Ottawa, Canada
Zdenka Babić   University of Banja Luka, Bosnia and Herzegovina
Zoran Jovanović   University of Niš, Serbia
Zoran Stanković   University of Niš, Serbia

CONFERENCE ORGANIZING COMMITTEE

St. Kliment Ohridski University, Bitola, North Macedonia:

M. Atanasovski, B. Arapinoski, G. Janevska, V. Dancevska, N. Mojsovska, J. Pargovski, M. Spirovski, I. Andreevski, D. Koltovska

University of Niš, Serbia:

A. Atanasković, T. Dimitrijević, J. Joković, D. Mančić, N. Maleš-Ilić, Z. Marinković, V. Marković, M. Milijić, O. Pronić-Rančić, B. Stošić, B. Veselić, L. Stoimenov, J. Milošević

Technical University of Sofia, Bulgaria:

L. Laskov, A. Mihaylova, S. Antonov, V. Stoynov, D. Mihaylova, K. Valkov, N. Hristova, D. Kireva-Mihova, T. Brusev, M. Todorov, K. Raynova, K. Stoyanova, M. Stoyanova, M. Nedyalkova, V. Dolapchieva

CONFERENCE SECRETARIAT

Conference Chairman:
Prof. Dr. Mitko Kostov
St. Kliment Ohridski University, Bitola, N. Macedonia
Makedonska Falanga 37, 7000 Bitola, N. Macedonia
phone: +389 47 207 702
e-mail: mitko.kostov@uklo.edu.mk

Technical Editor:
M. Atanasovski, St. Kliment Ohridski University, Bitola, N. Macedonia
e-mail: metodija.atanasovski@uklo.edu.mk

Secretary:
B. Arapinoski, St. Kliment Ohridski University, Bitola, N. Macedonia
e-mail: blagoja.arapinoski@uklo.edu.mk

TABLE OF CONTENTS

SESSIONS PER TOPICS

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS & TELECOMMUNICATION SYSTEMS AND TECHNOLOGY

I

SIGNAL PROCESSING & DIGITAL IMAGE PROCESSING

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES
&
INFORMATICS AND COMPUTER SCIENCE

ELECTRONICS

ENERGY SYSTEMS AND EFFICIENCY

CONTROL SYSTEMS

MEASUREMENT SCIENCE AND TECHNOLOGY

ENGINEERING EDUCATION & OTHER

PLENARY SESSION

Prognostics and Health Management of Solid Oxide Fuel Cell Systems: Challenges, Accomplishments and Trends

Pavle Boškoski, Boštjan Dolenc
Jožef Stefan Institute, Department of Systems and Control
Ljubljana, Slovenia
*pavle.boskoski@ijs.si

Abstract—There are several competing technologies with potential of replacing our dependence on fossil fuels. Fuel cells or hydrogen based technologies are one of them. Although the this technology is sufficiently mature there are still challenges that hinder its broad applicability. Our focus is on a particular type of fuel cells, so-called solid-oxide fuel cell (SOFC). Currently, this type of fuel cells provide the best efficiency (fuel to electricity conversion). Due to high temperatures of operation, the durability of SOFCs is affected by thermal stress and material degradation. We will focus on the current status of this technology, its main advantages as well as the latest achievements in the field of prognostics and health management (PHM).

Index Terms—fuel cells, hydrogen, health management

I. INTRODUCTION

Energy security is one of the main pillars of our modern society. The projections for the EU energy consumptions foresee a slight decrease in our energy demand, however oil will retain its top position in the energy share of the continent. This is predominantly due to the requirements of the transportation sector. Since European oil reserves (excluding Russia) are virtually nonexistent it is of great importance to reduce our dependance on oil due to two main factors. First there is the obvious ecological viewpoint that supports Europe's strong dedication towards significant reduction in green house gas emission. Second is from a purely financial standpoint, since oil imports amount to 15% of the overall imports by the EU. By analysing the EU energy consumption by sectors the three most energy demanding segments are transportation, residential consumers and district heating systems. Addressing the demands of these segments can significantly decrease our dependence on imported energy.

Currently we are witnessing a substantial growth in the installed renewable energy sources predominantly wind farms and solar power plants. The biggest deficiency of these systems is the intermittent nature of power generation. Therefore in order to exploit these power sources in full while in the same time to satisfy our demand of energy it is required to provide means of efficient energy storage and energy diistribution systems.

The distribution gird for electrical energy is very efficient. It estimated losses are in the range of 5% to 8% [1]. As a result a lot of effort is put into the development of energy storage devices that can harness the electrical energy. When analysing the available technologies the usual metrics involve two properties: energy density (Wh/kg) and power density (W/kg). On one side of the scale are batteries and fuel cells, which have high energy density but low power density. On the other side are super capacitors with somewhat lower energy density but quite high power density. Our focus is on how energy density devices i.e. batteries and fuel cells.

Currently the leading technology in batteries development is based on Lithium. From a geo-political viewpoint, Europe is in the same situation as it is with oil. The world's largest deposits of this element are in South America and China. Therefore in long term building our society solely on Lithium based batteries will make us highly dependent on resources that are outside of our borders.

Unlike batteries that are energy storage devices, fuel cells are energy conversion devices. They convert chemical energy stored in the hydrogen fuel into electrical energy through electrochemical combination of hydrogen and oxygen. Based on the hydrogen source, principle of operation and operating temperature, fuel cell technology distinguishes among SOFCs [2], low-temperature proton exchange membrane fuel cells (PEMFCs) [3], high temperature polymer electrolyte membranes (HT-PEMs) [4], direct methanol fuel cell (DMFC), sulphuric acid fuel cell (SAFC), molten carbonate fuel cell (MCFC), solid polymer fuel cell (SPFC), and alkaline fuel cell (AFC) [5]. Regarding fuel cells there are roughly three technology groups: low temperature PEM cells, high temperature MCFC and the so-called SOFC. Unlike batteries, that are capable of storing of electrical energy, fuel cells use hydrogen as a fuel. Consequently, when analysing the viability of fuel cell technology (or commonly referred to as hydrogen technology) vs. batteries there are several points that on first glance make the fuel cells as wrong choice. Key among them being production, compression and distribution of hydrogen. However making a more detailed analysis provide a completely different story.

The SOFC technology offers very high conversion efficiency that is close to 60%. Therefore using renewable energy sources in the time intervals when there is an excess of energy it is possible to produce a so called green hydrogen with very little losses.

The second issue, compression, is usually addressed from

the viewpoint of mechanical compression devices that are highly inefficient in particular for low density hydrogen. However, there are currently solutions that are based on the concepts of electrochemical compression capable of reaching 1000 bars of pressure without any moving components. Consequently, the compression process can be performed with efficiency of almost 80%.

The final issue is the transportation. Compared to other parts of the world, the EU has vast and extremely well developed gas pipeline system that is currently used for delivery of natural gas to almost every household on the continent. There are already examples where hydrogen is used as a supplement to the existing natural gas. The ultimate goal is to use segments of the pipeline as a delivery system for hydrogen to the end users.

This analysis shows that hydrogen based technologies are viable solution to the issues regarding the energy security. Even more, the SOFC technology can be built without any rare earth minerals thus completely rendering our dependence on foreign materials. In the remaining of the paper we will present the current state of development of the SOFC based fuel cells, open issues regarding their exploitation and future trends.

II. SOFC IN A NUTSHELL

Fuel cels consist of three adjacent layers: (i.) anode, (ii.) cathode, and (iii.) electrolyte. Fig. 1 displays the basic principles and essential components of a single SOFC. The main purpose of the electrolyte is to transport oxygen ions O^{2-} from the cathode to the anode, while at the same time preventing direct contact between anode and cathode chambers. The ions are formed via oxygen reduction reaction at the cathode, which is continuously fed with oxygen. On the other side, at the anode, the O^{2-} ions react with hydrogen in the process of hydrogen oxidation. In addition to H_2O, two electrons and some heat are released. The electrons travel through the external load to reach the cathode, where they participate in oxygen reduction reaction. In addition, one of the major advantages of the SOFC is its ability of internal reforming of the fuel. This property allows the fuel cells to operate not only on pure hydrogen, but on also other hydrogen-rich fuels, such as methane.

Theoretically, the voltage of a single SOFC can reach up to 1.2 V, depending on temperature, pressure and gas composition. The power output, however, is heavily conditioned with the active area of the fuel cell. Typically, the larger the active area of the cells, the higher the power output. However, the limiting factor for the size of the individual fuel cells is related to the thermo-mechanical stress induced due to thermal expansion coefficient (TEC) mismatch of the adjacent layers. With larger cells, temperature gradients over the cells become larger, thus causing higher stress on the materials, hence compromising their safe operation. Therefore, to increase power output, a number of fuel cells is connected in parallel to form a fuel cell stack.
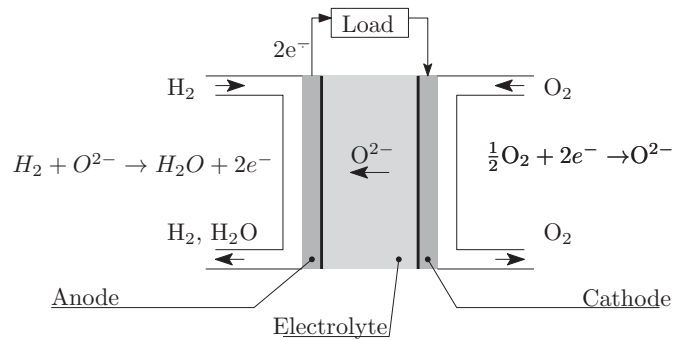
Fig. 1: The basic principles of a SOFC and its essential components.

Since the SOFCs operate at high temperatures, the stack is usually enclosed in an insulated housing to reduce heat losses. The stack then connects with a balance of plant (BOP) module, which carries out the pre-treatment of the incoming gases. Depending on the scale of the installation, the BOP consists of different interconnected components, such as heat-exchangers, blowers, fuel reformers, pipes, valves, etc.

III. PHM IN FUEL CELLS

The degradation occurring within the fuel cells inherently impedes the efficiency of power conversion. Apart from the fact a great deal of effort has been dedicated to the understanding of SOFC degradation mechanisms [6], only a relatively limited set of diagnostic approaches is available.

The PHM can be defined as a set of activities in which the main perspective is to enhance the effective reliability and availability of a product in its life-cycle conditions by detection of current and approaching failures [7].
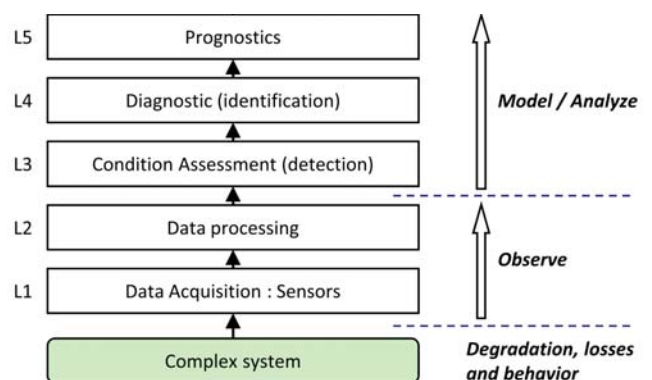
Fig. 2: PHM structure from [8]

Figure 2 shows basic self-contained building blocks of an PHM system, divided in seven layers. On the lower levels L1 the data is collected employing sensors and transducers. In the data processing level, features that carry information about the condition of an underlying system are extracted from the collected data. Further up, in the condition assessment layer, the features are evaluated in a sense that any deviation from

the normal operation is detected. If changes in the behaviour are detected, then the fault alarms are triggered. In the next stage, in the diagnostic layer, the patter of the triggered alarms are compared with known failure modes to possibly identify the cause of the triggered alarms. Having identified the faulty component, the prognostics layer aims to predict how the component in question will operate in future based on the historic data from the previous layers.

Within this framework, several passive fault detection and identification (FDI) approaches have been proposed to constantly monitor components that supply and pre-treat fuel before entering the reaction chambers of the fuel cells. That encompasses several analytical model-based approaches [9], [10], black-box approaches [11] and signal processing approaches [12]. However, yet the most predominate approaches to health assessment build on the use of electrochemical impedance spectroscopy (EIS). Characteristic for EIS is that applies local probing directly on the fuel cells in order to excite all the relevant dynamic modes related to chemical processes in the system to gain more detailed information about condition of the cells themselves. Although it has been around for several decades, the way it is used has not changed much.

Conventional EIS techniques use low-amplitude sinusoidal excitation, repeatedly performed at different frequencies, from which then the gain and phase of the points on the Nyquist curve are estimated. Such an approach suffers from too long probing time that is usually required to obtain high-quality EIS spectra. That means too long perturbation of the process in operating mode. Particularly critical is estimation of EIS curve at low frequencies as normally several periods of a sinusoid are required to extract precise information.

By applying excitation over a wide range of frequencies *simultaneously*, the authors in [13] showed that the same quality of the results as in conventional EIS can be obtained at the substantially shorter probing times (an order of magnitude). The evaluation of EIS curve is done by post-processing of the current and voltage signals by means of complex wavelet transform. Apart of the much shorter probing session, additional benefit is also much better resolution of the EIS curve obtained compared to the conventional EIS, as it is defined by the sampling rate. Savings in required probing times can gradually diminish when the required precision of the spectral reconstruction at low frequencies is increasing.

The measured EIS data can then be further processed to extract relevant information about the health status. Here, several approaches serve as tools the EIS data have to be interpreted either through the change of the pattern of the EIS curve, or by interpreting changes in the parameters of the equivalent circuit models (ECM) [14], [15] and distribution of relaxation timess (DRTs) [16].

A. Data acquisition

A block scheme of the data acquisition system consisting of several interconnected components is shown in Fig. 3. It includes programmable digital load, data acquisition device, a current probe, and controlling unit that connects the functionality of all the components.
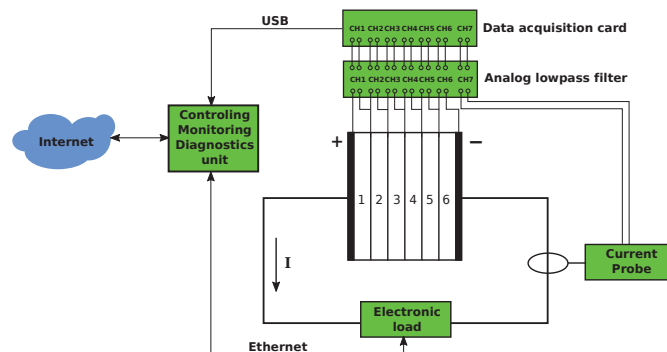
Fig. 3: Block scheme of the system for impedance characterisation of electrochemical energy devices

An electronic load is connected directly in series with the SOFC and is used to set DC current as well as to superimpose various excitation signals. This way, the current in the system is controlled by the load, and defined by the system operator. In more industrial circumstances, the electronic load can be also replaced by incorporating its functionality in power conditioning unit which is then connected to the electric grid. However, by doing so, one needs to keep in mind the limitations of connected power electronics when designing the excitation signals. The electrical current is measured using a non-invasive current probe which offers sufficiently high bandwidth and does not disturb the functionality of the system. The data acquisition hardware allows for individual cell voltages and the output of the current probe to be measured simultaneously. This way individual cells can be monitored at the same time and allow for more accurate isolation of the possibly faulty cell. Figure 4 show an excerpt from the data collected during one probing.
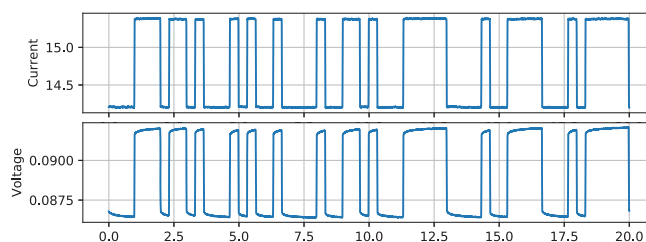
Fig. 4: An example of the data collected on a single SOFC.

The electronic load and the data acquisition hardware are connected to a computer that orchestrates the probing. First the electronic load is triggered to superimpose the excitation signals to the system while the the response from the cells and current probe are logged. Furthermore, the computer allows to design a schedule of experimental probing to facilitate automated and regular experiments. If needed the computer can also be used to process the data on sight, or send it to the cloud.

B. Data processing and feature extraction

The conventional frequency domain signal analysis performed with the Fourier transform, provides a detailed picture of the frequency components present in the signal but without any information regarding their time occurrence and duration. Time-frequency analysis offers a solution to this problem thus providing the information about the temporal details as well. Typical examples are the Short-time Fourier transform, Wigner-Ville distribution, wavelet transform etc.

Regardless of the selected method there is a theoretical limitation on the joined time-frequency resolution. Unlike other methods, the wavelet transform enables flexible selection of the desired time-frequency resolution by introducing the concepts of scaling. Wavelet transform is based on a set of specifically designed functions called wavelets. The continuous wavelet transform (CWT) of a square integrable function $f(t)$ is defined [17]

$$Wf(s,u) = \int_{-\infty}^{\infty} f(t)\Psi^*_{u,s}(t)dt \qquad (1)$$

where wavelet function $\Psi^*_{u,s}(t)$ is scaled by s and translated by u version of original mother wavelet $\Psi(t)$:

$$\Psi^*_{u,s}(t) = \frac{1}{s}\Psi(\frac{t-u}{s}) \qquad (2)$$

The key parameter in CWT is the selection of the wavelet function. EIS analysis requires information about the amplitude and phase of the excitation and response signals, therover a complex Morlet wavelet function is readily available [18]: The time and frequency localisation is determined through the parameters u and s respectively. More details regarding the properties of the Morlet wavelet and the application of CWT for EIS analysis can be found in [13].
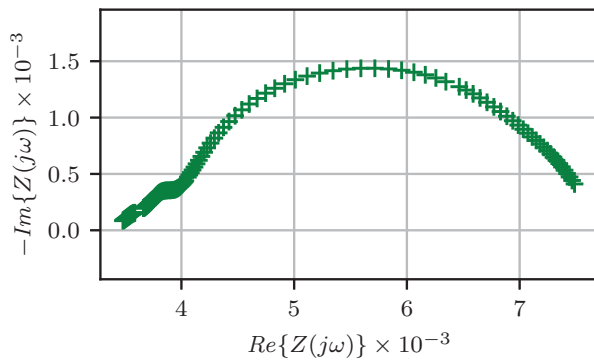
Fig. 5: An example of a Nyquist curve of a single SOFC.

The voltage/current signal pairs from Fig. 4 are transformed employing (1), devided and averaged over time to obtain Nyquist curve show in Fig. 5.

1) Feature extraction: Observing changes in Nyquist curves can be directly employed for detecting a change in a system and condition monitoring. However, a more concise information is required to facilitate the identification of

degradation mechanism. This can be achieved by modeling the measured Nyquist curve in one way or another and is often referred to as deconvolution of the EIS spectra.

Generally speaking, there are two approaches to deconvolution of the EIS spectra: (i.) ECM modeling through non-linear optimisation (ii.) non-parametric identification of the Nyquist curve through DRT [19]. From mathematical point of view the above approaches are equivalent. That is, having one of the two, one can easily derive the other one [20].

An ECM of fuel cell impedance can be described as a series of RQ elements connected in series, as shown in Fig. 6. It has
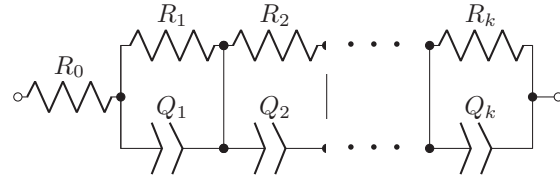
Fig. 6: General form of the ECM using RQ-elements

been shown that the degradation phenomena manifest in the changes of the parameters of the ECM. Indirect observations through DRT have recently been published [21]. Therefore, the *parameters* of the ECM can be employed to identify ongoing degradation mechanism.

There are several ways to estimate the parameters of any model. For instance, having a fuel cell impedance model structure

$$Z(j\omega) = R_s + \sum_i \frac{R_i}{\tau_i(j\omega)^{\alpha_i}+1} \qquad (3)$$

one can formulate an optimisation problem

$$\underset{\boldsymbol{\theta}}{\mathrm{argmin}} \quad c(\boldsymbol{\theta})$$

where $c(\boldsymbol{\theta})$ denotes a loss function which describes *goodness of fit* of the model with respect to measured Nyquist curve, and the $\boldsymbol{\theta}$ is a vector of model parameters. Following this approach one obtains a *point* estimate of parameter values.

In order to obtain richer insight regarding the accuracy of the estimates statistical based approaches can be employed. The main is as follows. Given a model and data D, the *posterior* distribution of the model parameters $\boldsymbol{\theta}$ can be estimated via Bayes rule as:

$$p(\boldsymbol{\theta}|D) = \frac{p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \qquad (4)$$

where $p(D|\boldsymbol{\theta})$ is *likelihood*, $p(\boldsymbol{\theta})$ is *prior*, and $\int p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is called marginal likelihood or model *evidence*.

By constructing likelihood function and defining prior probabilities (typically uninformative ones), the posterior of the model parameters can be inferred employing four different approached: (i.) analytically in the case of the tractable required mathematical operations, (ii.) employing numerical integration techniques (iii.) through Markov chain Monte Carlo (MCMC) simulations, (iv.) employing variational Bayes approximation.
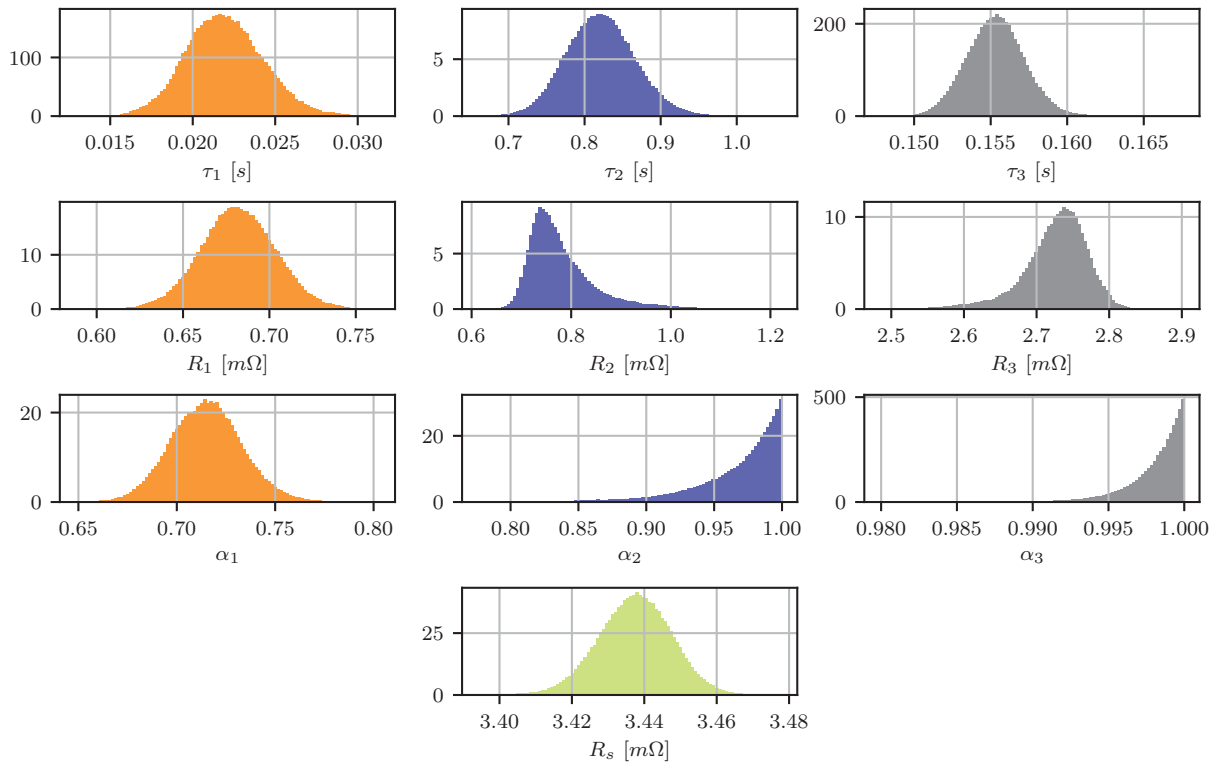
Fig. 7: Posterior probability density functions (PDFs) of the model parameters.

More on MCMC methods, and Bayesian inference, can be found in [22].

In this case a sequence of three RQ elements, $i \in \{1, 2, 3\}$, the posterior distributions of the model parameters are shown in Fig. 7. In particular, for each of the three RQ elements $i \in \{1 \ldots 3\}$, top row shows time constants τ_i, second row shows resistances R_i, and third row shows their corresponding α_i exponents. The bottom row shows serial resistance. Note that the parameters that belong to individual RQ element are denoted with the same colour.

C. Fault diagnosis

The onset of faults typically results into a change of the extracted feature values. The problem of reliable change detection in selected feature is outlined in Fig. 8. The right-hand figure shows the desired scenario i.e. the feature values exhibit steady growth. Once the value surpasses the pre-defined threshold an alarm is triggered. However, in the case of incipient faults, the extracted feature values become close to a pre-determined threshold value. Due to small variations one can observe numerous short periods when the feature value is higher than the threshold. This may lead to excessive rate of false alarms, and thus reducing the confidence in the the PHM system as a whole. This is a known problem that is usually solved either by employing some sort of filtering approaches or increasing the threshold values. Filtering approaches reduce the response time of the detection system, whereas increasing the threshold values reduces the its sensitivity. These problems can be resolved by employing robust statistical tests.
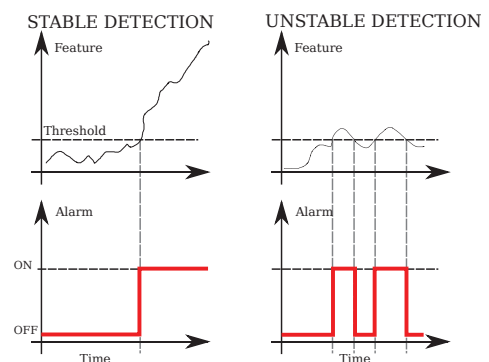
Fig. 8: Problem with detection detection stability

A change is established by quantifying dissimilarity among the features' statistics in current and nominal fault-free state.

By calculating Jensen-Rényi (JR) divergence among empirical distributions of the features, the need for known operating conditions and data records at various system failures is avoided. The approach was initially developed for vibration-based diagnostics for gears and bearings under constant and variable operating conditions, here is adopted for change detection in the parameters of an ECM, such as the one presented in previous section.

To this end, the change detection algorithm builds on generalised JR divergence to quantify the dissimilarity among two or more PDFs. The JR divergence JR_α^w quantifies dissimilarity among n PDFs:

$$JR_\alpha^w(\mathcal{P}_1, \ldots, \mathcal{P}_n) = H_\alpha\left(\sum_{i=1}^n w_i \mathcal{P}_i\right) - \sum_{i=1}^n w_i H_\alpha\left(\mathcal{P}_i\right),$$
(5)

In (5) H_α is Rényi entropy and $\sum_{i=1}^n w_i = 1$. The selection of weights w_i in (5) is in principle arbitrary. With w_i selected uniformly i.e. $w_i = 1/n$, the divergence reaches maximal value.

JR divergence quantifies shared information among n random variables. If they are identical, i.e. $\mathcal{P}_1 = \mathcal{P}_2 = \ldots = \mathcal{P}_n$, divergence is zero. However, if one of them deviates even slightly, the JR divergence becomes different from zero. Therefore, JR divergence carries information about dissimilarity among n PDFs.

Since we are seeking a way to detect a change in the statistics of the parameters of an ECM which were estimated in the previous section, JR divergence offers an elegant solution. To detect a change in the parameter, all that is needed is a reference set where the fuel cellwas operating under normal operation. Once the reference data is available, we can compare the online estimates of the model parameters with the reference ones. This is demostrated in Fig. 9. To be more precise, Fig. 9a shows how one of the parameters in from Fig. 7 behaved during a 600h long durability experiment of an SOFC short stack comprising of 6 cells. During this test, the data was collected every 6 hours. The figure was obtained by processing the data as described above in the paper. Note that there are three events that were intentionally triggered, which are marked by yellow strips. It is apparent that the selected condition monitoring parameter is affected by these changes. After the first event, the cell recovered, however, after the second experiment the cell remained permanently damaged. The change detection facilitated by employing JR divergence is shown in 9b. In this case, 10 PDFs were included in a reference data set, and compared with a set of 10 in an "online" mode. The JR divergence detects a first change in the parameter clearly, thus triggering the alarm. After this first event the cell stabilised and returned to normal operation. Further in the experiment, the the second event also caused severe change in the parameter. The algorithm clearly detected this change. However, after this second event the cell never recovered to its original state, hence the value of JR divergence remained at high level.

(a) Time evolution of the parameter

(b) JR divergence based change detection
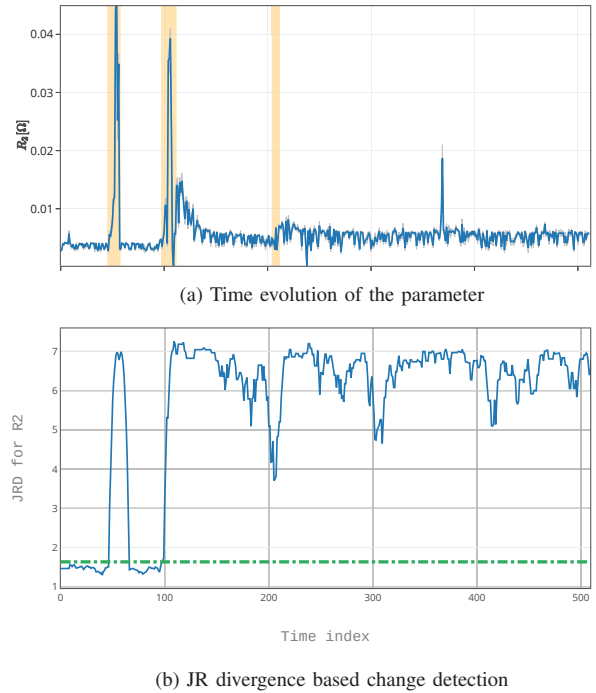
Fig. 9: Results of the proposed JRD based fault detection approach employed on a 3000 hours SOFC run-to-failure experiment.

D. Prognostics

Various approaches to condition monitoring have been proposed with the aim to detect the onset of fault and hence allow for the design of accommodation actions [6], [23]. However, for safety reasons and maintenance purposes, the information of particular practical value is how long will the system operate prior to its end of life (EOL). Relatively little attention has been devoted to the development of algorithms for predicting SOFC remaining useful life (RUL).

So far, some more work has been done in the area of proton exchange membrane (PEM) fuel cells. Although PEM and SOFC differ in many aspects, it is worth screening the main ideas applied to PEM. For example in [24], the authors discussed and summarised challenges related to PEM fuel cell prognostics and RUL predictions. The authors identify the importance of suitable health indicators for RUL prediction and address the definition of EOL point of fuel cells.

Works related to RUL prognosis in PEM aim at modelling temporal evolution of stack voltage, voltage related power output, or efficiency due to degradation [24]. The most common approach is to consider voltage as state of health (SOH) indicator and perform RUL upon evident drift in stack/cell voltage. Such a characterisation of SOH is often justified because voltage is directly associated with power output, and hence the efficiency of power conversion. Relatively few works define SOH differently, see e.g. [25]. Various empirical models are then employed in order to predict voltage drop in future

and estimate RUL. The employed models range from simple ones with linear, polynomial and exponential structures [26], to complex structures common in machine learning society [27]–[30].

However, from the practical point of view, such approaches become quickly unsatisfactory, especially under varying operating conditions. This issue was recently addressed and published by the authors [31]. In the afformentioned paper, the authors propose an an estimator of stack's internal resistance, such as the one in Fig. 7, by employing an Unscented Kalman filter (UKF).

1) RUL estimation: The main concept of RUL estimation is outlined in Fig. 10. At each time moment k, when current estimates of internal resistances become available, the model parameters are updated. Next, the model is used in open-loop simulations until time step $k + N$, when the modelled feature crosses a pre-defined failure threshold. RUL is then simply N number of steps multiplied by the sampling time.

The RUL plot displayed in lower part of Fig. 10 is obtained by plotting estimated RUL distribution at each time moment k. An example of such a plot is shown in Fig. 11. Explanation of the Fig.is as follows. The x-axis shows running time, while the y-axis displays the RUL distribution. *True* RUL is denoted by dashed thick line and it shortens linearly in time. The blue thick line is the estimate most probable value of RUL, while the colored contours plot corresponding distribution. In the beginning of the experiment, when only a little information about the degradation is available, the predictions are poor. With time passing, the stack degrades further and the model incorporates this new information. Thus, model becomes more accurate and is able to predict RUL more accurately. More details can be found in [31].
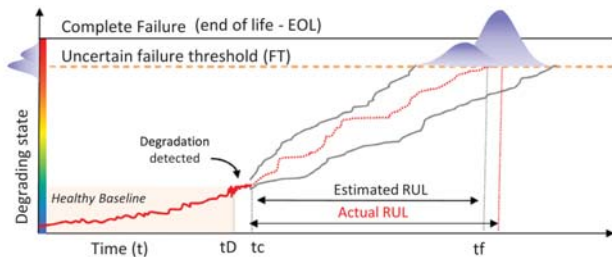
Fig. 10: RUL prediction concept

2) Degradation modelling: In order to be able to perform accurate RUL estimation accurate degradation models are required. Although, degradation is a very complicated process, sufficiently accurate RUL estimations can be achieved using quite simple approaches.

One of such models described in [32] reads:

$$r_d(FU, T, j) = \frac{0.59 FU + 0.74}{1 + e^{\left(\frac{T - 1087}{22.92}\right)}} \left(e^{2.64j} - 1\right). \qquad (6)$$

where r_d is the degradation rate, FU, T and j are fuel utilisation, temperature, and current density, respectively. These are all process variables, which can be manipulated and controlled

in order to minimise the degradation rate. Model (6) was recently adopted for predicting how the fuel cellstack will degrade in future [33]. The parameters, i.e. numeric values in (6) were estimated from the data employing Bayesian approach.

The prediction results are shown in Fig. 11. The results show that 800 hours before the end-of-life, the model accurately predicts the RUL. Furthermore, the variance (uncertainty) of the predictions are decreasing as the time progresses i.e. the model predictions are becoming more precise as we approach the end-of-life.

IV. Conclusion

The presented results addressed three aspects of PHM for SOFCs. A solution is provided regarding the feature extraction using fractional order models of SOFCs. Based on the statistical properties of these features a reliable change detection algorithm is proposed. The algorithm is based on multi-dimensional JR divergence. Finally, a computationally efficient algorithm for RUL estimation is presented.

Taking into account the maturity level of the SOFC technology, we should be witnessing an increased number of installations of such systems. By doing so we will become more self-reliable in the context of energy security, since all the raw materials required for these systems are readily available in our region. Furthermore, the fuel cell technology, in particular SOFC, provides an approach for local fuel production. All these benefits makes the hydrogen technology a viable candidate in the process of seeking alternative to our dependence of fosill fuels.

References

[1] "Energy, transport and ghg emissions trends to 2050."
[2] S. C. Singhal and K. Kendall, *High Temperature Solid Oxide Fuel Cells: Fundamentals, Design, and Applications*. Elsevier, 2003.
[3] F. Barbir, *PEM fuel cells: Theory and practice*. Elsevier, 2013.
[4] Q. Li, D. Aili, and H. Jensen, Eds., *High Temperature Polymer Electrolyte Membrane Fuel Cells*. Springer, 2016.
[5] A. Kirubakaran, S. Jain, and R. Nema, "A review on fuel cell technologies and power electronic interface," *Renew. Sustainable Energy Rev.*, vol. 13, pp. 2430–2440, 2009.
[6] L. Barelli, E. Barluzzi, and G. Bidini, "Diagnosis methodology and technique for solid oxide fuel cells: A review," *Int. J. Hydrogen Energy*, vol. 38, pp. 5060–5074, 2016.
[7] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, and N. Zerhouni, "Prognostics and health management of pemfc – state of the art and remaining challenges," *International Journal of Hydrogen Energy*, vol. 38, no. 35, pp. 15 307 – 15 317, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S036031991302274X
[8] M. Lebold and M. Thurston, "Open standards for condition-based maintenance and prognostic systems," *5th Annual Maintenance and Reliability Conference*, 2001.
[9] D. Marra, C. Pianese, P. Polverino, and M. Sorrentino, *Models for Solid Oxide Fuel Cell Systems: Exploitation of Models Hierarchy for Industrial Design of Control and Diagnosis Strategies*. Springer, 2016.
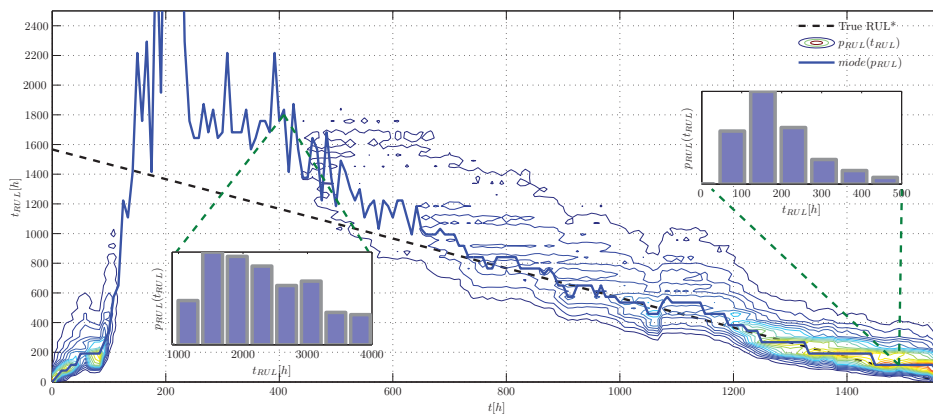
Fig. 11: RUL predictions for the entire duration of the experiment. The contour plot contains three elements: (i.) experimental RUL, (ii.) RUL distribution for each time moment k with support t_{RUL}, and (iii.) mode of the corresponding RUL distribution marking the most probable t_{RUL}.

[10] A. Sorce, A. Greco, L. Magistri, and P. Costamagna, "FDI oriented modeling of an experimental SOFC system, model validation and simulation of faulty states," *Appl. Energy*, vol. 136, pp. 894 – 908, 2014.

[11] M. Sorrentino, D. Marra, C. Pianese, M. Guida, F. Postiglione, K. Wang, and A. Pohjoranta, "On the use of neural networks and statistical tools for nonlinear modeling and on-field diagnosis of solid oxide fuel cell stacks," *Energy Procedia*, vol. 45, pp. 298 – 307, 2014, aTI 2013 - 68th Conference of the Italian Thermal Machines Engineering Association.

[12] E. Pahon, N. Y. Steiner, S. Jemei, D. Hissel, M. Péra, K. Wang, and P. Moçoteguy, "Solid oxide fuel cell fault diagnosis and ageing estimation based on wavelet transform approach," *International Journal of Hydrogen Energy*, vol. 41, no. 31, pp. 13 678 – 13 687, 2016.

[13] P. Boškoski, A. Debenjak, and B. Mileva Boshkoska, *Fast Electrochemical Impedance Spectroscopy As a Statistical Condition Monitoring Tool*, ser. SpringerBriefs in Applied Sciences and Technology. Springer International Publishing, 2017.

[14] P. Polverino, M. Sorrentino, and C. Pianese, "A model-based diagnostic technique to enhance faults isolability in solid oxide fuel cell systems," *Applied Energy*, vol. 204, pp. 1198 – 1214, 2017.

[15] S. M. R. Niya, R. K. Phillips, and M. Hoorfar, "Process modeling of the impedance characteristics of proton exchange membrane fuel cells," *Electrochimica Acta*, vol. 191, pp. 594 – 605, 2016.

[16] M. Heinzmann, A. Weber, and E. Ivers-Tiffée, "Advanced impedance study of polymer electrolyte membrane single cells by means of distribution of relaxation times," *Journal of Power Sources*, vol. 402, pp. 24 – 33, 2018.

[17] "A wavelet tour of signal processing (third edition)," M. Stéphane, Ed. Academic Press, 2009.

[18] D. Iatsenko, P. V. E. McClintock, and A. Stefanovska, "Nonlinear mode decomposition: A noise-robust, adaptive decomposition method," *Phys. Rev. E*, vol. 92, p. 032916, Sep 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.92.032916

[19] F. Ciucci and C. Chen, "Analysis of electrochemical impedance spectroscopy data using the distribution of relaxation times: A bayesian and hierarchical bayesian approach," *Electrochimica Acta*, vol. 167, pp. 439 – 454, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0013468615007203

[20] R. M. Fuoss and J. G. Kirkwood, "Electrical properties of solids. viii. dipole moments in polyvinyl chloride-diphenyl systems*," *Journal of the American Chemical Society*, vol. 63, no. 2, pp. 385–394, 1941.

[21] P. Caliandro, S. Diethelm, and J. Van herle, "Triode solid oxide fuel cell operation under sulfur-poisoning conditions," *Fuel Cells*, vol. 17, no. 4, pp. 457–463, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/fuce.201600196

[22] J. K. Kruschke, *Doin Bayesian Data Analysis*, 2010.

[23] P. Polverino, C. Pianese, M. Sorrention, and D. Marra, "Model-based development of a fault signature matrix to improve solid oxide fuel cell systems on-site diagnosis," *J. Power Sources*, vol. 280, pp. 320–338, 2015.

[24] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, and N. Zerhouni, "Degradations analysis and aging modeling for health assessment and prognostics of PEMFC," *Reliability Engineering & System Safety*, vol. 148, pp. 78–95, 2016.

[25] X. Zhang and P. Pisu, "An unscented kalman filter based approach for the health monitoring and prognostics of a polymer electrolyte membrane fuel cell," in *Annual Conference of Prognostics and Health Management Society*, 2012.

[26] J. Kimotho, T. Meyer, and W. Sextro, "PEM fuel cell prognostics using particle filter with model parameter adaptation," in *IEEE Conference on Prognostics and Health Management*, 2014, pp. 1–6.

[27] R. Silva, R. Gouriveau, S. Jemei, D. Hissel, L. Boulon, K. Agbossou, and N. Steinere, "Proton exchange membrane fuel cell degradation prediction based on adaptive neuro-fuzzy inference systems," *Int. J. Hydrogen Energy*, vol. 39, pp. 11 128–11 144, 2014.

[28] S. Morando, S. Jemei, R. Gouriveau, N. Zerhouni, and D. Hissel, "Fuel cells remaining useful lifetime forecasting using echo state network," in *IEEE Vehicle Power and Propulsion Conference*, 2014, pp. 1–6.

[29] K. Javed, R. Gouriveau, and N. Zerhouni, "Data-driven prognostics of proton exchange membrane fuel cell stack with constraint based summation-wavelet extreme learning machine." in *6th international conference on fundamentals and development of fuel cells*, 2015, pp. 1–8.

[30] A. Hochstein, H. Ahn, Y. Leung, and M. Denesuk, "Switching vector autoregressive models with higher-order regime dynamics," in *IEEE Conference on Prognostics and Health Management (PHM)*, no. 1-10, 2014.

[31] B. Dolenc, P. Boškoski, M. Stepančič, A. Pohjoranta, and Đ. Juričić, "State of health estimation and remaining useful life prediction of solid oxide fuel cell stack," *Energy Convers. Manag.*, vol. 148, pp. 993 – 1002, 2017.

[32] V. Zaccaria, D. Tucker, and A. Traverso, "A distributed real-time model of degradation in a solid oxide fuel cell, part i: Model characterization," *J. Power Sources*, vol. 311, pp. 175–181, 2016.

[33] B. Dolenc, "Characterisation of coupling functions between process and degradation dynamics in solid oxide fuel cells," *Philosophical Transactions A, to appear*, 2019.

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS & TELECOMMUNICATION SYSTEMS AND TECHNOLOGY

Comparison of Wave Digital and Circuit Models of Microstrip Single-Stub L-Tuners

Biljana P. Stošić and Nebojša S. Dončov

Abstract – **A simple wave digital models of microstrip single-stub tuners have been presented in this paper. Advanced Design System (ADS) software and MATLAB/Simulink environment are used to design and simulate the investigated microstrip circuits and their models. A comparative study is carried out, the results of which are summarized here. The comparison of the simulation results of the investigated models shows that there is a good agreement between them.**

Keywords – **Microstrip circuit, circuit models, wave digital models, single-stub tuners.**

I. INTRODUCTION

Recently, much work has been concentrated on the development of wave digital (WD) approach for modeling and analysis of different physical systems. Research on application of WD structures for electromagnetic (EM) field simulation is reported in many literatures, e.g. Bilbao [1] and Russer et al. [2]. Maggioni in [3] presented an application of Advanced Design System (ADS) to simulations of different microstrip structures based on their wave digital network (WDN) representations. In [4], Franken et al. gave a framework for the automated generation of the wave digital structures, and the reference circuit is assumed to comprise arbitrary connection types.

The concepts of WDNs have their origins in the field of filter design, where they are designated more specifically as wave digital filters [5]-[8].

Basically, it is important to analyze microstrip structure quickly, to get proper information in short time. In practice, many models have been in use: EM models, electric circuit models, etc. EM models are highly accurate, as they aim to model interactions in whole structure; however this makes them extremely complex and time consuming. Electric circuit models however are less complex and provide sufficient representation of a structure.

The modeling of microwave structures comes with challenges, one of which is to obtain the wave digital model of the different junctions with several arms (i.e. transmission lines). This study is carried out to develop WD models of microstrip single-stub L-tuners and to compare those models with the other ones. Various microwave devices require the use of tuners in experimental verification of their performances.

The authors are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mails: biljana.stosic@elfak.ni.ac.rs, nebojsa.doncov@elfak.ni.ac.rs.

This paper describes modeling and analyzing procedures for microstrip circuits based on use of one-dimensional wave digital approach. Frequency responses are obtained by direct analysis of the block-based networks formed in Simulink toolbox of MATLAB environment. This wave-based method allows an accurate and efficient analysis of different microwave structures.

Amplitude-frequency responses of S_{21} parameter for different symmetrical microwave circuits are discussed in the previous paper [9]. In this paper, a comparative study of S_{21} and S_{11} parameters obtained from wave digital and circuit models is carried out. Asymmetric microwave circuits with very thin conductor lines are modeled and analyzed.

II. WAVE DIGITAL NETWORK

Some general properties and basic aspects of the one-dimensional wave digital approach that are of considerable importance, related to this work, will be discussed here briefly.

In order to generate a method for construction of algorithm that efficiently implements microstrip structures utilizing cross-junction opened stubs, their layouts have to be observed. Fig. 1 illustrates the known configuration of considerable interest - a microstrip structure with cross-junction opened stubs; it is assumed that this structure consists only of uniform segments assigned from UTL1 to UTL4.

Fig. 2 illustrates the equivalent WDN of the known microstrip circuit associated with Fig. 1. In accordance with the general method outlined in Section II in [9], this WDN is created. The WDN is a general resultant network, i.e. structure model comprises $n_t = \sum_{k=1}^{4} n_k$ unit elements, one four-port parallel adaptor and two two-port series adaptors.
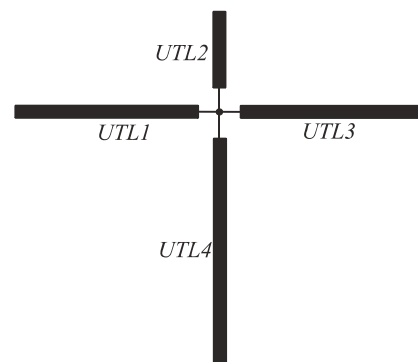
Fig. 1. Microstrip circuit with cross-junction opened stubs: symbolical representation of uniform segment connection
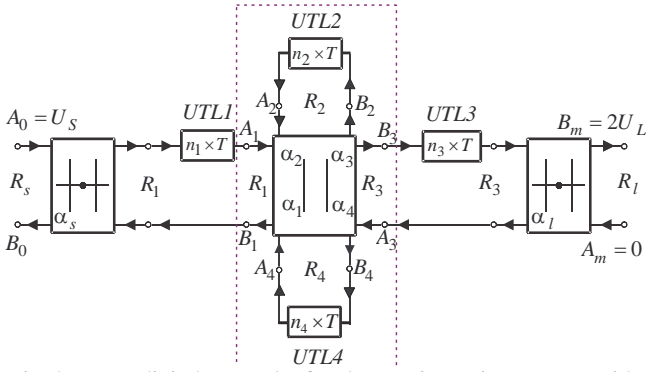
Fig. 2. Wave digital network of a planar microstrip structure with cross-junction opened stubs

(a)

(b)

Fig. 3. General Simulink models of WDN from Fig. 2:
(a) S_{11} and S_{12} parameters, and (b) S_{21} and S_{22} parameters

The equation set for four-port parallel adaptor, symbolically presented in Fig. 2, is:

$$B_k = (\alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3 + \alpha_4 A_4) - A_k, \qquad (1)$$

where the adaptor coefficients

$$\alpha_k = \frac{2 G_k}{\sum_{j=1}^{n} G_j}, \ k = 1, 2, 3, 4, \qquad (2)$$

depend on port conductances $G_k = 1/R_k$, with R_k being port resistances, and

$$\sum_{k=1}^{4} \alpha_k = 2. \qquad (3)$$

According to the last equation, it is possible to eliminate one coefficient, i.e. α_4, in order to reduce number of adders in resultant wave digital network of this adaptor. In this case, the new equations are:

$$B_4 = A_4 - \sum_{k=1}^{3} \alpha_k (A_4 - A_k), \qquad (4)$$

$$B_k = B_4 + (A_4 - A_k), \ k = 1, 2, 3. \qquad (5)$$

III. MATLAB/Simulink Model Implementation

MATLAB is an excellent tool for simulating structure, and for creating the valuable "proof of concept". Block diagram is a representation of physical structure using blocks. Individual blocks can be put together to represent the structure in block diagram form. Individual blocks can be the basic blocks or they can be subsystems. They are considered in the text given below.

Simulation of Simulink model represents a series of MATLAB and Simulink commands and functions which are used for its creation. Response is obtained directly in the time domain, and Fourier transformation is used for frequency response calculation. A major part here is formed Simulink model (slx-file), whereby the slx-file is run by m-file that is provided for initialization, response calculation and plotting.

A correct equivalent wave-based model depends on structure geometry. The models represented here are upgraded to the previous models given in [9] in order to calculate all S-parameters, not just S_{21}.

Fig. 3 illustrates the general Simulink model formed in accordance with the present wave digital network shown in Fig. 2. As stated hereinabove in association with Fig. 2, Fig. 3 comprises the blocks **Line_1**, **Stub_2**, **Line_3** and **Stub_4** representing models of uniform segments, the block **ADP_T1S2T3S4** representing four-port parallel adaptor, as well as the blocks **ADP-In** and **ADP-Out** representing two-port series adaptors. The two-port adaptors at the ends are used for matching source and load resistances to the rest of the WDN. The two-port adaptor coefficients are

$$\alpha_s = (R_s - Z_{c1}) / (R_s + Z_{c1}), \qquad (6)$$

$$\alpha_l = (Z_{c3} - R_l) / (Z_{c3} + R_l). \qquad (7)$$

Simulink model of four-port parallel adaptor with port 4 being dependent, depicted in Fig. 4, is formed in accordance with Eqs. (4)–(5).
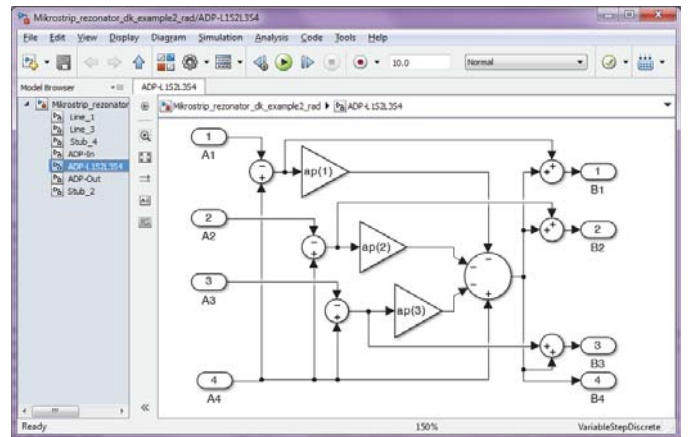
Fig. 4. Simulink model of four-port parallel adaptor with port 4 being dependent

IV. RESULTS AND DISCUSSION

In radio frequency (RF) engineering, stub tuners are often very convenient because they are simple to implement and cheap to manufacture: they only require more of the same material used to make the transmission line. The stub can simply be fabricated as a part of transmision media onto the PCB (Printed Circuit Board) along with the rest of the circuit. The goal that stubs are designed to accomplish is to cancel out the reactive component of the load to be matched, thus they will only work at a specific frequency.

Validation of the proposed circuit modeling is demonstrated and discussed in this section as follows:

- simulating the developed wave digital models,
- extracting the model S-parameters,
- plotting parameters.

In order to evaluate these results, a comparison is carried out. Different microwave structures based on microstrip technology are simulated using the same characteristics of the substrate. The structures chosen for the comparison are made on cheap FR-4 substrate of dielectric thickness $h = 0.8\,\text{mm}$, relative permittivity $\varepsilon_r = 4.5$, loss tangent $\tan\delta = 0.02$, and metallization thickness $t = 35\,\mu\text{m}$.

The circuits depicted in next figures have been manufactured by the other researchers and their S-parameters have been measured with an Agilent N5227A vector network analzyer and the results are shown in [10]. The measured results are then de-embedded in order to neutralize the influence of the SMA connectors. Circuits with very thin microstrip lines having high characteristic impedances are modeled, i.e. $0.3\,\text{mm}$ and $0.6\,\text{mm}$ in Figs. 5a and 6a, respectively. The effect of the manufacturing tolerances of narrow lines can influence the power division between the ports.

The models are build in both ADS and MATLAB/Simulink and illustrated in Figs. 5 and 6. The models are than simulated, and plots of the magnitudes of the reflection coefficient (S11) and transmission coefficient (S21) versus frequency are generated. The comparison between the results of simulation is given in Table I. Table I summarizes the shift in resonant frequencies discussed for the electromagnetic simulator, circuit simulator, WD model and measured results.

The agreement between this simulated results and the measured ones is very good despite a frequency shift and an increase in the insertion losses. The degradation between results could be attributed to increased dielectric losses in the substrate fabrication error. The proposed approach is implemented on a processor Intel(R) Core(TM) i5-3470 CPU @ $3.20\,\text{GHz}$. A time for a response calculation from WD models is very low, i.e. $6.6\,s$ and $5.2\,s$, respectively. ADS Momentum requires more than $20\,min$ for response calculation. Generally, the WD method provides the fast simulations versus complex and time consuming 3D models.

TABLE I
SHIFTING OF RESONANT FREQUENCIES FOR SIMULATED AND MEASURED RESULTS

Simulation/Measured /WD results	Resonant frequency [GHz]	
	Example 1	Example 2
WD model	2.6	1.45
ADS circuit model	2.95	2.25
ADS Momentum	2.75	1.8
Measured [10]	2.8	1.85

(a)

(b)

(c)

(d)
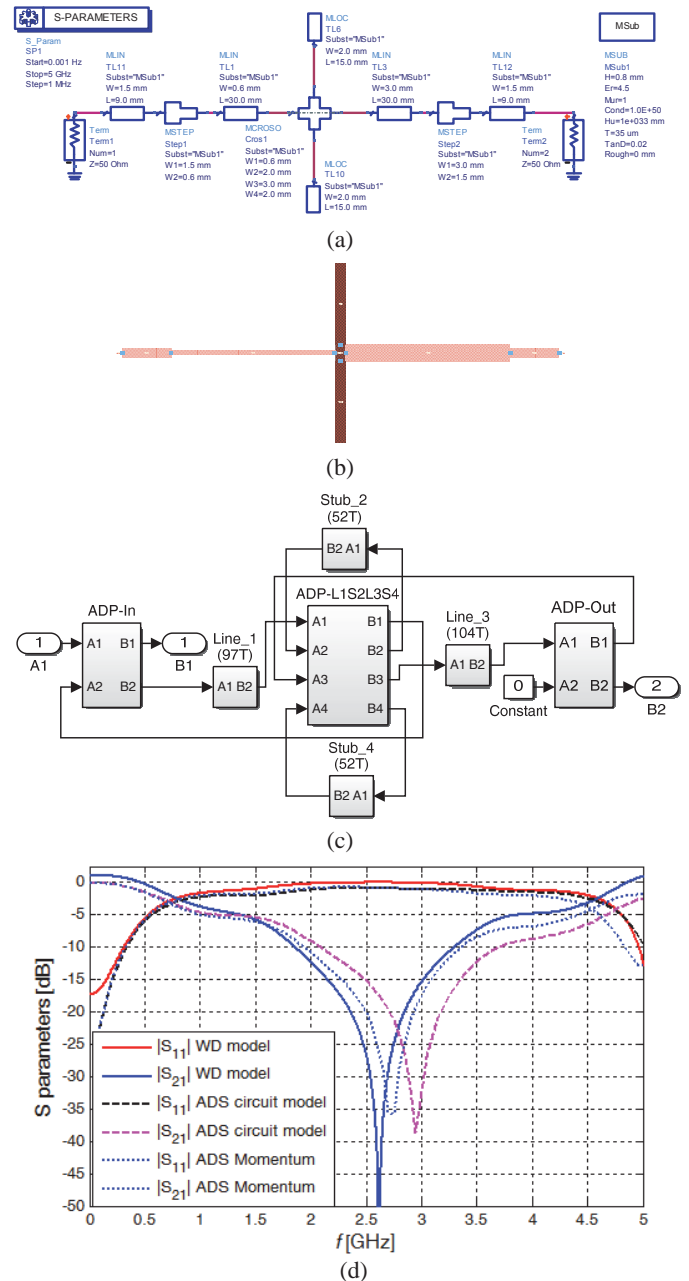
Fig. 5. Microstrip single-stub L-tuner (Example 1): (a) circuit model, (b) layout, (c) wave digital model, (d) result comparison

(a)

(b)

(c)

(d)
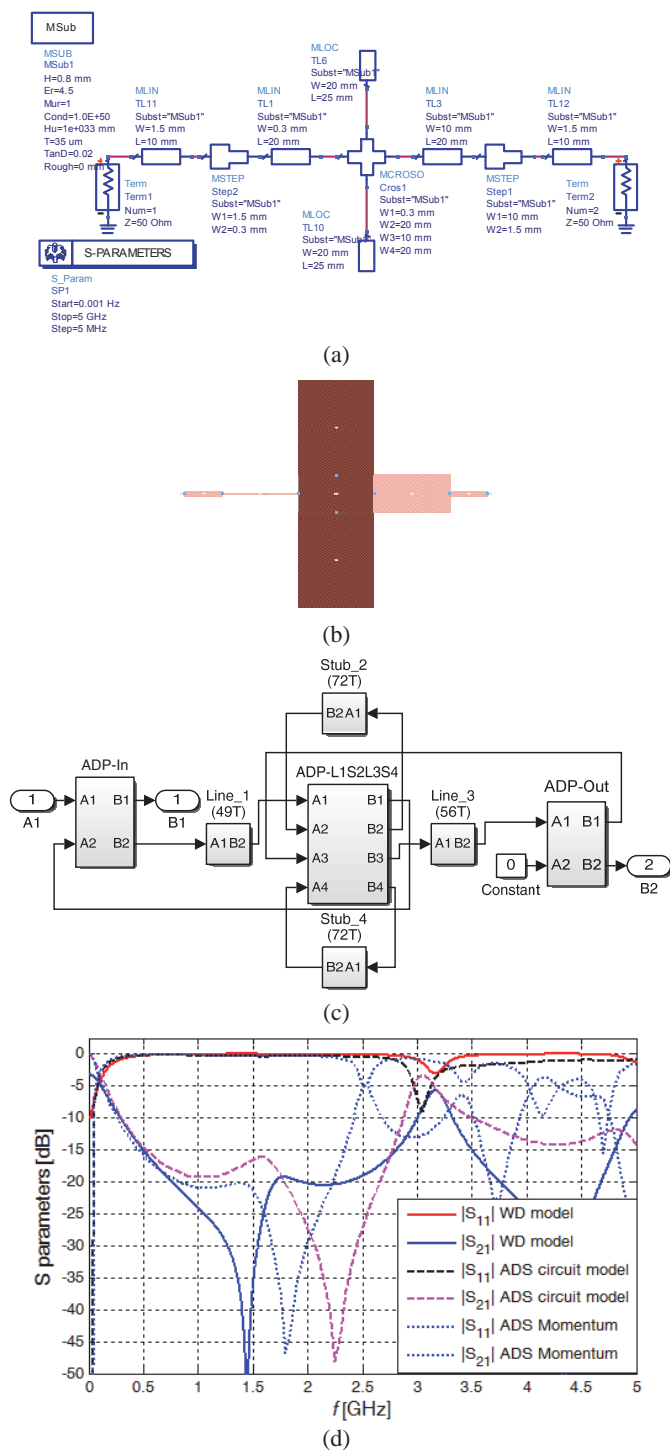
Fig. 6. Microstrip single-stub L-tuner (Example 2): (a) circuit model, (b) layout, (c) wave digital model, (d) result comparison

V. CONCLUSION

Single stub tuners have the ability to match any load impedance that has a positive real component, but they require placement at a specific distance from the load to maintain matching. This paper presents an investigation of both wave digital model and circuit model for microstrip single-band tuners. The objective of the proposed models is to reduce time

of simulations of different time-consuming electromagnetic structures.

Using ADS software the equivalent circuit models are simulated. Using MATLAB/Simulink enviroment wave digital models are generated and simulated. The development of the model would aid in reducing the time taken to design microstrip circuit in electromagnetic simulator, due to the fact that the developed WD elements can be built-in MATLAB simply by using library elements. Synthesis of wave digital model of structure with more or less elements is easy task with simple addition or substraction of existing blocks.

The results obtained from both wave digital and circuit models are in good agreement with one another, as well as the measured results. It can be observed also how dimensions of the microstrip lines of the circuit geometry contributes to the respective resonant frequencies.

REFERENCES

[1] S. Bilbao, *Wave and Scattering Methods for Numerical Simulation*, Wiley, Hoboken, New Jersey, 2004.
[2] J.A. Russer, Y. Kuznetsov, and P. Russer, "Discrete-time Network and State Equation Methods Applied to Computational Electromagnetics", Microwave Review, vol. 16, no. 1, pp. 2-14, 2010.
[3] F. Maggioni, "Time Domain Electrical Simulation using Equivalent Digital Wave Networks in ADS", ADS User Group Meeting, Rome, May 13, 2009.
[4] D. Franken, J. Ochs, and K. Ochs, "Generation of Wave Digital Structures for Networks Containing Multiport Elements", IEEE Transaction on Circuits and Systems I: Regular Papers, vol. 52, no. 3, pp. 586-596, 2005.
[5] A. Fettweis, "Digital Circuits and Systems", *IEEE Transactions on Circuits and Systems*, vol. CAS-31, no. 1, January, 1984, pp. 31-48.
[6] A. Fettweis, "Wave Digital Filters: Theory and Practice", *Proc. IEEE*, vol. 74, 1986, pp. 270-327.
[7] W.K. Chen, *The Circuits and Filters Handbook*, CRC Press, 1995 (Wave Digital Filters, pp. 2634-2661).
[8] M.V. Gmitrovic, *Microwave and Wave Digital Filters*, Faculty of Electronic Engineering, Niš, 2007.
[9] 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, December 2011, pp. 31-38.
[10] B. Bukvić, A. Ilić, and M.M. Ilić, "Comparison of Approximate and Full-Wave Electromagnetic Numerical Modeling of Microstrip matching Networks", 2015 International Conference on Electromagnetics in Advanced Applications (ICEAA), Italy, Torino, September 7-11, 2015, pp. 76-79.

A Neural Approach for Lumped Element Circuit Based Inverse Modeling of RF MEMS Switches

Rohan Dhuri[1], Tomislav Ćirić[2], Olivera Pronić-Rančić[2],
Vera Marković[2], Zlatica Marinković[2]

Abstract – **In this paper a novel neural approach for inverse modeling of RF MEMS switches is presented. The approach is developed for a lumped element circuit. Artificial neural networks are applied to determine one lateral dimension of the switch bridge for given value of the lumped element circuit model inductance and the other bridge lateral dimension. The approach is validated through comparison of the switch dimensions obtained by the inverse model and the corresponding target values.**

Keywords –Artificial neural networks, inverse modeling, RF MEMS switches.

I. INTRODUCTION

RF switches are among the most important components of RF transceivers. In the recent years RF MEMS switches have been becoming widely used owing to the advantageous properties that they have comparing to their mechanical and semiconductor counterparts. They are very small, extremely linear, can be integrated and allow easy re-configurability or tunability of a system [1].

An RF MEMS switch has an actuation section and an electrical element which can be categorized according to [1] - [4] :

- Actuation scheme: to electrostatic, magnetostatic, piezoelectric and thermal;
- Geometrical configuration: to vertical configuration and horizontal configuration;
- Electrical configuration: to ohmic contact / series circuit and capacitive / shunt circuit. The series switch has a switch series in the signal path. The shunt switch has a switch parallel to the signal path [4].

In this work a capacitive shunt RF MEMS switch is considered [5].

Design of RF MEMS switches requires reliable and accurate models. Modeling is performed simultaneously in standard full-wave electromagnetic simulators and in mechanical simulators. The simulations are quite time consuming, which is especially emphasized when optimization of the switch dimensions is to be performed to meet the desired electrical or mechanical properties.

In the case of series shunt switches, important dimensions to be optimized are lateral dimensions of the switch bridge, as

[1]R. Dhuri is with ALTEN GmbH, Munich, Germany E-mail: rohan.dhuri3@gmail.com.
[2] T. Ćirić, O. Pronić-Rančić, V. Marković and Z. Marinković are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia
E-mail: cirict78@gmail.com, olivera.pronic@elfak.ni.ac.rs, vera.markovic@elfak.ni.ac.rs, zlatica.marinkovic@elfak.ni.ac.rs

it will be described in details in Section II. RF MEMS switch modeling based on the artificial neural networks (ANNs) [6] has been proposed as a suitable alternative to the standard modeling approach [7]-[17]. Having in mind that the ANNs have a very short response time, by exploiting the ANN models the process of simulation and optimization of the RF MEMS switches can be speed up. Moreover, as a further improvement of the design and modelling efficiency, inverse models of RF MEMS switches based on ANNs have been developed [12]-[17]. The inverse models are trained to calculate the switch bridge dimensions for given switch resonant frequency and/or switch actuation voltage.

In this paper the inverse modeling approach is extended to the lumped element equivalent circuit model. Namely, inverse modeling assumes determination of the switch bridge dimensions for given values of the equivalent circuit elements.

The paper is organized as follows. In Section II the capacitive RF MEMS switch modeled in this work is described and the description of the considered equivalent circuit model is given. The proposed inverse ANN modeling approach is presented in Section III. The numerical results and discussion are given in Section IV. Section V contains concluding remarks.

II. MODELED DEVICE

The RF MEMS switch analyzed in this work is a coplanar capacitive shunt switch. Top view of the fabricated switch is shown in Fig. 1. a) and the cross-sectional view is shown in Fig. 1. b). The switch is fabricated at FBK in Trento in an 8 layer silicon micromachining process [5].

The signal line below the bridge is made of a thin aluminum layer. Adjacent to the signal line, the DC actuation pads made by polysilicon are placed. The bridge is a thin membrane connecting both sides of the ground. The switch has a structured membrane with a solid center part and a fingered part close to the anchors for adjusting electrical and mechanical parameters. The length of the fingered part, L_f,

and solid part, L_s, can be changed in order to change the resonant frequency of the switch. The actuation mechanism used here is the electrostatic actuation. For a certain value of the voltage applied to the DC pads, the bridge comes down and touches a coplanar waveguide (CPW) centerline. The inductance of the bridge and the fixed capacitance between signal line and bridge form a resonant circuit to the ground. [14].

a)

b)

c)
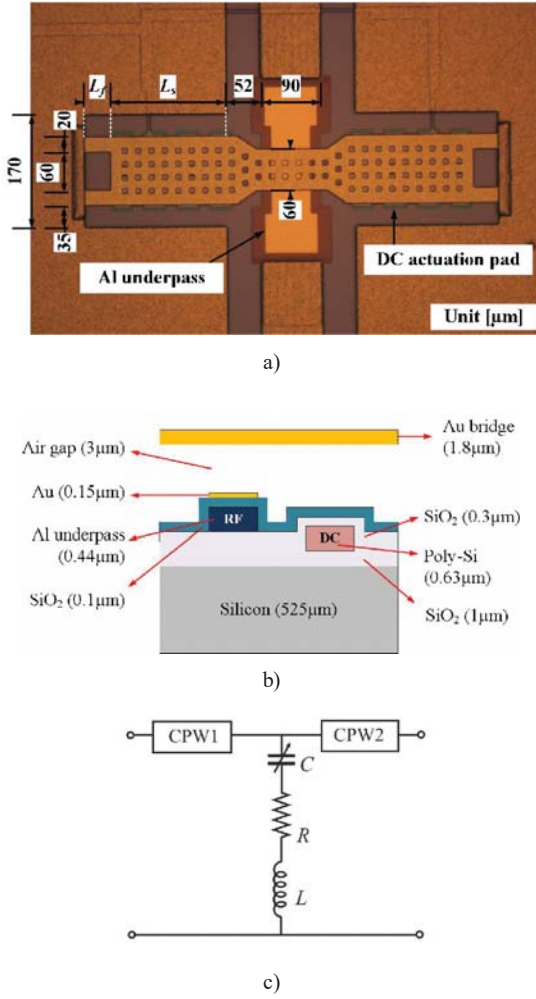
Fig. 1. a) Top-view of the realized switch; b) schematic of the cross-section with 8 layers in FBK technology [5]; c) equivalent circuit of the RF MEMS switch RLC lumped element model

This process is related to the switch features and mechanical/material properties, such as a DC pad size and location, a bridge spring constant and residual stress, bridge shapes or supports, etc.

The considered switch can be represented by a simplified equivalent circuit model, as it is shown in Fig. 1.c), [15] It consists of the resistance R, the inductance L and the capacitance C, as well as two 50Ω CPW lines.

The switch resonant frequency can be calculated as:

$$f_{res} = \frac{1}{2\pi\sqrt{LC}} . \qquad (1)$$

The switch capacitance in the membrane down-state, considered in this case, is calculated from the layout using the following expression [1]:

$$C = \frac{\varepsilon_0 \varepsilon_r A}{t_d} , \qquad (2)$$

where ε_0 is the dielectric permittivity, ε_r is the relative dielectric permittivity, t_d is the distance between the two plates forming the capacitance and A is the surface of the plates forming the capacitance. As can be seen, the capacitance does not depend on the bridge lateral dimensions, i.e., it is constant with the changes of the bridge lateral dimensions. The other two elements, R and L, depend on the bridge lateral dimensions L_s and L_f.

During the design of switch, if a resonant frequency is given, which is usually the case, the inductance can be straightforwardly calculated from Eq. 1 as:

$$L = \frac{1}{4\pi^2 f_{res}^2 C} \qquad (3)$$

The only parameter which is to be determined by optimizations in a circuit simulator is the resistance.

As far as the determination of the bridge lateral dimensions is concerned, they can be determined from the given inductance as explained in the following section.

III. PROPOSED INVERSE ANN MODEL

As explained in the previous parts of the paper, the switch dimensions depend on the lateral dimensions, i.e., the bridge solid part (L_s) and the bridge fingered part (L_f). Once the switch inductance is calculated for the given resonant frequency, the lateral dimensions can be determined by ANN based models, as shown in Fig. 2.

It should be mentioned that the inductance is not unique for different values of the dimensions, i.e., one inductance value matches several values of the bridge solid and fingered parts, but is unique to a combination of L_s and L_f. Having that in mind, it is not possible to find both L_s and L_f simultaneously by using a single ANN. Therefore, the two ANNs have to be used to determine L_s and L_f separately. One dimension has to be fixed to determine the other dimension. The ANN shown in Fig. 2 a) is used to determine the bridge solid part for given value of the bridge fingered part and the ANN shown in Fig. 2 b) is used to determine the bridge fingered part for the given bridge solid part.

To train the ANNs, it is necessary to calculate the equivalent circuit inductance for different combinations of the bridge lateral dimensions. This can be done in the full-wave simulators, where the S-parameters are calculated for different combinations of the bridge lateral dimensions. The resonant frequency is found for each case and is further used to calculate the equivalent circuit inductance. To have a complete equivalent circuit model, the resistance is determined by the optimizations in a circuit simulator. Moreover, an additional model for the resistance versus the lateral dimensions can be modeled by an ANN, as shown in [15].

16

a)

b)
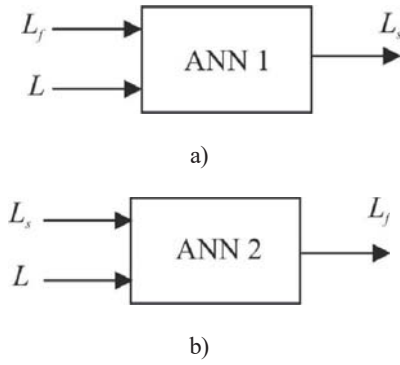
Fig. 2: ANN inverse model for a) L_s , b) L_f

IV. NUMERICAL RESULTS AND DISCUSSION

The training of the ANNs was performed by different training algorithms, and it was found that for the available data the best results in terms of learning (modeling accuracy for the training values) and generalization (modeling accuracy for the test values) were achieved by using the Bayesian regularization algorithm.

For all models, the ANNs with different number of hidden neurons were trained and compared, and the ones giving the best accuracy were chosen as the final ones. The accuracy was tested by estimating the absolute test error as well as the relative test error.

The training of ANN 1 was done using ten data samples, as shown in Table I. The chosen ANN has two hidden layers of 25 neurons each. The test was done for the training data (Table I) and the test data (Table II). In both cases, the absolute error is within the limit of 3μm, which is within the fabrication tolerances. The relative errors are less than 1%.

The ANN 2 was trained using the same combinations of the lateral dimensions as in the case of the ANN 1. The ANN showing the best accuracy has two hidden layers of 50 neurons each. The test statistics for the training and test input combinations are given in Table III and Table IV, respectively. The chosen ANN has two hidden layers of 50 neurons each. As in the previous case, the absolute error is within the limit of 3μm for both training and testing samples. The relative error is except in a very few cases less than 1%.

The inverse modeling where the resonant frequency is taken as the ANN input instead the inductance can be seen in [12]-[15]. However, that alternative way to perform inverse modeling requires additional simulations (either in a full-wave EM simulator or preferably by using an ANN model trained to determine the resonant frequency for the given lateral dimension) to get more training data needed to achieve the acceptable absolute error limit of 3μm. The inverse model using inductance as the ANN input does not need any additional simulations and gives a good generalization with use of only 10 samples for training compared to 814 samples used for training of the inverse model using resonant frequency as shown in [12].

TABLE I: TEST STATISTICS FOR THE TRAINING SET FOR THE ANN MODEL FOR L_s DETERMINATION

L_f [μm]	L [pH]	L_s - target [μm]	L_s - ANN [μm]	Abs. error [μm]	Rel. error [%]
50	11.35	50	49.82	0.18	0.36
100	17.04	50	51.02	1.02	2.04
50	27.21	200	198.93	1.07	0.54
100	33.38	200	197.75	2.25	1.13
50	39.85	300	302.55	2.55	0.85
100	46.63	300	300.05	0.05	0.02
50	52.17	400	399.43	0.57	0.14
100	59.97	400	402.90	2.90	0.73
50	65.24	500	501.71	1.71	0.34
100	71.23	500	497.91	2.09	0.42

TABLE II: TEST STATISTICS FOR THE TEST SET FOR THE ANN MODEL FOR L_s DETERMINATION

L_f [μm]	L [pH]	L_s - target [μm]	L_s - ANN [μm]	Abs. error [μm]	Rel. error [%]
25	30.06	250	252.68	2.68	1.07
75	36.69	250	251.05	1.05	0.42
25	41.93	350	348.05	1.95	0.56
75	49.28	350	347.11	2.89	0.83

TABLE III: TEST STATISTICS FOR THE TRAINING SET FOR THE ANN MODEL FOR L_f DETERMINATION

L_s [μm]	L [pH]	L_f - target [μm]	L_f - ANN [μm]	Abs. error [μm]	Rel. error [%]
50	11.35	50	49.96	0.04	0.08
50	17.04	100	100.06	0.06	0.06
200	27.21	50	50.29	0.29	0.58
200	33.38	100	99.39	0.39	0.39
300	39.85	50	50.25	0.25	0.50
300	46.63	100	100.76	0.76	0.76
400	52.17	50	48.76	1.24	2.48
400	59.97	100	100.16	0.16	0.16
500	65.24	50	51.42	1.42	2.84
500	71.23	100	99.05	0.95	0.95

TABLE IV: TEST STATISTICS FOR THE TEST SET FOR THE ANN MODEL FOR L_f DETERMINATION

L_s [μm]	L [pH]	L_f - target [μm]	L_f - ANN [μm]	Abs. error [μm]	Rel. error [%]
250	30.06	25	23.93	1.07	4.28
250	36.69	75	74.31	0.69	0.92
350	41.93	25	22.72	2.28	9.12
350	49.28	75	73.41	1.59	2.13

V   CONCLUSION

Inverse ANN models developed to determine the bridge dimensions of RF MEMS switches could be very helpful in the RF MEMS switch design. They can be used for fast determination of the dimensions to get the desired inductance or resonant frequency. The inverse modeling approach proposed in this paper is combined with the equivalent circuit model. Namely, an ANN is used to determine one of the two bridge lateral dimensions, for the given value of the other lateral dimension and the equivalent circuit inductance calculated for the desired resonant frequency. According to the test statistics on the training data and on the test data not used for the model development, the difference between the determined and target values is smaller than the deviations from the nominal values which might appear during the fabrication process. Moreover, it was concluded that this approach enables achieving the same accuracy as the accuracy achieved with the previously proposed models, where the resonant frequency is used as an ANN input instead of the equivalent circuit inductance, but with a significantly smaller training data set.

ACKNOWLEDGEMENT

REFERENCES

[1]   G. M. Rebeiz, *RF MEMS Theory, Design, and Technology*. New York: Wiley, 2003.

[2]   M. Daneshmand, R. R. Mansour: "Redundancy RF MEMS Multiport Switches and Switch Matrices", *Journal of Microelectromechanical Systems,* vol. 16 no. 2, 2007, pp. 296–303.

[3]   K.A. Jose Vijay, K. Varadan, K.J. Vinoy, *RF MEMS and Their Applications,* Wiley, 2003.

[4]   G. Schiavone, Marc PY Desmulliez, and Anthony J Walton: "Integrated magnetic MEMS relays: Status of the technology", *Micromachines*, vol. 5, no.3, 2014, pp. 622–653.

[5]   S. DiNardo, P. Farinelli, F. Giacomozzi, G. Mannocchi, R. Marcelli , B. Margesin, P. Mezzanotte, V. Mulloni, P. Russer, R. Sorrentino, F. Vitulli, L. Vietzorreck, "Broadband RF-MEMS based SPDT", *Proc. European Microwave Conference 2006, Manchester*, Great Britain, September 2006.

[6]   Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Boston, MA: Artech House, 2000.

[7]   Y. Lee, D. S. Filipovic, "Combined full-wave/ANN based modelling of MEMS switches for RF and microwave applications," *Proc. of IEEE Antennas and Propagation Society International Symposium*, July 2005, vol. 1A, pp. 85-88.

[8]   Y. Mafinejad, A. Z. Kouzani, K. Mafinezhad , "Determining RF MEMS switch parameter by neural networks," *Proc. of IEEE Region 10 Conference TENCON 2009*, Jan. 2009, pp. 1-5.

[9]   Y. Lee, Y Park, F. Niu, D. Filipovic, "Artificial neural network based macromodeling approach for two-port RF MEMS resonating structures," *IEEE Proceedings of Networking, Sensing and Control*, March 2005, pp. 261 – 266.

[10] Y. Gong, F. Zhao, H. Xin, J. Lin, Q. Bai, "Simulation and Optimal Design for RF MEMS Cantilevered Beam Switch," *Proc. of International Conference on Future Computer and Communication* (FCC '09), June 2009, pp. 84-87.

[11] T. Kim, Z. Marinković, V. Marković, M. Milijić, O. Pronić-Rančić, L. Vietzorreck, "Efficient Modelling of an RF MEMS Capacitive Shunt Switch with Artificial Neural Networks," *Proc. of URSI-B 2013 International Symposium on Electromagnetic Theory*, Hiroshima, Japan, May, 2013, pp. 550-553.

[12] Z. Marinković, T. Ćirić, T. Kim, L. Vietzorreck, O. Pronić-Rančić, M. Milijić, V. Marković, "ANN Based Inverse Modeling of RF MEMS Capacitive Switches", *11th Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services TELSIKS 2013*, Niš, Serbia, October 16-19, 2013, pp. 366-369

[13] Z. Marinković, A. Aleksić, T. Ćirić, O. Pronić-Rančić, V. Marković, L. Vietzorreck, „Inverse electro-mechanical ANN model of RF MEMS capacitive switches - applicability evaluation," *XLX Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST 2015*, Sofia, Bulgaria, June 24-26, 2015, pp. 157-160.

[14] Z. Marinković, T. Kim, V. Marković, Marija Milijić, O. Pronić-Rančić, Tomislav Ćirić, 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, July 2016..

[15] Z. Marinković, V. Marković, T. Ćirić, L. Vietzorreck, O. Pronić-Rančić,"Artifical neural networks in RF MEMS switch modeling", *Facta Universitatis, Series: Electronics and Energetics*, vol. 29, no. 2, pp. 177–191, 2016.

[16] T. Ćirić, Z. Marinković, O. Pronić-Rančić, V. Marković, L. Vietzorreck, "ANN approach for modeling of mechanical characteristics of RF MEMS capacitive switches - an overview", *Microwave Review*, vol. 23, no. 1, pp. 25-34, June 2017.

[17] T. Ćirić, R. Dhuri, Z. Marinković, O. Pronić-Rančić, V. Marković, L. Vietzorreck, "Neural Based Lumped Element Model of Capacitive RF MEMS Switches", *Frequenz*, vol. 72, no. 11-12, pp. 539-546, November 2018.

Design of CPW-fed asymmetrical slot array for K-band applications

Marija Milijic[1] and Branka Jokanovic[2]

Abstract – **This paper discusses the design of an antenna array consisting of rectangular slots positioned asymmetrically relative to CPW feeding line. The proposed asymmetrical slots offer greater flexibility in antenna design and easier control of crucial antenna parameters such as gain, side lobe suppression and bandwidth. Additionally, the proposed antenna is intended for frequency range 24.25-27.5 GHz and therefore it is suitable for future great capacity broadband 5G technologies.**

Keywords – **Antenna array, Asymmetrical slot antennas, CPW-fed antennas, CPW T-junction.**

I. INTRODUCTION

With the rapid growth of the wireless communication system, sub-6 GHz frequency range has become overcrowded. Therefore, the operation bandwidth of the upcoming 5G networks will be at higher frequencies compared with previous generations of mobile communication networks [1]. Several promising millimetre-wave bands have been released by the International Telecommunication Union (ITU) for the 5G wireless communication system that include the 24.25 – 27.5 GHz, 37 – 40.5 GHz, 66 – 76 GHz bands [2]. Meanwhile, the Federal Communications Commission (FCC) has considered the spectrum of approximately 11 GHz above 24GHz for flexible, mobile and fixed wireless broadband for the next-generation 5G networks and technologies in the United States [3]. Among the available millimetre wave spectrum, frequency around K band is extensively highlighted for 5G because of lower atmospheric absorptions and relatively lesser attenuations while these effects become more prominent at higher frequencies [4].

Furthermore, 5G generation of mobile networks are intended to connect hundreds different devices creating crucial demands for services of great broadband capacities and transmission speeds [1]. Consequently, new challenges to design of millimetre-band antennas have come out requiring antennas composed of dozens of radiating elements. Antenna arrays feature better radiation characteristics, combined with the reduced sizes and higher gains. CPW (coplanar waveguide) - fed antennas have received much attention for finding applications in 5G mobile communication systems due to their wide bandwidth, low cost, light weight, small size, and ease of fabrication [5-7].

[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
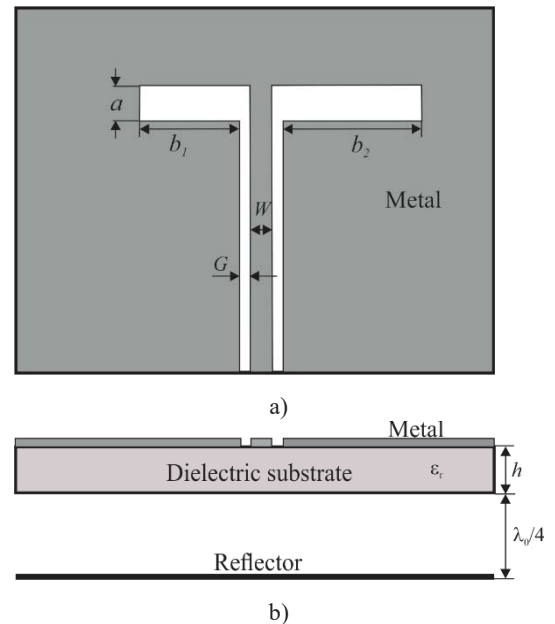
Fig. 1. CPW-fed asymmetrical rectangular slot: a) top view b) side view.

This paper presents a centrally offset fed rectangular slot array designed for wide-band applications in 5G frequency range 24.25-27.5 GHz. The slots are located asymmetrically relative to CPW feeding line in order to easier control of crucial antenna parameters such as gain, side lobe suppression and bandwidth. Aside from planar structure and inexpensive fabrication, given simulated results report that the proposed antenna array has 1.85 GHz bandwidth in terms of radiation pattern and 1.2 GHz bandwidth in terms of S_{11} parameter with 16 dBi average gain that make this class of antennas a good candidate for a variety of 5G communication applications.

II. CPW-FED ASYMMETRICAL SLOT ANTENNA

Geometrical view of the proposed CPW-fed asymmetrical slot antenna is shown in Fig. 1. It is modelled using a substrate with thickness of h=0.508 mm and relative permittivity of ε_r=2.54. The rectangular slot is with width a and arms of different length b_1 and b_2 causing that the CPW feed line is not positioned in the middle of slot. The widths of the strip and gap (W and G) of the CPW feed line, whose impedance is around 120 Ω, are 0.3 mm and 0.375 mm, respectively. The antenna's metal surface is of size 15 x 14 mm². At the distance $\lambda_0/4$ =2.89 mm from the slot antenna there is a reflector plate whose dimension are the same as the antenna's dimensions (λ_0 is wavelength in vacuum at the centre frequency f_c=25.875 GHz).
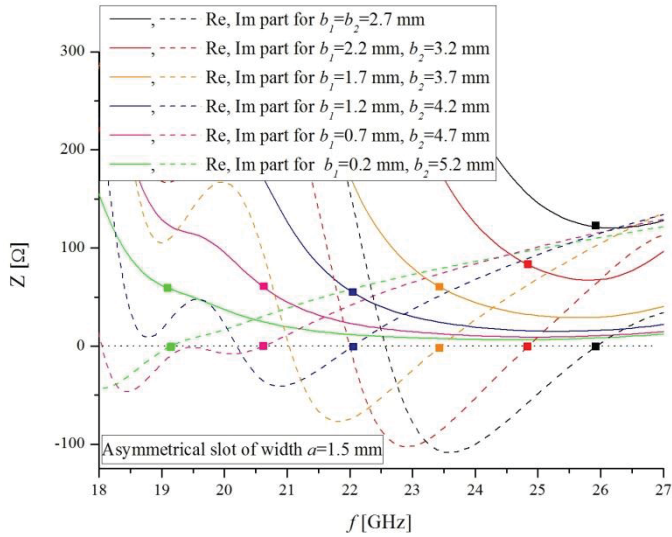
Fig. 2. De-embedded impedance of the CPW-fed asymmetrical slot antenna versus frequency. The slot width, a and its overall length, $b=b_1+b_2$ are fixed.

TABLE I:

SECOND RESONANCE AND IMPEDANCE OF CPW-FED ASYMMETRICAL SLOT ANTENNA VERSUS DIFFERENCE Δb BETWEEN ITS ARMS' LENGTHS.

$\Delta b=b_2-b_1$[mm]	2nd resonance[GHz]	Impedance [Ω]
0	25.9	120
1	24.85	80
2	23.45	60
3	22.05	55
4	20.6	60
5	19.15	55

To estimate the performance of the proposed CPW-fed slot antenna, it is simulated using the WIPL-D software [8]. Its impedance is determined by its de-embedded S_{11} and S_{21} parameters using formulas [9]:

$$Z = -\frac{1}{Y_{21}(S_{11}, S_{21})} \quad (1)$$

$$Y_{21}(S_{11}, S_{21}) = -\frac{2S_{21}}{(1+S_{11})^2 - S_{21}^2} \frac{1}{Z_0} \quad (2)$$

where Z_0=120 Ω is impedance of CPW feed line. The simulation results of impedance for asymmetrical slot with width a=1.5 mm is presented in Fig. 2. It can be observed that with constant slot length only by changing the arms' length it can cover entire K-band within impedance range 55-120 Ω. The asymmetrical slot, whose total length and arm's lengths are both changeable, offers a bigger opportunity to fit impedance at desired frequency than rectangular or bow-tie slot whose length can be only adjusted [10,11]. The numerical values from Fig. 2 are presented in Table I together with the second resonance and slot impedance at second resonance for all considered difference Δb between slot arms' length b_2 and b_1.

a)

b)
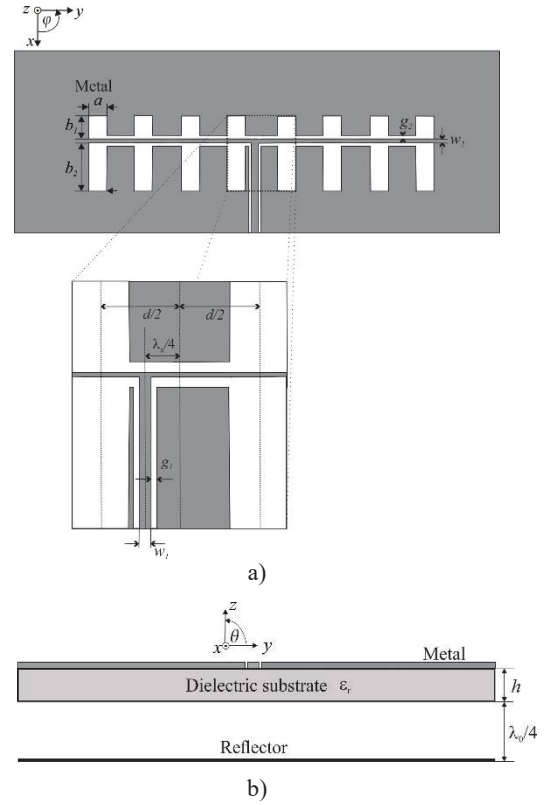
Fig. 3. CPW-fed asymmetrical slot array: a) Top view b) Side view.

III. THE ARRAY OF CPW-FED ASYMMETRICAL SLOTS

The configuration of the proposed CPW-fed asymmetrical slot array is depicted in Fig. 3. A dielectric substrate of constant ε_r=2.54 and thickness h=0.508 mm is used. The centre frequency f_c=25.875 GHz is calculated as the central value of range 24.25 – 27.5 GHz, recommended by ITU [2]. The antenna array is at the distance $\lambda_0/4$ =2.89 mm from the reflector plate whose dimension are the same as the array's dimensions. Unlike the microstrip antennas with a backside ground plane, slot antennas require the reflector plane to be at a distance equal to the quarter of the free space wave-length. It should ensure that the antenna radiates only in half the space. The array is split into two identical four-slot subarrays placed at the different distances from the CPW T-junction. In order to provide the in-phase feeding of both subarrays, the CPW T-junction is moved off the symmetry axes (see Fig. 3) for $\lambda_g/4$, where λ_g=9 mm is CPW line wavelength at the centre frequency f_c. The difference between lengths of CPW lines between T-junction and sub-arrays results in shift between their S-parameters in Smith chart (Fig. 4). The 60 Ω CPW line (featuring the strip w_1=0.9 mm and gap g_1=0.1 mm) is used to enable feeding for both sub - arrays dividing power into two 120 Ω CPW feed lines featuring the strip w_2=0.3 mm and gap g_2=0.375 mm for feeding every four slots sub - array.
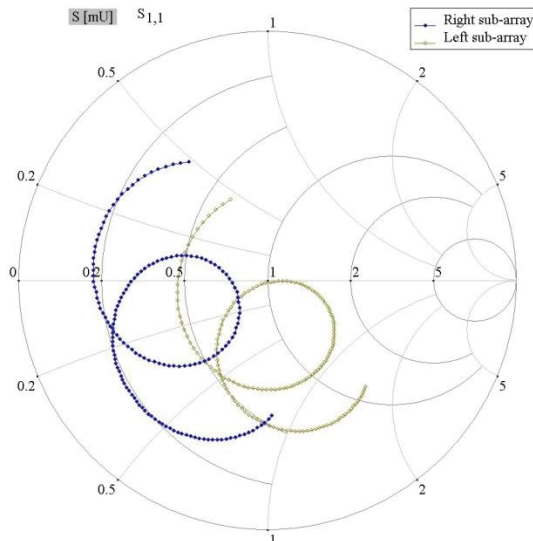
Fig. 4. The S-parameter Smith chart for four-element sub-arrays calculated at the CPW-T junction. S-parameters are normalized to the impedance of CPW feeding line (120 Ω).

The dimensions of metal plate are 77mm x 25mm x 0.508 mm. Each slot has a rectangular shape with a=2 mm width and arms of different length b_1=2 mm and b_2=4.35 mm. Also, the slots are positioned at mutual distance d=9 mm. The antenna array is designed, simulated and analysed using WIPL-D software [8].

IV. SIMULATION RESULTS AND DISCUSSION

The radiation patterns for lower, central and upper frequency are presented in Figs. 5 - 6 resulting in the range from 25.48 GHz to 27.23 GHz. The obtained simulation results point out that arrays with centrally feeding feature more wideband characteristics than the series-fed arrays [12] which is one of the crucial demands for 5G communication systems. The investigated frequency domain has the maximum gain fluctuation 1.6 dB from the gain at the central frequency which is 16.85 dBi. Side lobe suppression varies from 10 dB at the edge frequencies to 13 dB at the central frequency, which is expected for an uniform antenna array. Furthermore, the comparisons between E-plane co-polar and cross-polar radiation patterns as well as H- plane radiation pattern for φ=0° for three considered frequencies are shown in Fig. 5.

Further presented result is S_{11} parameter versus frequency in Fig. 7. It can be observed that S_{11} is below -10 dB for frequencies between 25.475 GHz and 26.65 GHz. Bandwidth which is quoted in terms of return loss, is less than bandwidth determined by radiation pattern.

The Table II presents the review of the overall characteristics of examined array with asymmetrical slot antennas fed by CPW transmission line at three considered frequencies.
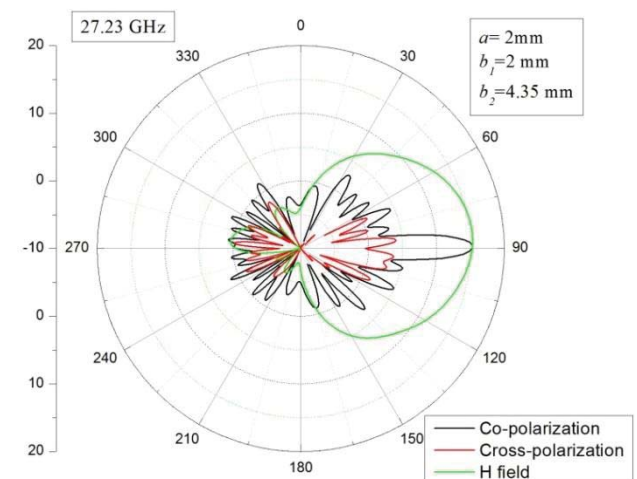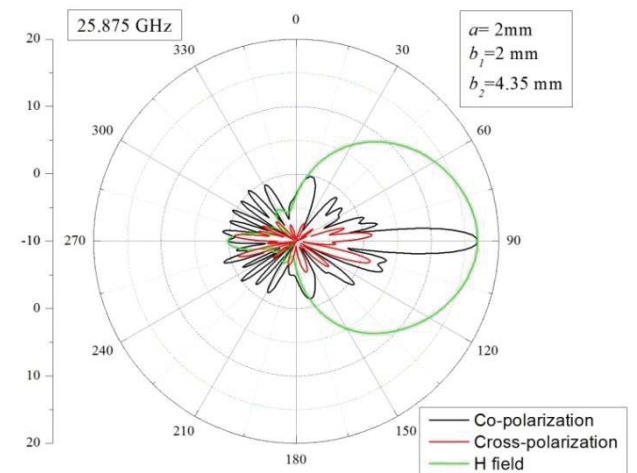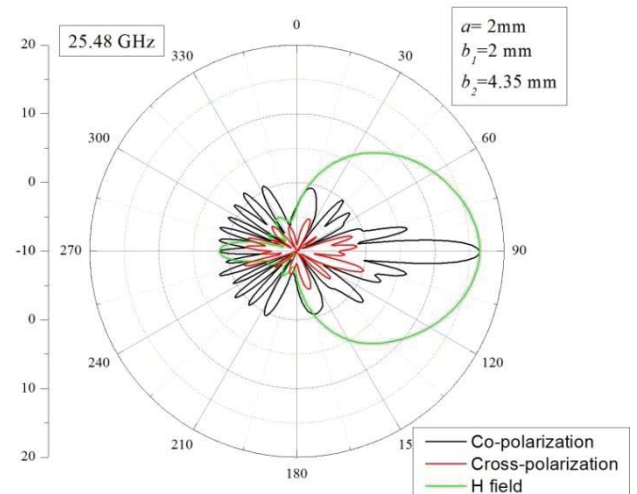
Fig. 5. E-plane co- and cross-polar and H-plane radiation pattern for the CPW-fed asymmetrical slot array at: a) 25.48 GHz b) 25.875 GHz c) 27.23 GHz.
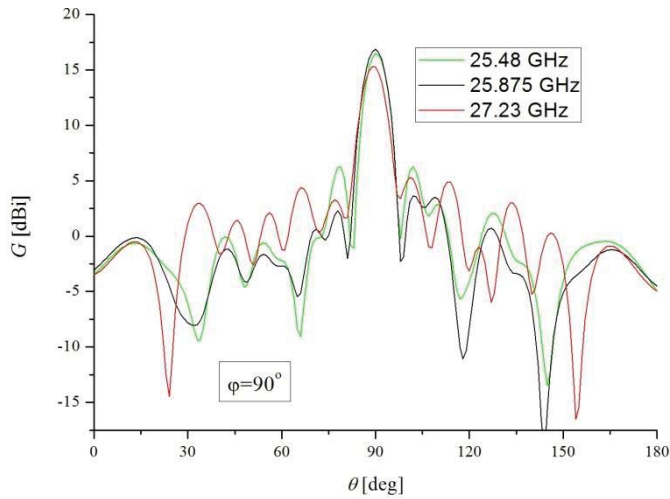
Fig. 6. Radiation pattern in the E-plane for the CPW-fed asymmetrical slot array at 25.48 GHz, 25.875 GHz and 27.23 GHz (φ=90°).
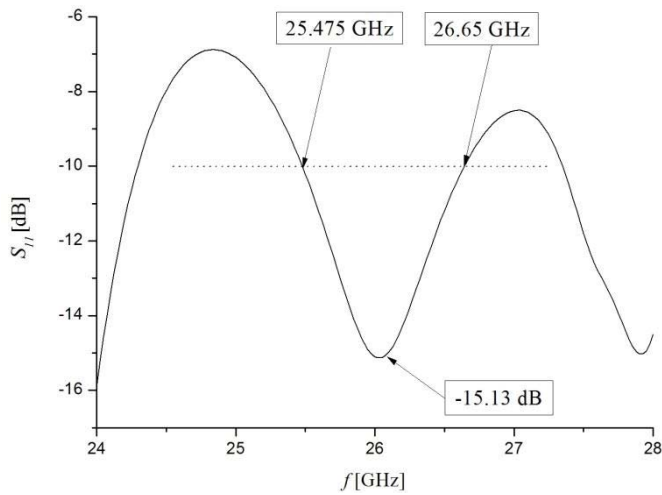
Fig. 7. S_{11} parameter of the CPW-fed asymmetrical slot array versus frequency.

TABLE II:
THE CHARACTERISTICS OF THE ASYMMETRICAL SLOT ANTENNA ARRAY FED BY CPW TRANSMISSION LINE.

Frequency [GHz] Characteristics	25.48	25.875	27.23
Gain [dBi]	16.48	16.85	15.25
SLS [dB]	10	13	10
Co/cross-polar ratio [dB]	> 18	> 18	> 13.5
3dB-beamwidth [°]-E plane	6.5	7	7.5
3dB-beamwidth[°]-H plane	51	53	45

V. CONCLUSION

An antenna design is one of the major considerations to realise mm-wave based 5G applications. Modern wireless communication system requires low profile, light weight, high gain, ease of installation, high efficiency and simple in structure antennas. The design of a wideband compact antenna is a challenging task especially at the operating higher frequencies around 28-GHz. The key features of a CPW-fed antenna are low profile, relative ease of construction, low weight, comfortable to planar and non-planar surfaces, low cost, simple and inexpensive manufacturing. These advantages make them popular in many wireless communication applications.

In this paper, a wideband compact array of CPW-fed asymmetrical rectangular slots is proposed. The centrally feeding provides a wide bandwidth from 25.48 to 27.23 GHz. Furthermore, the simulation results show excellent characteristics of the proposed antenna array, concerning average gain of 16.2 dBi and very symmetrical radiation pattern.

Further research will be conducted by introducing a rat-race CPW coupler and its comparison with CPW T-junction, presented in this paper. The antenna with better simulated results is intended to be fabricated and measured.

ACKNOWLEDGEMENT

REFERENCES

[1] 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.

[2] 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.

[3] M. J. Marcus, 5G and "IMT for 2020 and beyond," *IEEE Wireless Commun.*, vol. 22, no. 4, pp. 2-3, Aug. 2015.

[4] Q. Zhao, J. Li, "Rain attenuation in millimeter wave ranges", 7thInt. Symp. Antennas Propag. & EM Theory, pp. 1-4, Oct. 2006.

[5] 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.

[6] W. Tu, "Analysis and Design of Coplanar Waveguide-Fed Capacitively Coupled Slot Antennas", 2015 International Workshop on Antenna Technology, Seoul, Republic of Korea, 4-6 March 2015.

[7] M. Yang, X. Yin, H. Zhao, "Wideband Coplanar Waveguide-Fed Slot Antenna Array with Via-Wall Structure", Proc. of 2016 10th European Conference on Antennas and Propagation, Davos, Switzerland, 10-15 April 2016.

[8] WIPL-D Pro, WIPL-D Team

[9] David M. Pozar, "Microwave Engineering", John Wiley & Sons, Inc., 2005.

[10] M. Milijic, B. Jokanovic, "K-Band CPW-Fed Rectangular Slot Dipoles for 5G Applications", 5th International Conference IcETRAN 2018, pp. 982-986, Palić, 11-14.06.2018.

[11] M. Milijic, B. Jokanovic, "Parametric Analysis of Wideband CPW-fed Bow-Tie Slot Dipole", 53rd International Scientific Conference ICEST 2018, pp. 93-96, Sozopol, Bulgaria, June 28 - 30, 2018.

[12] M. Milijic, B. Jokanovic, "Radiation bandwidth of series-fed slot arrays for 5G and IoT applications", 2018 26th Telecommunications Forum (TELFOR), pp. 462-465, Belgrade, Serbia, November, 20-21, 2018.

Channel selection in 2.4 GHz ISM band for IEEE 802.15.4 networks

Slađana Đurašević[1] and Uroš Pešović[2]

Abstract – **IEEE 802.15.4 standard represents one of the most used communication technologies for the Internet of Things. This standard shares 2.4 GHz band with WLAN and Bluetooth devices, which number has been drastically increased in the past decade. This paper presents analysis of usage of IEEE 802.15.4 channels in ISM band.**

Keywords –**IEEE 802.15.4, Channel selection, ISM band, Coexistence**

I. INTRODUCTION

License-free bands represent part of the radio spectrum which can be freely used by anyone, where the only constraint is limited transmitting power. With the rapid increase in the number of wireless devices which used license-free bands, these bands almost reached the limits of their capacity in highly urbanized environments [1]. This is especially the case for the 2.4 GHz ISM band which is used by IEEE 802.15.1 Bluetooth devices, IEEE 802.11 Wireless LANs, IEEE 802.15.4 WPAN devices and cordless phones. IEEE 802.15.4 standard represents one of the key wireless transmission technologies for implementation of Internet of Things (IoT). This technology, targeted for smart devices in homes, can be significantly affected by other types of networks in ISM band. Term coexistence, represents the ability of devices to operate under the presence of device which uses different wireless standards in the same frequency band [2]. Coexistence of these wireless standards in the ISM band can be achieved using proactive and reactive techniques. Proactive techniques require cooperative channel sharing and complete control of deployed networks which is usually not the case. Also, most radio standards are not designed to detect other network transmissions and cooperatively share channels. Reactive approaches are typically focused on techniques which are used by the device itself to be able to coexist without need to influence interfering devices.

II. IEEE 802.15.4 STANDARD

IEEE 802.15.4 standard is designed for Low Power Wireless Personal Area Networks (LP-WPAN) and defines

Physical (PHY) Layer and Medium Access Control (MAC) Layer. The PHY layer is responsible for wireless transmission and reception of packets and control of radio transceiver, while the MAC layer provides fair and efficient access to the wireless channel. The physical layer of the IEEE 802.15.4 standard defines several radio bands, where the 2.4 GHz ISM band is the most commonly used worldwide. This band is divided into sixteen IEEE 802.15.4 channels, which are separated by 5 MHz and each channel has a bandwidth of 2 MHz and data rate of 250 kbps. The maximum transmitting power is 0 dBm (1mW), which is much lower compared to maximum transmitting power of 100 mW used for IEEE 802.11 networks. In order to improve coexistence IEEE 802.15.4 devices use Direct Sequence Spread Spectrum DSSS which spreads every bit with eight chips which significantly improve immunity on bit errors. As we can observe from Fig.1, IEEE802.15.4 achieves the best BER (Bit Error Rate) performance compared to the other standards which operate in 2.4 GHz band.
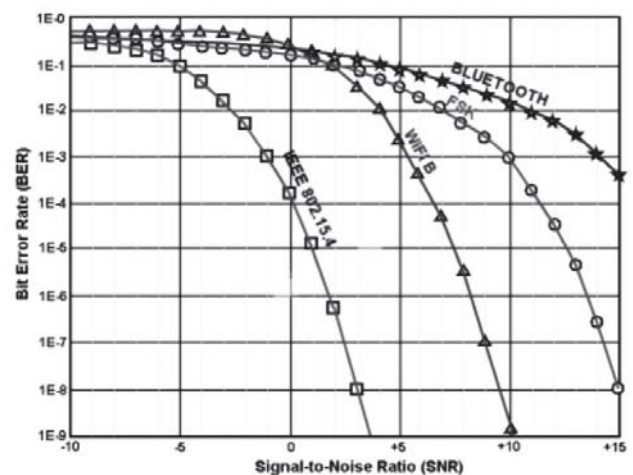
Fig. 1. BER for wireless standards in 2.4 GHz ISM band [3]

MAC Layer uses CSMA/CA (Carrier Sense Medium Access with Collision Avoidance) to achieve efficient medium access. This technique requires the device's PHY layer to first check the channel state using Clear Channel Assignment (CCA) before starting transmission of a packet. There are three methods of performing CCA: either by detecting IEEE 802.15.4 carrier signal (Carrier Sense – CS) or by performing Energy Detection (ED) in the channel, or combination of both. CCA reports a busy medium upon detecting a signal with the modulation and spreading

[1]Slađana Đurašević is with the Faculty of Technical Sciences Čačak, University of Kragujevac, Svetog Save 65, Čačak 32000, Serbia, E-mail: sladjana.djurasevic@gmail.com
[2]Uroš Pešović is with the Faculty of Technical Sciences Čačak, University of Kragujevac, Svetog Save 65, Čačak 32000, Serbia, E-mail: uros.pesovic@ftn.kg.ac.rs

characteristics of IEEE 802.15.4 or upon detecting energy level within the bandwidth of an IEEE 802.15.4 channel which is above the Energy Detection (ED) threshold. Carrier sense CCA is intended to prevent collision with IEEE 802.15.4 devices, while ED can be used to prevent collisions with another type of networks which share the same frequency. ED is also used by MAC during the initial formation of the network to estimate the state of all sixteen channels and choose the channel with the lowest energy level.

MAC layer can operate in one of two operating modes: Non-Beacon Enabled and Beacon Enabled mode. Non-Beacon Enabled mode is event-driven medium access mechanism, where device attempts medium access as soon as the transmission is requested. Prior transmission, device performs blind backoff CSMA/CA, where it waits for a random period of time. When this period expires, MAC requests CCA service to ensure that the channel is free in order to start transmission. If CCA reports a busy channel, the device shall wait for another random period before trying to access the channel again. If the number of unsuccessful random backoff exceeds the maximum permissible number of retries, the packet transmission is aborted.

Beacon Enabled mode represents scheduled medium access mechanism in which device can initiate channel access during the appropriate time slot intended for transmission. Using beacon frame, the PAN coordinator announces current channel access timetable, shown in Fig.2.
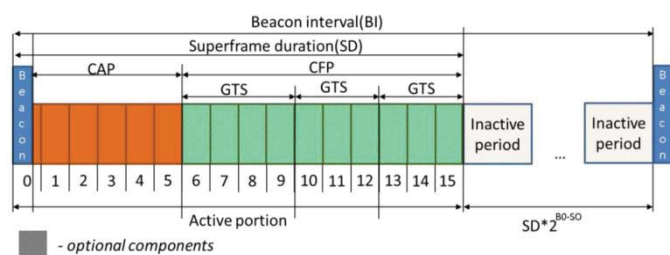
Fig. 2. Superframe structure [4]

Active communication period, called superframe, is divided into 16 slots, which are grouped into Contention Access Period (CAP) and Contention Free Period (CFP). During CAP devices are contenting for medium access using slotted CSMA/CA mechanism. The device, which wants to start transmission, locates a boundary of the time slot, waits for a random period, after which it performs CCA. Backoff time is used to prevent devices from simultaneous channel access, which can lead to a collision of transmitted packets. CFP provides up to seven Guaranteed Time Slots (GTS) troughs which device can transmit time sensitive traffic. GTS is pre-allocated to the specific devices, after device issues GTS reservation request from PAN coordinator. If the device does not use its GTS for the extended period, PAN Coordinator will assign these GTS slots on another device request. Superframe is followed by an optional inactive period in which devices go to low-power sleep mode until the arrival of the next beacon frame.

III. COEXISTENCE TECHNIQUES

Proactive techniques tend to prevent interference using physical separation of networks, frequency separation and time separation. Physical separation relies on careful network planning in order to achieve spatial separation between different networks. Due reduced signal strength over distance, networks will be less influenced by one another which can improve their coexistence. Physical separation doesn't work well in locations with dense wireless networks. It is only practical, if the user has complete control of networks deployment in certain broader area which is usually not the case.
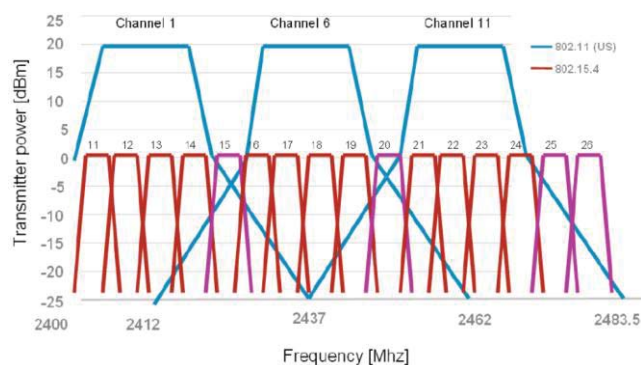
Fig. 3. Channel separation in 2.4 GHz ISM band [5]

Frequency separation reduces interference between two networks by operating on different frequencies. Since ISM band is divided into different channels by different wireless standards, it is necessary to select channels for different types of networks in such manner that their operating frequencies don't overlap. IEEE 802.11 and 802.15.4 devices operate in fixed channels, where transmit power of IEEE 802.11 devices is a hundred times stronger compared to 802.15.4 devices, which can be significantly affected. In order to achieve frequency separation, it's necessary to exploit frequency gaps between non-overlapping IEEE 802.11 channels in which IEEE 802.15.4 devices can operate without interference. These IEEE 802.15.4 channels are 15, 20, 25 and 26 for US and 15, 16, 21 and 22 for Europe as shown by Fig.3.

TABLE I
ISM BAND CHANNELS

Wireless standard	Non-overlapping channels	Channel bandwidth	Channel separation
IEEE 802.11b	4	22	5
IEEE 802.11g	3	20	5
IEEE 802.11n	1	40	5
IEEE 802.15.1	79	1	1
IEEE 802.15.1 BLE	37	2	2
IEEE 802.15.4	16	2	5

IEEE 802.15.1 networks operate in entire bandwidth using the pseudorandom channel-hopping technique. Regular

Bluetooth hops to one of the 79 channels every 625 µs, while newer Bluetooth Low Energy hops to one of 37 channels every 3 ms, with adaptive hopping capability to avoid channels with interference. Due to the unpredictability of channel hopping scheme, frequency separation is not feasible with Bluetooth networks.

Time separation technique enables devices which operate in different standards to send and receive data in different time slots, in order to avoid collisions. This technique exploits low utilization which is common for low power networks, so they can utilize the medium for a small amount of time. In order achieve time separation, it's necessary to implement centralized control of different networks. In case of intense traffic time, separation can increase packet delays and collisions as shown in Fig.4..
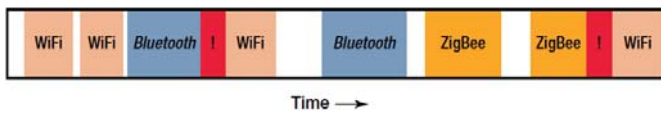
Fig. 4. Time separation in 2.4 GHz ISM band [6]

Proactive techniques require cooperative channel sharing and complete control of deployed networks which is usually not the case. Also, most radio standards aren't designed to detect other network transmissions and cooperatively share channels. Thus, wireless standards need to be upgraded in such a way that they can detect and avoid congested channels. One of the modifications of standard was proposed in 2015. IEEE 802.15.4e implements Time Slotted Channel Hopping. Using this technique all devices operates in one channel during certain time slot, and when that slot expires, all devices switch to another channel where they continue communication. Thus IEEE 802.15.4 devices can avoid using overcrowded channels by adaptive channel hopping as shown in papers [7-11].

IV. EXPERIMENTAL RESULTS

IEEE 802.15.4 channels are observed in 2.4 GHz ISM band with PICDEM Z development board [12]. This board uses MRF24J40MA [13], 2.4 GHz IEEE 802.15.4 compliant transceiver, realized on prefabricated Printed Circuit Board (PCB) with dipole antenna. Board is placed vertically in order achieve omnidirectional radiation pattern of transceiver antenna in the horizontal plane. PICDEM Z is controlled from personal computer using Radio Utility Driver Program [14], which logs ED data from all channels received via RS232 port. Transceiver is put into reception mode in which it performs cyclic ED on each channel. Result of ED is RSSI (Received Signal Strength Indicator) value which is averaged on four received Bytes. When one channel is selected, after expiry of oscillator stabilization time ED is performed 32 times in succession, which is equivalent to duration of transmission of packet of maximum size of 128 Bytes. Maximum detected ED value is then send to console output, after which ED is performed on following channel.

Channel state is observed in three scenarios: public building (faculty), residential building and downtown house during the peak traffic period: noon at public buildings and evening in
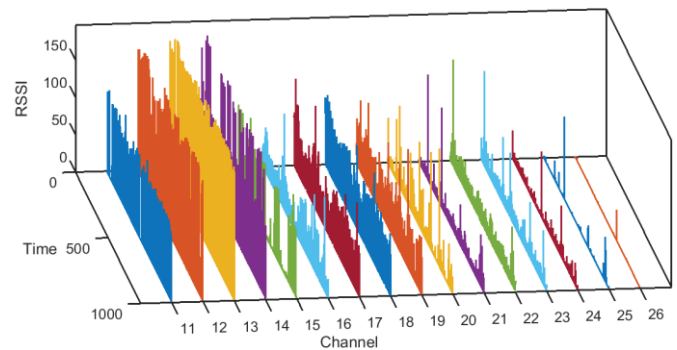
Fig. 5. IEEE 802.15.4 channels in public building

residential buildings. All sixteen channels are monitored for duration of 1000 cycles where each cycle lasts 400ms.

Results for public building, presented in Fig.5, shows one dominant IEEE 802.11 networks which occupies same bandwidth as IEEE 802.15.4 channels 11 to 14. Also we can observe several less influencing IEEE 802.11 networks on other channels. Result shown that, in this case, channel 26 is the least affected by other types of wireless transmission and it can be used for IEEE 802.15.4 operation.

Results for residential building presented in Fig.6, shows several dominant IEEE 802.11 networks which occupies same bandwidth as IEEE 802.15.4 channels 13 to 23. Also we can observe several less influencing IEEE 802.11 networks on other channels. This is the results of number large number of WLANs which are typically used as home LAN infrastructure in apartments in Serbia. Result shown that, in this case, all channels are occupied and it will be challenging task to implement error-free operation of IEEE 802.15.4 network.
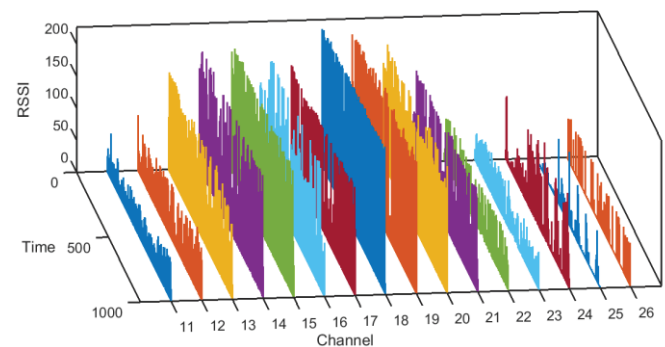
Fig. 6. IEEE 802.15.4 channels in residential building

Results for downtown house presented in Fig.7, shows several IEEE 802.11 networks which occupies same bandwidth as IEEE 802.15.4 channels 22 to 23 and 12 to 15. Interference of these networks is much smaller compared to other two scenarios due physical separation between houses which is dozen meters. In this scenario IEEE 802.15.4 networks can operate without interference on most channels.
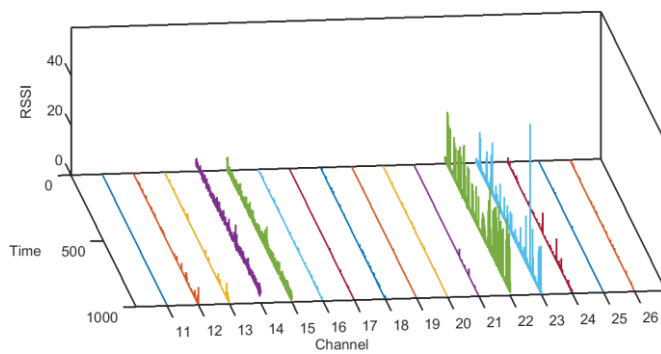
Fig. 7. IEEE 802.15.4 channels in downtown house

V. Conclusion

Results show that IEEE 802.15.4 networks could hardly coexist with other types of networks in highly urbanized environments, such as public and residential buildings. In such scenarios it's necessary to use advanced mechanisms with adaptive channel hopping.

Acknowledgement

References

[1] V. Valenta, R. Marsalek, G. Baudoin, M. Villegas, M. Suarez, "Survey on Spectrum Utilization in Europe: Measurements, Analyses and Observations. 5th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications, Jun 2010, Cannes, France. pp. ISBN: 978-963-9799-94-3.

[2] "Co-existence of IEEE 802.15.4 at 2.4 GHz", NXP Application Note JN-AN-1079, Revision 1.1, 8-Nov-2013

[3] J.T. Adams, "An introduction to IEEE STD 802.15.4", IEEE Aerospace Conference, 4-11 March 2006,Big Sky, MT, USA, ISSN: 1095-323X

[4] K. Mikhaylov, N. Plevritakis, J. Tervonen, "Performance Analysis and Comparison of Bluetooth Low Energy with IEEE 802.15.4 and SimpliciTI", Journal of Sensor and Actuator Networks 2(3) pp.589-613, September 2013

[5] White Paper: "Integrating an Industrial Wireless Sensor Network with Your Plant's Switched Ethernet and IP Network", Cisco Systems, Inc. 2009

[6] White paper: "How to Ensure IoT Devices Work in Their Intended Environment - Locate And Identify Interference", Keysight Technologies, 2019

[7] S. Grimaldi, A. Mahmood, M. Gidlund, "Real-Time Interference Identification via Supervised Learning: Embedding Coexistence Awareness in IoT Devices", IEEE Access (Volume: 7) 2018, pp: 835 - 850, ISSN: 2169-3536

[8] H. Dakdouk, E. Tarazona, R. Alami, R. Feraud, G. Z. Papadopoulos, P. Maille, "Reinforcement Learning Techniques for Optimized Channel Hopping in IEEE 802.15.4-TSCH Networks", MSWIM '18 Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 99-107, Montreal, Canada, 28.10 - 02.11, 2018

[9] S. Hammoudi, S. Harous, Z. Aliouat and L. Louail, "Time slotted channel hopping with collision avoidance", Int. Journal Ad Hoc and Ubiquitous Computing, Vol. 29, Nos. 1/2, 2018 85

[10] A. Elsts, X. Fafoutis, R. Piechocki, and I. Craddock, Adaptive Channel Selection in IEEE 802.15.4 TSCH Networks, 2017 Global Internet of Things Summit (GIoTS), 6-9 June 2017 Geneva, Switzerland, ISBN: 978-1-5090-5873-0

[11] N. H. Mahalin, H. S. Sharifah , S.K. Syed Yusof, N. Fisal, R.A Rashid, "RSSI Measurements for Enabling IEEE 802.15.4 Coexistence with IEEE 802.11b/g", TENCON 2009 - 2009 IEEE Region 10 Conference, 23-26 Jan. 2009, Singapore, ISSN: 2159-3450

[12] PICDEM Z Demonstration Kit User's Guide, Microchip Technology Incorporated,2008,

[13] Microchip, "MRF24J40MA Datasheet, 2.4 GHz IEEE Std. 802.15.4™RF Transceiver Module", Microchip Technology Incorporated,2008,

[14] MRF24J40 Radio Utility Driver Program, Application Note AN1192, Microchip Technology Incorporated,2009,

INVESTIGATION OF LTE/LTE-R FUNCTIONALITIES IN TRAIN RADIO

Adi Koruni[1] and Ivaylo Topalov[2]

Abstract – **High-speed railways (HSRs) improve the quality of rail services, yield and help to create socioeconomically balanced societies. To handle increasing traffic, ensure passenger safety, and provide real-time multimedia information, a new communication system for HSR as well as for conventional rail (CR) is required. In the last decade, public networks have been evolving from Global System for Mobile Communications (GSM) with limited capabilities, to third (3G) and fourth-generation (4G) broad-band systems that offer higher data rates e.g., long-term evolution (LTE). This requires development for the railways of upgraded or new train dispatchers and drivers terminals with LTE/LTE-R functionalities. In this paper, the necessary functionalities in broadband wireless access for train radio have been discussed.**

Keywords – **LTE, LTE-R, Dual radio, GSM-R, Functionality, Cab Radio**

I. INTRODUCTION

For high-speed railway (HSR) and conventional rail (CR) to evolve the current Global System for Mobile Communications (GSM) - railway (GSM-R) technology with the next-generation railway-dedicated communication system providing improved capacity and capability is now in EU an upcoming task. [1] A reliable broadband communications system is essential for different HSR and CR components, such as train control and safety-related communications. Since 2014, a project of the International Union of Railways (UIC), known as the Future Railway Mobile Communication System (FRMCS), has started to assess and shape the future of HSR mobile communications and to identify suitable candidate technologies to use once the currently used GSM-R has become obsolete. HSR applications have strict requirements for quality-of-service (QoS) measures, such as data rate, transmission delay, and bit error rate (BER). Due to these factors as well as a desire to use mature and low-cost technology, HSR communications generally use off-the-shelf technologies and add applications to meet specific services and demands.

GSM-R [2] is a successful example, based on the GSM standard and used on over 70,000 km of railway lines (including over 22,000 km of HSR lines) all over the world [3]. The GSM communications systems are being decommissioned as the public communication market is

[1]**Adi Koruni is with the "Todor Kableshkov University of Transport of Sofia, 158 Geo Milev Str., Sofia 1574, Bulgaria, E-mail: office@vtu.bg**

[2]**Ivaylo Topalov is with the "Todor Kableshkov University of Transport of Sofia, 158 Geo Milev Str., Sofia 1574, Bulgaria, E-mail: office@vtu.bg**

evolving toward the Third Generation Partnership Project (3GPP) long-term evolution (LTE) [4]. A new system is thus required to fulfill HSR operational needs, with the capability of being consistent with LTE, offering new services but still coexisting with GSM-R for a long period. The selection of a suitable wireless communication system for HSRs needs to consider such issues as performance, service attributes, frequency band, and industrial support. Compared with third-generation (3G) systems, fourth-generation (4G)f LTE has a simple flat architecture, high data rate, and low latency, making it an acknowledged acceptable bearer for real-time HSR applications. Fifth-generation (5G) systems, although currently discussed in 3GPP, will be available only after 2020 and, therefore, are not suitable for the HSR time frame [5]. In view of the performance and level of maturity of LTE, LTE - railway (LTE-R) will likely be the next generation of HSR communication systems [6], [7], and the future vision for HSR and CR wireless technologies will thus rely on it.

II. GSM-R SYSTEM

GSM-R is essentially the same system as the GSM but with railway-specific functionalities. It uses a specific frequency band around 800/900 MHz, as illustrated [8].

In addition, the frequency bands 873 – 876 MHz (uplink) and 918–921 MHz (downlink) are used as extension bands for GSM-R on a national basis, under the name Extended GSM-R (E-GSM-R). GSM-R is typically implemented using dedicated base stations (BSS) close to the rail track. The distance between two neighboring BSS is 7–15 km but in China it is 3–5 km because redundancy coverage is used to ensure higher availability and reliability. GSM-R has to fulfill tight availability and performance requirements of the HSR radio services.

III. GSM-R SERVICES

The GSM-R network serves as a data carrier for the European Train Control System (ETCS), which is the signaling system used for railway control. The ETCS has three levels of operation and uses the GSM-R radio network to send and receive information from trains. On the first level, ETCS-1, the GSM-R is used only for voice communications. On the other two levels, ETCS-2 and ETCS-3, the GSM-R system is used mainly for data transmissions. The GSM-R is very relevant to ETCS-2 and ETCS-3, where the train travels at a speed up to 350 km/h, and it is thus necessary to guarantee a continuous supervision of train position and speed. When the call is lost, the train has to automatically reduce the speed to 300 km/h (ETCS-1) or lower [9].

In Fig.1 [10] is summarized the future possible services provided by LTE-R, which is based on UIC technical reports of China Railway and ERA. It is noteworthy that broad-band wireless access for passengers inside high-speed trains is not provided by LTE-R because of its limited bandwidth.
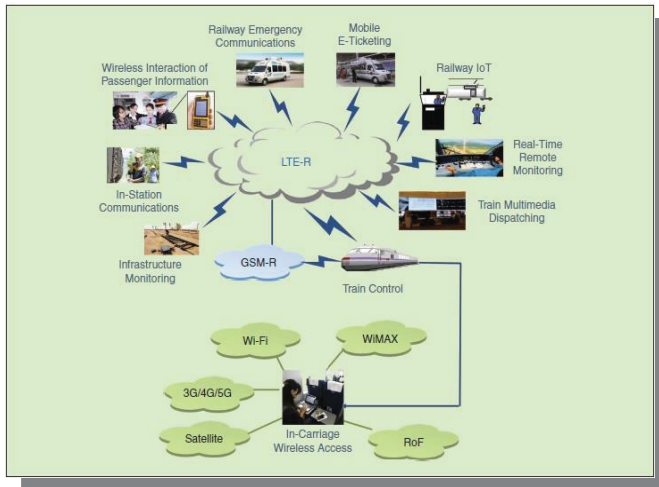
Fig.1 The LTE-R services

IV. LTE-R CHALENGES

> **There are several challenges associated with LTE/R.**

1) HSR-specific scenarios: In the LTE standard of [11], a channel model for HSR is resented that only includes two scenarios, open space and tunnel, and uses a nonfading channel model in both scenarios.

2) High mobility: High-speed trains usually run at a speed of 350 km/h, and LTE-R is designed to support 500 km/h. The high velocity leads to a series of problems. First, high velocity results in a non-stationary channel because, in a short time segment, the train travels over a large region, where the field strength change significantly.

3) Delay spread: Delay dispersion leads to a loss of orthogonality between the OFDM sub-carriers, and a special type of guard interval, called the cyclic prefix (CP), should be employed. The delay dispersion determines the required length of LTE CP supports both short (4.76 ms) and long (16.67 ms) CP schemes.

4) Linear coverage: In HSRs, linear coverage with directional antennas along the rail track is used, where the directional BS antennas orientate their main lobe along the rail track so that it is power efficient. The linear coverage brings some benefits, e.g. with the known location of a train, it is possible to design distance/ time-based beam forming algorithms with good performance.

V. THE FUTURE CAB RADIO INCLUDED IN LTE-R

The Future Cab Radio generations will be based on a packet oriented data transmission system, using or derived from current available commercial networks. Based on its long years of experience, in Funkwerk, they has been developed a LTE based Cab Radio system providing 2G/3G/4G telephony and data services as well as VoLTE1 and IP-based data-services (Fig.2)[12]. It operates on an Android based system. Due to the implementation of this market leading operating system the sage of this device is nearly self-explanatory. The dispatching radio is designed for use in Rail vehicles. In the vehicle there are two directional antennas supporting the MIMO technology which is applied in LTE for speed increases of download and upload data stream. The equipment was tested in Huawei's lab successfully in 2014. On this base a dual-mode GSM-R/LTE Cab Radio is based on Funkwerk's well-known GSM-R Cab Radio family "MESA".
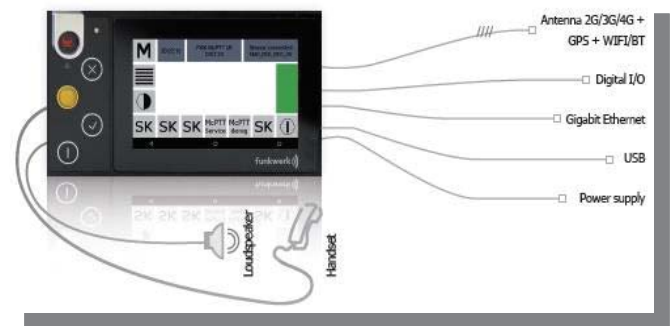
Fig.2 LTR Cab radio

It can be operated using public available SIM cards and is an all-in-one unit for fixed installation in vehicles. Besides the interfaces for 2G/3G/4G network "access further interfaces like Gigabit Ethernet, WIFI, Bluetooth", as well as a handset, a loudspeaker and generic I/O connection, are available. The installation of 3rd party applications is possible and therefore it can be smoothly integrated into an existing or new infrastructure of systems supporting Funkwerk McPTT (Mission critical Push to Talk) application. Customer specific applications or adjustments can be provided by Funkwerk or by 3rd parties too. The system is designed for the support of upcoming LTE releases and will be adapted to future services like group communication services.

VI. LTE FEATURES FOR SUPPORTING REQUIRED RAILWAY FUNCTIONALITIES

Railway services demands specific functionalities to train radio systems. For instance, GSM standard was enhanced with the Advanced Speech Call Items (ASCI) functionalities. Proposed LTE features and mechanisms to implement the railway functionalities are shown in Table 1.

TABLE I

PROPOSED LTE FEATURES TO SUPPORT GSM-R
RAILWAY FUNCTIONALITIES

GSM-R Functionality	LTE Feature
Voice Group Call Service (VGCS)	**LTE** IMS based **VoIP (VoLTE)** + IMS based **Push to talk Over Cellular (PoC)** + **Enhanced Multimedia Broadcast Multicast Service (eMBMS**
Voice Broadcast Calls (VBS)	**VoLTE** + **PoC** and/or **eMBMS**: IP multicast of voice and video services
Priority and Pre-emption (eMLPP)	**Access Class Barring** mechanisms + **Policy Control Rules** + **QoS** mechanisms **(ARP)**.
Functional Addressing (FN)	**Session Initiation Protocol** (SIP) **Addressing**
Location Depending Addressing (LDA, eLDA)	**Localization Services** in LTE (Release 10)
Railway Emergency Calls (REC, e-REC)	**Emergency** and critical safety **voice services over IMS in LTE.**
Fast Calls Set-up	IMS based **PoC** + **Access Class Barring**
Data Exchange (SMS, Shunting)	IMS based **SMS Service**

The IP Multimedia Subsystem (IMS) is enhanced-services architecture for delivering any service, reaching any customer regardless of how they connect to the network [13]. The main drawback is the fact that it is far from ready for deployment. The Session Initiation Protocol (SIP) is a protocol that facilitates the formation, modification and execution of communication sessions between individual or multiple participants. Locating call recipients and talking with them on their different user agents is accomplished using a SIP address [14]. The Push to Talk Over Celular (PoC) deployed into an IMS Service Delivery Platform can be an efficient way to provide the VGCS functionality in

LTE standard while fulfilling railway services delay requirements. However, Voice Group Call Services (VGCS) is radio cell based in contradiction to PoC functionalities (Mandoc [15]).

VII. TOP CHALLENGES OF LTE

In this section, main technical challenges of LTE for supporting railway functionalities with their specific QoS requirements are described:

➢ **Voice service provision over IP LTE networks**

GSM-R functionality for delivering voice services is considered a key core functionality for railway operation. The voice service provision in the LTE standard is a major challenge that must be carefully assessed and analyzed. Regarding the technological solutions for delivering voice over LTE, the solutions based on using the IP Multimedia Subsystem (IMS), the hybrid voice over LTE via Generic Access (VoLGA) solution and the CSFB solution are the most promising ones [16].

➢ **Handover in LTE**

LTE standard support hard handover mechanisms, which reduces the complexity of the LTE network architecture. However, the hard-handover (HHO) mechanism does not guarantee any data packet losing in handover process. LTE HHO must fulfill the railway service QoS and RAMS requirements, especially in high speed scenarios. It is necessary that the HHO mechanisms supported in LTE, lossless and seamless, minimize the packet loss or avoid it completely with fast connection and re-association time.

➢ **Quality of service mechanisms and access control in LTE networks**

Service prioritization, the ability to preempt users, and Quality of Service are all crucial to a future railway communication system. In railway environments, LTE must assure the delivery of data and voice packets while meeting a combination of delay, jitter, dropped call rate or data error rate, handover interruption time, maximum call setup time and maximum/guaranteed bit rate requirements.

➢ **Network performance in high speed environments**

The impact of high speed in LTE performance and capacity should be assessed carefully. The most important issues that should be evaluated are:

− *The Doppler shift effect in LTE downlink and uplink channel performance.*
− *The effect of high speed in resource scheduling.*
− *Handover mechanisms in LTE.*

➢ **Service RAMS requirements**

The described in [17] RAMS requirements Availability (interval availability), Regular Maintenance, Down Time, Corrective Maintenance, Preventive Maintenance, Maintenance Window, Mean Time to Restore (MTTR), Preventive Maintenance, RAM Program, Service Failure, System Lifecycle and Up Time has to be defined before implementation of LTE-R cap radio into HSR networks.

VIII. DUAL-MODE GSM-R/LTE CAB RADIO

The dual mode cab radio (Fig.3) is not limited to provide video / voice communication over LTE, it also supports voice communication between GSM-R network and LTE network. Since the bandwidth demands are increasing in the transport fields the dual mode cab radio will be one better choice for all Railway and Public Transport operators in the future.

➤ **APPLICATION SCENARIOS:**

- Real-Time Video communication over LTE between train driver and the dispatcher terminals, and maintenance staffs with handhelds.
- Voice communication over LTE between train driver and the dispatcher terminals, and maintenance staffs with handhelds.
- Voice communication between GSM-R network and LTE network
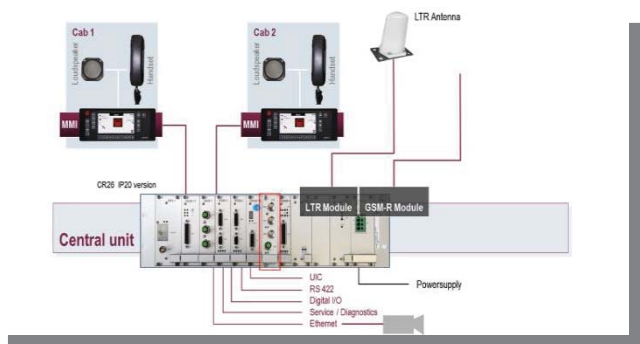
Fig.3 Dual-mode GSM-R/LTE Cab Radio Overview

IX. CONCLUSION

In this paper, the top challenges for the LTE system to become the future railway communication system are identified and discussed. These are related to the LTE mechanisms and features to implement the required railway functionalities, LTE technical requirements, the convergence to an all IP network, spectrum harmonization, network deployment considerations and LTE capabilities to meet the service RAMS requirements.

ACKNOWLEDGEMENT

REFERENCES

[1] UIC. (2016). High speed. [Online]. Available: http://www.uic.org/spip.php?mot8

[2] UIC GSM-R Functional Group, "GSM-R functional requirement specification (FRS)," UIC, Paris, France, UIC EIRENE Technology Report, UIC Code 950, version 7.3.0, 2012.

[3] UIC. (2016). GSM-R. [Online]. Available: http://www.uic.org/spip.php?rubrique851

[4] 3GPP TS 36.201, "Evolved Universal Terrestrial Radio Access (E-UTRA);LTE physical layer; General description," 3GPP, Sophia Antipolis, France, 3GPP Technical Specification, version 12.2.0, Release 12, 2015.

[5] D. Taylor, N. Lofmark, and M. McKavanagh, "Survey on operational communications—study for the evolution of the railway communications system," final report for the European Railway Agency, 2014.

[6] G. Barbu, "Broadband communication with moving trains, technology state of the art," UIC E-Train Technology Report, 2010.

[7] K. D. Masur and D. Mandoc, "LTE/SAE—the future railway mobile radio system: Long-term vision on railway mobile radio technologies," UIC Technology Report, 2009.

[8] Electronic Communications Committee within the European Conference of Postal and Telecommunications Administrations,"Practical mechanism to improve the compatibility between GSM-R and public mobile networks and guidance on practical coordination," Copenhagen, Denmark, EEC Report 162, 2011.

[9] UIC GSM-R Operators Group, "GSM-R system requirements specification (SRS)," UIC, Paris, France, UIC EIRENE Technology Report, UIC Code 951, version 15.3.0, 2012.

[10] Y. Zhou, Z. Pan, J. Hu, J. Shi, and X. Mo, "Broadband wireless communications on high speed trains," in Proc. 2011 20th Annual Wireless and Optical Communications Conf. (WOCC), Newark, NJ, pp. 1–6.

[11] 3GPP TS 36.104, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception," 3GPP, Sophia Antipolis, France, 3GPP Technical Specification, version 11.8.2, Release 11, 2014

[12] http://www.funkwerk.com/wp-content/uploads/2017/05/CRLT-20_EN_V25.pdf

[13] Badard, B., Diascorn, V. and Boulmier, G. et al, "Migration to VoIP over mobile networks: Technical challenges and economic opportunity analysis," Telecommunications Network Strategy and Planning Symposium (NETWORKS), 2010 14th International, vol., no., pp. 1-7, 27-30 Sept. 2010.

[14] Chen Wu; "User ID provisioning for SIP registration in IMS," Education Technology and Computer (ICETC), vol.2, no., pp. V2-206-V2-210, 22-24 June 2010.

[15] Mandoc, D., "LTE/SAE – The Future Railway Mobile Radio System: Long-Term Vision on Railway Mobile Radio Technologies". Technical Report. UIC, November 2009.

[16] Paisal, V., "Seamless voice over LTE," Internet Multimedia Services Architecture and Application (IMSAA), 2010 IEEE 4th International Conference on, vol., no., pp. 1-5, 15-17 Dec. 2010.

[17] EuropeRailStandard.pdfaintainability.pdf

Enabling Adaptivity in IoT-based Smart Grid Architecture

Nenad Petrović and Đorđe Kocić[1]

*Abstract –***The increasing usage of electric power in recent years has led to evolution of the existing electric grid infrastructure towards Smart Grid architecture. In this paper, we explore how state-of-art information and communication technologies can be combined to enable adaptivity within the Smart Grid relying on affordable IoT devices. As outcome, we propose an architecture and present some implementation and evaluation aspects.**

Keywords – **IoT, Smart Grid, Edge Computing, semantic technology, data analysis.**

I. INTRODUCTION

In the previous century, the usage of electrical power has been one of the main key-enablers of rapid technological progress [1]. However, the demand for electrical energy of today's consumers is becoming much higher, while, on the other side, the availability of non-renewable resources is limited, pushing the traditional energy distribution systems to their limits [1-4]. In such conditions, the quality of the transferred power is dramatically affected that could lead to serious problems and even catastrophic results. Moreover, a constant pressure for switching to renewable, sustainable and cheaper energy resources exists. Therefore, there is a need for evolution of the existing energy distribution systems and increase of their flexibility while making them adaptable.

In recent years, a lot of effort is being put in process of transformation of the existing energy distribution systems, relying on state-of-the-art information and communication technologies, towards the so-called *Smart Grid* infrastructure. Smart Grid is defined as a next generation power grid, implemented as a two-way cyber-physical system with embedded computational intelligence, leveraging the collected information in order to provide clean, safe, secure, reliable, resilient, efficient, economic and sustainable electrical energy to end-users [1-5]. One of its main characteristics is the ability to detect the events that occur anywhere in the grid and react by adopting the corresponding strategy in order to respond the changing demands or recover itself in case of anomalies, in near real-time.

In this paper, it is examined how the synergy of cutting-edge information and communication technologies and paradigms can enable adaptivity within Smart Grid relying on Internet of Things (IoT) devices. As an outcome, we propose an architecture leveraging the mentioned concepts and present proof-of-concept implementation with some evaluation aspects.

II. BACKGROUND AND RELATED WORK

[1]Nenad Petrović and Đorđe Kocić are with the Faculty of Electronic Engineering at University of Nis, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: nenad.petrovic@elfak.ni.ac.rs

Architecture model: In literature, many descriptions of Smart Grid infrastructure components and architecture model exist [3-5]. However, some common elements are identified and we summarize them as follows in Table I.

TABLE I
SMART GRID COMPONENTS AND THEIR ROLES

Component	Role
Energy subsystem	Power generation, transmission and distribution
Communication layer	Usage of wired and wireless communication technology to enable information exchange between components
Metering devices	Devices recording electrical and non-electrical measurement values
Computational intelligence	Knowledge extraction from the collected data and decisioning
Applications and services	Various software used by operators or consumers providing visualization, monitoring and/or control

Internet of Things (IoT): IT refers to a system of interconnected devices used in everyday life, residing in our environment with goal to perform the automation of a particular domain – from healthcare and home appliances to military systems. These devices are equipped with different kinds of sensors and modules enabling them to collect certain type of information. Moreover, they can be equipped with actuators in order to be able to affect the environment as a response. In most cases, their processing power is quite limited and they often need to communicate with other devices (or servers) in order to achieve their goal. For communication, a variety of technologies is used, both short- (such as Bluetooth) and long-range (Wi-Fi, 4G). It is identified that IoT has great potential in Smart Grid applications, as measurement and actuation devices have to be distributed throughout residential and industrial objects in order to collect the necessary data and provide the response to the events that have occurred. IoT devices perfectly fit that purpose, considering their small size, affordability and connectivity [2].

Data analysis: In Smart Grids, it is necessary to analyze the enormous amount of data acquired by IoT and metering devices to extract knowledge and meaningful patterns in order to detect or predict occurrence of particular events within the environment and react accordingly. For that purpose, various data mining and machine learning techniques are leveraged. In most of the existing work, the focus of their application is on anomaly detection and load forecasting [3]. In these use cases, clustering, classification and regression are widely used, acting on various measurements beside electrical signals – temperature, weather, rainfall and location data [3-7]. Anomaly detection is of particular importance in this context, as it

provides the ability of discovering failures and malfunctions, so Smart Grid can respond to fix them, making them self-healing. In [4], a clustering algorithm together with association rule discovery was applied in order to detect anomalous events, such as overcurrent. On the other side, to optimize the demand scheduling, the accurate energy usage pattern of the consumers is essential. For that reason, the demand forecasting plays an important role. In [6] and [7], regression based on support vector machines was used for load forecasting.

Edge computing: Acquiring the data coming from IoT devices and their sensors is of utmost importance for monitoring and decisioning in Smart Grids. However, in this case, the traditional Cloud computing approach does not give satisfactory results, as offloading the enormous amount of data generated by IoT devices equipped with a variety of sensors to the Cloud for processing would introduce huge latency. On the other side, as Smart Grids grows, the number of connected IoT devices increases, introducing delay even further. As Smart Grid has to act to the environment changes and events in near real-time, such delay is intolerable [8, 9]. The idea of Edge computing is to move the computation and data processing closer to data sources, in order to enable faster response time [8]. For example, in [9], it has been shown that Smart Grid system monitoring performance can be increased up to 10 times by moving the computation closer to the location where the data was generated.

Semantic technology: The role of semantic technology is to encode the meaning of data separately from its content and application code. This way, it is enabled that both machines and people can understand the data, exchange it and perform reasoning. In context of semantic technologies, ontologies are used to describe the shared conceptualization of a particular domain. Semantic descriptions are stored within the triple stores. RDF is often used for the representation of semantic data within triple stores. It consists of classes, their properties and relationships expressed in forms of triplets (*subject, predicate, object*). SPARQL is a language used for querying the RDF semantic triple stores. By executing queries against the triple store, it is possible to retrieve the results that can be further used by reasoning mechanisms in order to infer new knowledge based on the existing facts. In IoT systems, semantic technology is used for various purposes. It is widely adopted in cases when it is needed to achieve interoperability of heterogeneous devices [10]. In [11], a lightweight semantic framework was used for semantic annotation of the results obtained by computer vision algorithms in order to enable reasoning about the events that occurred within IoT-based video surveillance system and act accordingly. In this paper, we want to adapt the similar approach to [11] within Smart Grid architecture.

Domain-specific language (DSL): It is a programming language specialized for solving problems from a particular application domain. If the considered problem belongs to target domain, that problem is solved more conveniently than using general purpose programming languages. Their notation could be textual or visual (within modeling tools). Domain-specific languages are being adopted in IoT systems in order to decrease cognitive load introduced by the device heterogeneity and complexity of the underlying infrastructure. The domain-specific language scripts are further automatically translated to lower-level device-specific commands. For example, in [12], EDL domain specific language is used to describe the experiments carried out using robotic IoT devices. In context of

Smart Grids, visual tools based on domain-specific languages would enable much more convenient control and management for operators. This way, the implementation of complex scenarios is enabled by eliminating the need to deal with in-depth implementation details of the involved devices.

III. IMPLEMENTATION OVERVIEW

In this section, we propose Smart Grid architecture putting together the previously described technologies and present several aspects of system implementation.

System architecture and working principles: The measurement of electrical quantities is performed by smart devices, referred to as *Smart Meters*. They can either be microcontrollers, low-power single-board computers (such as Raspberry Pi) or even smartphones (as in our case). The advantage of using smartphones is the availability of built-in sensors and inputs (such as audio jack). Moreover, their rechargeable batteries and wireless mobile network availability give the ability to use smartphone devices conveniently even in less accessible areas. After that, the collected data is analyzed relying on data mining and machine learning techniques. The obtained results are semantically annotated, so the semantic reasoning can be performed against them in order to draw conclusion about the events that occurred. According to these results the corresponding actions are taken in order to adapt the Smart Grid to current consumption demands. The adaptation plan can be specified by operators, using a visual modeling tool utilizing a domain-specific notation. Finally, the device-specific commands are generated in order to respond to the changes detected in Smart Grid. The illustration of working principle is given in Fig. 1.
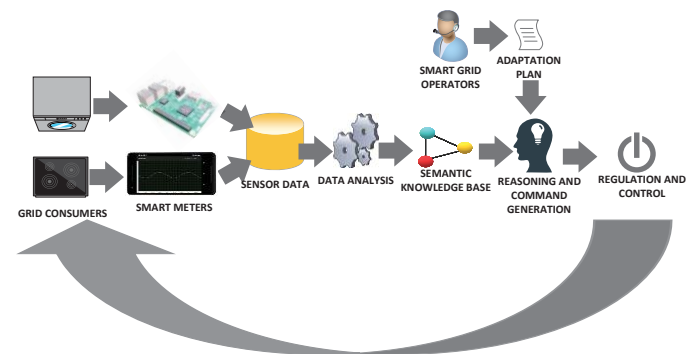
Fig. 1. Overview of automated adaptation process in Smart Grid

Android-based Smart meters: In [13], we have presented a method for acquiring electric measurement utilizing affordable Android-based devices. Voltage and current signals are acquired via voltage and current transformers from the power grid and both converted into voltage signals, which are then scaled down further to audio signal levels using variable resistors. After that signal goes directly into the devices 3.5 mm audio jack. Since many Android based devices support stereo microphone input, it makes them an ideal two-channel measuring platform for power signals. Sound card of the device performs analog to digital (A/D) conversion, so the data can be further processed by standard digital signal processing

methods, such as FFT-based algorithms. Moreover, it is useful to also record other data coming from device, such as temperature, timestamp and location. The collected measurements are sent to Edge server, via MQTT (Message Queuing Telemetry Transport)[1], a lightweight, publish-subscribe-based ISO-standard messaging protocol, working on top of TCP/IP. The messages are sent as JSON-encoded string. The smart measurement system is illustrated in Fig. 2.
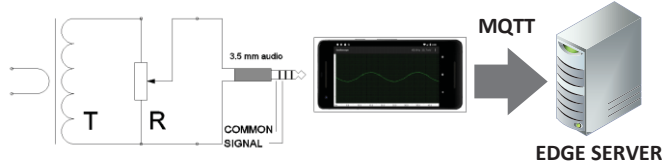
Fig. 2. Android based smart measurement

Data analysis mechanisms: For implementation of data analysis mechanisms, we rely on TensorFlow[2] for Python, an open-source library used for machine learning. We decide to use it, as it supports execution on GPU. Two mechanisms are implemented: 1) anomaly detection based on classification 2) load forecasting based on regression. In the first case, the training data contains voltage and frequency measurements with label (ok/anomaly), while in the second case it consists of average daily consumption (dependent variable) and average temperature (independent variable).

Semantic framework: A domain ontology (illustrated in Fig. 3) is defined in order to semantically annotate the results obtained during the process of data analysis. This way, it is possible to draw conclusion about the events that occurred by executing SPARQL queries and interpreting their results. Different types of events are considered: device failure, voltage anomaly, idle state of the consumer device etc. Moreover, for each of the events, it can be defined which are possible actions that could be taken in order to react to detected events, such as voltage regulation, turning off the device, switching the device to power saving mode.
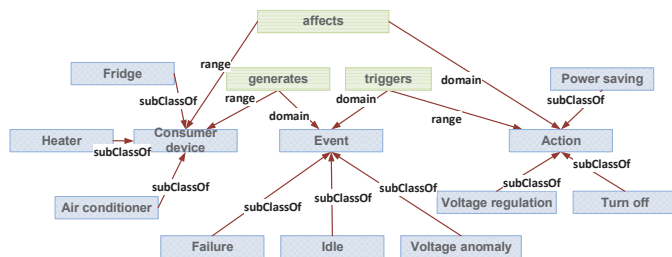
Fig.3. Domain ontology describing control within Smart Grid

Visual modelling tool for grid operators: It was developed using Node-RED[3] as a basis. It is an intuitive and extendable framework that is used for wiring together IoT devices, APIs and online services in novel ways, providing a browser-based editor with drag-and-drop user interface using the wide range of modeling elements (nodes). The domain-specific notation within the tool is described by a metamodel shown in Fig. 4. A model of a modeling language which defines the structure and constraints for a family of models. In this paper, it is used to

define a set of actions that need to be taken over the target devices in order to adapt Smart Grid to the environment changes when pre-defined environment conditions (related to the change that occurs) are satisfied. The conditions could be either some specific events or relational expressions with respect to a given pre-defined threshold.
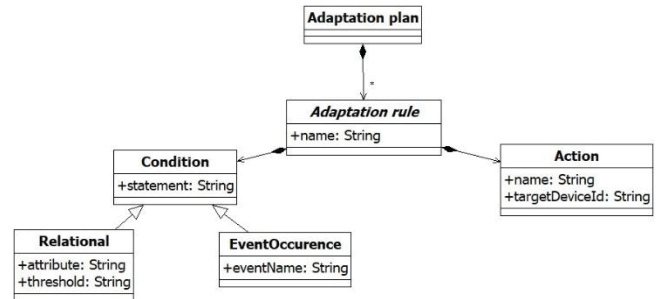
Fig. 4. Adaptation plan metamodel given in UML notation

Reasoning and command generation: It is performed according to the algorithm shown in Listing I as pseudo-code. Each condition from the adaptation plan is translated to SPARQL query and executed against the semantic knowledge base. If it returns results, then the corresponding code is generated and appended to the command script.

LISTING I
CODE GENERATION ALGORITHM

Input: sensor measurements, adaptation plan
Output: commands
Steps:
 1. Retrieve all the adaptation rules from the adaptation plan;
 2. Analyze sensor data;
 3. Semantically annotate results;
 4. For each of the adaptation rules
 5.   If(condition is true)
 6.   then generate command targeting adaptation_rule.targetDeviceId;
 7. end for each
 8. end

IV. EVALUATION AND RESULTS

In this section, we present experimental results achieved utilizing the described framework from two different perspectives. First, we consider the accuracy of anomaly detection and load forecasting mechanisms. Moreover, the adaptability responsiveness of the implemented system is analyzed considering the processing time necessary for each step. The evaluation was performed on a server equipped with AMD Ryzen 7 1700X octa-core CPU running at 3.80GHz, 64GB of DDR4 RAM and NVIDIA Quadro P2000 GPU with 4GB of VRAM.

In Table II and Table III, the achieved results using the presented approach for anomaly detection and load forecasting implemented using TensorFlow are given. As it can be noticed

---

[1] http://mqtt.org/
[2] https://www.tensorflow.org

[3] https://nodered.org/

observing the achieved results, both data analysis mechanisms have satisfactory performance in cases of different training/test set ratios, performing better in case of larger training sets.

TABLE II
ANOMALY DETECTION RESULTS

Training set size	Test set size	Correct/Test size [%]
75	150	89,93
100	150	92,58
125	150	95,17

TABLE III
LOAD FORECASTING RESULTS

Training set size	Test set size	Relative error [%]
75	150	13,51
100	150	12,89
125	150	12,21

In Table IV, an overview of the achieved processing times for data analysis and code generation are shown for various cases of adaptation plan length (in number of rules) and consumer devices involved. Each rule targets a distinct device, while data analysis was performed for measurements collected during 300 seconds. For data analysis, the processing times for anomaly detection considering both the CPU and GPU execution are provided (70% training, 30% test set). According to the results, the time spent for data analysis increases with larger number of devices involved, as more devices generate larger amount of data. Moreover, the time needed for code generation also increases, as number of SPARQL queries that will be executed during the code generation process depends on number of adaptation rules involved. Another observation is that the code generation time for the same number of rules may vary, as it depends on number of adaptation rule conditions that are true, so the commands will be generated only then, as in the second and third case shown in Table IV. Finally, the execution of anomaly detection is up to 3 times faster when executed on GPU, instead of CPU, while the speed-up increases with amount of data, which is beneficial when adaptivity has to be performed in near real-time for huge number of devices.

TABLE IV
PROCESSING TIME OVERVIEW

Number of rules	Data analysis (CPU) [s]	Data analysis (GPU) [s]	Code generation[s]
1	1,74	0,93	1,62
2	2,68	1,06	2,44
2	2,64	1,09	1,91
3	3,71	1,24	3,08

V. CONCLUSION AND FUTURE WORK

In this paper, the enabler technologies for enabling adaptivity in IoT-based Smart Grid architecture were discussed and some implementation and evaluation aspects presented. It can be concluded that IoT-based technology has huge potential in this use case. While the achieved results in case of anomaly detection are comparable to a similar solution [4], the load forecasting performs slightly worse compared to [6, 7]. It can be concluded that more data was needed for training in case of load forecasting. However, our plan is to further work on the implementation, considering the adoption on Big Data technologies, evaluation on larger data sets, optimization of real-time performance, security and evaluation of various implementation variants.

REFERENCES

[1] A. H. Bagdadee, L. Zhang, "Smart Grid: A Brief Assessment of the Smart Grid Technologies for Modern Power System", Journal of Engineering Technology, vol. 8, no. 1, pp. 122- 142, 2019.

[2] M. Jaradat et al., "The Internet of Energy: Smart Sensor Networks and Big Data Management for Smart Grid", Procedia Computer Science 56 (2015), pp. 592–597, 2015.

[3] B. Rossi, S. Chren, "Smart Grids Data Analysis: A Systematic Mapping Study", ArXiv 2018, pp. 1-26, 2018.

[4] B. Rossi, S. Chren, B. Buhnova, T. Pitner, "Anomaly Detection in Smart Grid Data: An Experience Report", IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016), pp. 1-6, 2016.

[5] H. Shahinzadeh, J. Moradi, G. B. Gharehpetian, H. Nafisi, M. Abedi, "IoT Architecture for Smart Grids", 2019 International Conference on Protection and Automation of Power System (IPAPS), pp. 22-30, 2019.

[6] M. Božić, M. Stojanović, Z. Stajić, "Short-Term Electric Load Forecasting Using Least Square Support Vector Machines", Facta Universitatis, Series: Automatic Control and Robotics vol. 9, no. 1, pp. 141-150, 2010.

[7] A. B. M. S. Ali and S. Azad, "Demand Forecasting in Smart Grid", Green Energy and Technology, pp. 135–150, 2013.

[8] W.Z. Khan et al., Edge computing: A survey, Future Generation Computer Systems (2019), pp. 1-44, 2019.

[9] Y. Huang et al., "An Edge Computing Framework for Real-Time Monitoring in Smart Grid", 2018 IEEE International Conference on Industrial Internet (ICII), pp. 99-108, 2018.

[10] R. Agarwal et al., "Unified IoT ontology to enable interoperability and federation of testbeds", 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), pp. 70-75, 2016.

[11] N. Petrovic, "Surveillance System Based on Semantic Video and Audio Annotation Leveraging the Computing Power within the Edge", XIV International SAUM 2018, pp. 281-284, 2018.

[12] K. Kolomvatsos, M. Tsiroukis, S. Hadjiefthymiades, "An Experiment Description Language for Supporting Mobile IoT Applications", FIRE Book, European Commission, River Publishers, pp. 461-486, 2016.

[13] Đ. Kocić, N. Petrović, "Application of Android Based Devices in Analog Electric Signal Measurement", YuInfo 2019, Kopaonik, Serbia, pp. 1-5, 2019.

Traffic Grooming on Designing Elastic Optical Networks

Suzana Miladić-Tešić[1], Goran Marković[2] and Valentina Radojičić[3]

Abstract – **In the design of next generation optical networks, the concept of elasticity has been proposed. This means optical paths provisioning using just enough spectrum to best fit the user demands and diverse type of services. Between adjacent optical paths guard bands are needed and consequently a large portion of spectrum is wasted. To further improve the resource usage, traffic grooming technique could be applied. Grooming technique allows the establishment of optical tunnels carrying a several connections in a contiguous block of spectrum without inserting guard bands in between, therefore minimizing the spectrum usage or the number of transmitters. Solving the grooming problem together with the routing and spectrum allocation, as a key issue in elastic optical networks (EON), is a highly challenging task particularly in the case of large problem instances. In this paper we consider grooming capability at the optical layer. Using metaheuristic approach, we solved the traffic grooming problem with static traffic demands, therefore suitable on designing EON. The proposed algorithm aims to minimize the total spectrum usage while serving all traffic demands. Significant spectrum savings are obtained compared to the non-grooming case.**

Keywords – **elastic optical network, optical grooming, metaheuristic approach**

I. INTRODUCTION

Today's huge volume traffic demands cannot be efficiently satisfied with traditionally deployed wavelength division multiplexing (WDM) technology due to its coarse bandwidth granularity and rigid spectrum allocation. Based on O-OFDM (Optical Orthogonal Frequency Division Multiplexing) technology, spectrum-efficient, data-rate flexible and energy-efficient optical network architecture was analyzed in [1] to meet different traffic granularity needs. In such a network, flexible data rates are supported through bandwidth variable transponders (BVT) at the network edge and the bandwidth variable optical cross-connects (BV-OXC) in the network core. The term flexibility or elasticity refers to the ability of a network to dynamically adjust its resources such as the optical bandwidth and the modulation format, according to the requirements of each connection.

ITU (International Telecommunication Union) updated its G.694.1 recommendation [2] to include the flexible grid option based on a frequency slot (FS) concept. The frequency slot presents the minimum frequency range of an optical signal could take. The available optical spectrum is then divided into smaller granularity FSs and the optical connections (flexible lightpaths) are allocated a proper number of slots (bandwidth on demand). Unlike the current WDM frequency channels of 100 or 50 GHz width, a FS in EON could be of finer granularity, such as 25 GHz, 12.5 GHz or even 6.25 GHz.

Network performances could be further improved if some spectrum management techniques such as traffic grooming are been applied and combined with elasticity property. Traffic grooming enables grouping of traffic demands with the same source into one transmitter and switching them together over the network forming in such a way an optical tunnel with the capacity equal to the capacity of a transmitter. In this paper, we assumed a grooming approach to aggregate traffic directly at the optical layer and such eliminating the O/E/O (Optical/Electrical/Optical) conversions. Solving the traffic grooming (TG) problem together with the routing and spectrum allocation (RSA), namely TG-RSA is a highly challenging task particularly in the case of large networks. We applied the bee colony optimization (BCO) metaheuristic approach to solve the static optical TG-RSA problem.

The paper is organized as follows. Section II describes the optical grooming problem in EON. Section III is dedicated to the proposed BCO metaheuristic applied to the researched TG-RSA problem. Simulations and results to demonstrate the benefits of optical grooming versus non-grooming case are given and discussed in Section IV. Concluding remarks are given in Section V.

II. OPTICAL GROOMING PROBLEM IN EON

A. Basic concept

Optical layer grooming has been considered by the research community in case of static traffic scenario [3-5] as well as for dynamic scenario [6, 7]. The problem was mostly solved using ILP (Integer Linear Programming) formulations for small size networks and using heuristic approaches for large networks.

The benefits of grooming technique are related to spectrum and transmitters' savings. Transmitters' savings arise from the fact that grouping several same source demands leads to a better utilization of capacity and therefore their less number. For switching demands, guard bands are needed to avoid interference. If a separate optical tunnel is needed to provision each of these demands, spectrum wastage by guard bands may occurred. Because of the orthogonality, demands starting from the same source within the same tunnel are not separated by guard bands and they are only needed between different optical tunnels. Therefore, more spectrum slots could be saved by eliminating the guard bands.

[1]Suzana Miladić-Tešić is with the Faculty of Transport and Traffic Engineering at University of East Sarajevo, Vojvode Mišića 52, Doboj 74000, B&H, E-mail: suzana.miladictesic@sf.ues.rs.ba

[2]Goran Marković is with the Faculty of Transport and Traffic Engineering at University of Belgrade, Vojvode Stepe 305, Belgrade 11010, Serbia, E-mail: g.markovic@sf.bg.ac.rs

[3]Valentina Radojičić is with the Faculty of Transport and Traffic Engineering at University of Belgrade, Vojvode Stepe 305, Belgrade 11010, Serbia, E-mail: valentin@sf.bg.ac.rs.

The grooming benefits are illustrated in Fig. 1. Connections with the same source (node **a** for example) are grouped since they share some links from source to destination and guard bands are not needed in this case. When a connection needs to be separated from the optical tunnel at any intermediate node (such as node **b** or **c** in Fig. 1), it is dropped or switched optically using BV-OXC and continue its way to the destination.
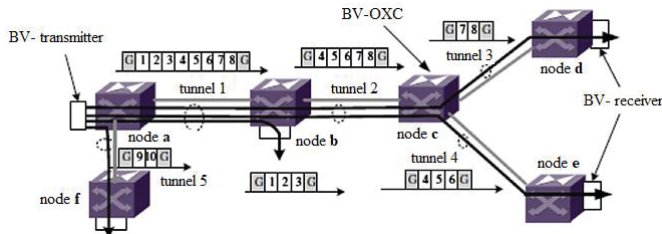
Fig. 1. Optical layer grooming in EONs [3]

B. Problem statement

Optical grooming, besides the main RSA constraints, includes the constraints related to the transmitter resources. Hence, the main constraints configuring in optical grooming problem are the following: 1) RSA spectrum contiguity constraint - to establish a flexible lightpath with the capacity of f frequency slots, the constraint assumes that f consecutive frequency slots must be allocated to establish a flexible lightpath, 2) RSA spectrum continuity constraint- the same FSs must be allocated on each physical link on a chosen route and 3) The number of groomed connections is limited by the capacity of BVT.

We solved the static traffic grooming problem in EON which typically appears during the network planning or designing phase. The problem could be defined in the following way: for a given traffic demand matrix (specified by the number of requested FSs between network node pairs), available FSs on network links and the capacity of transmitters, minimize the total spectrum usage with the assumption that all connection requests have to be satisfied in the network by aggregating multiple optical connections with the same source into one transmitter. By grooming the connections that share the longest common route from the same source node, the number of guard bands is minimized. As a result, the spectrum gain is maximized. Spectrum gain, H, defined in [3] has been incorporated in our model:

$$H = 2G \times X - (Y - Z) \times W'$$ (1)

where G presents the guard band ($1G$ on each side of the tunnel), X refers to the common route length (number of common links), Y refers to the length of the path that is considered for grooming (path that has the same source node), Z is the length of the first path placed in a tunnel with a capacity of ($W + 2G$) and W' is the number of FSs for the connection request Y. These terms are going to be explained in more detail in section III.

For the routing subproblem, fixed-alternate routing (FAR) method is used, which assumes that k- shortest paths are calculated in advance for each node pair. For the frequency allocation subproblem, the first fit (FF) policy [8] is applied which assumes that FSs are indexed and a list of indexes of available and used slots is maintained. The policy always attempts to choose the lowest indexed slot from the list of available slots and allocates it to the lightpath to serve the connection request [8]. The objective function F is assumed as follows:

$$F = \min \sum_l FS_l, \forall l \in L$$ (2)

and presents the total spectrum usage (the number of occupied frequency slots on all network links L) while FS_l presents the number of occupied slots on link l.

III. BCO METAHEURISTIC APPLIED TO THE OPTICAL GROOMING PROBLEM

C. BCO metaheuristic

Metaheuristics as a global search method examine solutions produced by a heuristic algorithm and move to better ones in a sophisticated manner [9]. BCO metaheuristic is at first proposed by Lučić and Teodorović [10] to solve complex transportation engineering problems. An overview of BCO with its applications could be found in [11, 12]. It is a population-based stochastic random-search technique that is inspired by the foraging habits of natural bees looking for nectar sources.

During the search process, bees generate and evaluate their individual partial solutions by steps. Each step consists of two phases: the forward and backward pass. Every bee starts with the forward pass to discover its partial solution and after that makes the backward pass (flies back to the hive) to evaluate and compare the quality of its solution. Bees exchange the information about the quality of their solutions in the hive. The quality of the bee's solution corresponds to the considered objective function value. During each step, every bee decides with a probability whether to discard the created partial solution and become an uncommitted follower or to continue to expand its current solution with or without recruiting other bees. The bees have a certain level of loyalty to their partial solutions, depending on its solution quality. Bees that are loyal to their solutions form the set of recruiter bees advertising their solutions, while other bees from the colony become uncommitted followers. The sets of recruiter bees and uncommitted followers are changed from one backward pass to another. Each uncommitted bee accepts previously created partial solution of the recruiter bee, but in the next forward pass every bee is free to continue the solution exploration independently of other bees. The best-found solution among B created solutions in the iteration is saved. The algorithm iterates through the pre-specified number of iterations and the best solution obtained during all iterations is chosen as the final solution [10-12]. Therefore, the BCO

optimization consists of the following phases: initialization, partial solutions generation, solutions comparison and recruitment phase.

The initialization phase requires the input data such as: the number of iterations I, number of demands (requests) r to be tested in each algorithm step, number of bees B in population, physical network topology given by the set of nodes N and set of physical links L, set of k- shortest paths for each node pair, guard band value, capacity of transmitters U and the traffic demand matrix.

During each forward pass (or algorithm's step) s =1, 2, ..., S, every bee investigates a given number of demands chosen from the set of all demands in a random manner. By choosing a specific demand, bees attempt to establish a lightpath between one real source-destination node pair in the optical network and find the same source node demands. In each new step the bee chooses the demands that have not been previously tested. The exploration procedure is performed until all traffic demands are tested during an iteration.

D. Grooming procedure

We explored the application of BCO metaheuristic while solving the offline optical grooming problem in EON and denoted our algorithm as BCO-TG-RSA.

The grooming procedure is incorporated in the generation of partial solutions and contains the following steps during an iteration:

Step 1 - For each shortest path of the randomly chosen demand, the network is searched for the first possible placement with the capacity $(W + 2G)$, where W is the number of FSs requested for a given demand. The maximum index of the occupied FSs is determined for every route/path and the one with lowest starting spectrum slot index is chosen to establish the lightpath. We called this route as the *referent one* and it has been denoted as Z in the Eq. 1. This connection is the first one in a potential optical grooming tunnel.

Step 2- For every chosen demand in the previous step, connections having the same source and their k-shortest path routes are searched. The routes having common links (at least one link) with the referent one, starting from the source node are investigated for grooming (denoted as Y in the Eq. 1) following the spectrum gain given by the Eq. 1 and availability of frequency slots on all links of the route. Connections where spectral gain $H > 0$ are groomed and added into the tunnel with the capacity W'.

Step 3- If the spectrum gain is $H = 0$, only the first chosen demands with their *referent routes* are placed in the network, without grooming with some of the other requests.

IV. SIMULATIONS AND RESULTS

E. Simulation settings

To evaluate the performances of the proposed BCO-TG-RSA grooming algorithm, simulation experiments are carried out on two network topologies: scenario 1 and 2 (shown in Fig. 2) with bidirectional fibers.

The input data are: traffic demand for each (s, d) pair in terms of FSs number was randomly generated between 1 and $M \in [4, 8, 12, 16]$ with the uniform distribution, where M is the maximum required capacity in number of FSs between each node pair, guard band value $G = 1$, capacity of a transmitter $U = 16$ FS, the number of k- shortest paths $k = 3$ (using Yenn's algorithm), the number of requests tested in each step of the algorithm $r = 2$ and the maximum number of iterations $I = 10$. 10 different traffic scenarios are generated randomly and simulated. The objective is to minimize the total spectrum usage while serving all the requests. We used the population of $B = 3$ bees for scenario 1 (due to small network size) and $B = 5$ bees for scenario 2. All simulation tests have been performed by running the programming code that is implemented in Python, using PC with the processor on 2.71 GHz and installed RAM of 8 GB.
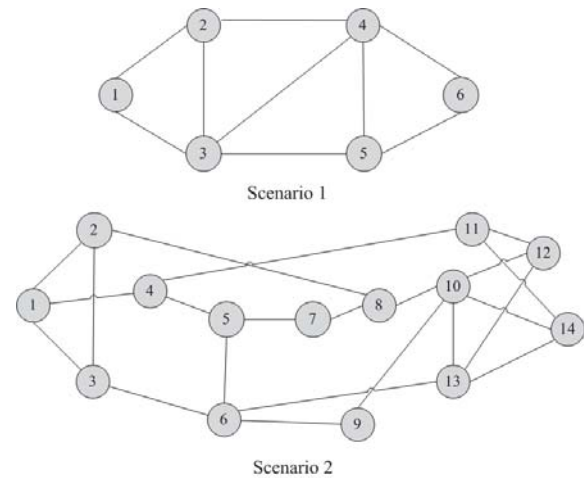
Fig. 2. Network topologies used for simulations

F. Results

After performing extensive simulations, it could be stated that grooming technique leads to a significant spectrum savings in case of large networks examples with complex traffic scenarios. Considering scenario 1, spectrum savings are within 5 %- 6 % versus non-grooming case. Note, that this is a small network example with simple traffic scenario and therefore grooming benefits cannot be fully expressed. Due to space limit, we presented in more detail the results for scenario 2 where grooming benefits are significantly higher.

Fig. 3 (a) shows the total spectrum usage (occupied frequency slots on all network links) of the grooming and non-grooming case for scenario 2. Different values of M are chosen in order to study the relationship between grooming efficiency and service granularity. It could be seen that optical grooming with the objective of minimizing the total spectrum usage achieves significant 5 %- 20 % of spectrum savings compared to the non-grooming case. Therefore, it is highly recommended to be applied in the design of EON.

We also analyzed the relation between spectrum savings and traffic granularity. The results of maximal and average

spectrum savings relative to non-grooming case are shown in Fig. 3 (b). The highest spectrum savings are obtained when M = 4- average 19.72 %. The lowest spectrum savings are obtained when M was set to be 16- average 4.8 %. This leads to a conclusion that spectrum savings decrease as the traffic granularity grows. The obtained results show that spectrum savings are mostly achieved for small traffic demands. The reason is that grooming opportunities are higher in this region and unused transmitter capacities are easier to be exploited. Also, it should be noted that grooming benefits increase as the transmitter's capacity increase as well as the guard band size.
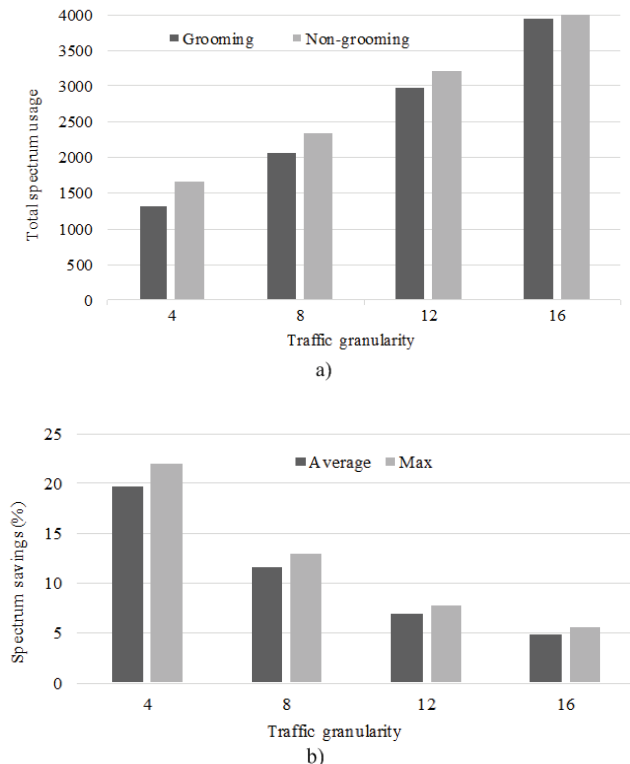
Fig. 3. a) Total spectrum usage and b) spectrum savings for different values of M

The simulations showed that, the greater number of bees does not necessarily lead to the improvement of the solution quality and that solution quality does not change significantly. Our experiences show that the population of $B = 5$ bees is enough to obtain high-quality solution within acceptable computational time with tens of seconds. Also, we performed simulations for higher number of iterations ($I = 20, 30, 50$) but due to the solution improvement and its repeating in iterations, we assumed $I = 10$. Therefore, the algorithm is able to find the same solution quality within the smaller number of iterations.

V. Conclusion

In light of the above, we believe that optical grooming has great potential for spectrum savings for future elastic optical networks and hence its designing. Our results show that optical grooming achieves significant spectrum savings compared to the non-grooming scenario, especially for small traffic demands granularity and for complex networks and traffic scenarios. Therefore, the algorithm such as the one we proposed is highly recommended for solving the complex grooming RSA problem in such networks. To the best of our knowledge, it is the first application of BCO method to the TG-RSA problem.

References

[1] G. Zhang, M. De Leenheer, A. Morea, B. Mukherjee, "A survey on OFDM-based elastic core optical networking", IEEE Commun. Surv. Tutor., vol. 15, no. 1, pp. 65-87, 2013.

[2] ITU-T G.694.1, "Spectral grids for WDM applications: DWDM frequency grid", v2.0, ITU, 2012.

[3] G. Zhang, M. De Leenheer, B. Mukherjee, "Optical traffic grooming in OFDM-based elastic optical networks [Invited]", J. Opt. Commun. Netw., vol. 4, no. 11, pp. B17–B25, 2012.

[4] P. D. Choudhury, N. Agarwal, T. De, "Spectrum and splitter utilization efficient traffic grooming routing and spectrum assignment in elastic optical networks", IEEE SPICES, Conference Proceedings, pp. 1-6, India, 2017.

[5] K. D. R. Assis, R. C. Almedia, A.F. Santos, "Grooming benefits on designing elastic optical networks: An exact formulation", SBFoton IOPC, Conference Proceedings, pp. 1-6, India, 2018.

[6] P.S. Khodashenas, J. Comellas, S. Spadaro, J. Perello, "Dynamic source aggregation of subwavelength connections in EON", Photonic Netw. Commun., vol. 26, no. 2-3, pp. 131-139, 2013.

[7] Z. Fan, Y. Qiu, C-K. Chan, "Dynamic multipath routing with traffic grooming in OFDM-based elastic optical path networks", J. Lightw. Technol., vol. 33, no. 1, pp. 275-281, 2015.

[8] B.C. Chatterjee, N. Sarma, E. Oki, "Routing and spectrum allocation in elastic optical networks: a tutorial", IEEE Commun. Surv. Tutor., vol. 17, no. 3, pp. 1776-1800, 2015.

[9] E. A. Varvarigos, K. Christodoulopoulos, "Algorithmic aspects in planning fixed and flexible optical networks with emphasis on linear optimization and heuristic techniques", J. Lightw. Technol., vol. 32, no. 4, pp. 681-693, 2014.

[10] P. Lučić, D. Teodorović, "Computing with bees: Attacking complex transportation engineering problems", Int. J. Artif. Intell. Tools, vol. 12, no. 3, pp. 375-394, 2003.

[11] T. Davidović, D. Teodorović, M. Šelmić, "Bee colony optimization part I: The algorithm overview", YUJOR, vol. 25, no. 1, pp. 33-56, 2015.

[12] D. Teodorović, M. Šelmić, T. Davidović, "Bee colony optimization part II: The application survey", YUJOR, vol. 25, no. 2, pp. 185-219, 2015.

MAC-layer Protocol for UWB-Based Single-Tag Indoor Localization System

Milica Jovanovic[1], Igor Stojanovic[1], Sandra Djosic[1] and Goran Lj. Djordjevic[1]

Abstract – **In this paper we present a MAC-layer protocol for UWB-based indoor localization system composed of a set of fixed anchor nodes and single mobile tag. The protocol we propose solves the problem of delivering ranging data from the mobile tag to the location server through the multi-hop sink-tree network of anchor nodes. The proposal regulates the process of network formation and employs a TDMA scheme wherein each anchor node is statically assigned a time slot for ranging and data forwarding, thereby avoiding collisions and improving system scalability.**

Keywords – **ultra-wide band, indoor localization, medium access control, sink-tree network.**

I. INTRODUCTION

Accurate location information is essential for the location-based services in many context-aware applications [1]. Among variety of localization technologies, the ultra-wide band (UWB) localization is considered to be the most promising radio-based localization technology available today, which offers the potential of achieving a centimetre level localization accuracy even in indoor multipath environments [2][3]. To modulate the information, the UWB uses ultra-short pulses with duration of less than $1\ ns$, which are transmitted over a large bandwidth in the frequency range from 3.1 GHz to 10.6 GHz [4][5]. The most important characteristic of UWB is large bandwidth in comparison with prevalent narrowband systems (e.g., Wi-Fi, and Bluetooth LE). Due to the inverse relationship between the time-of-flight (TOF) estimation error and signal bandwidth, the distance between two UWB transceivers can be measured with a high precision with excellent immunity against multipath fading [6].

In this paper, we consider a simple yet practical application scenario of single person localization in a multi-room indoor environment. The application setup includes a number of battery powered fixed anchor nodes distributed throughout the area of interest, a mobile tag node carried by a person to be localized, and a centralized location server. The tag periodically performs UWB ranging with surrounding anchors and sends the ranging data to the location server. Although UWB signal can penetrate one wall, it usually cannot propagate through multiple walls, so the full coverage of the entire localization space (e.g., typical residential apartment) is not possible. Therefore, the delivery of ranging data is the main problem with this set up because of limited range of UWB communication in the indoor environment. In order to

solve this problem, we propose a MAClayer protocol that enables the tag to deliver ranging data to the location server through anchor nodes organized in a multi-hop sink-tree network.

Regarding MAC design for UWB indoor localization systems, several proposals have been presented in the recent past. An analysis of scalability of different MAC schemes in terms of tag density was presented in [7]. In [8] a WiFi-UWB MAC protocol is presented, in which the time difference of arrival (TDoA) based UWB indoor localization system is deployed on top of a WiFi ad-hoc mesh network. In order to allow simultaneous localization of multiple tags, and avoid collisions between UWB messages, the protocol adopts a TDMA approach with on-demand slot assignment. In contrast to previous works, our proposal is a pure UWB MAC protocol adapted to time-of-flight (ToF) based UWB localization and highly optimized for single-tag application scenarios with low to moderate location update rate requirements.

The rest of the paper is organized as follows. Section II explains the UWB-based localization, and a commonly used ranging method. Section III introduces new UWB-based single-tag indoor localization system, and describes its architecture. Section IV presents the proposed MAC protocol design. Section V discusses some key aspects of the proposed localization system and indicates potential directions for future research.

II. UWB-BASED LOCALIZATION

Indoor localization system typically consists of a set of reference nodes, so called anchors, placed at fixed locations in the area of interest, and a *tag* node carried by the person or attached to the object that needs to be localized. The UWB localization process involves two phases: (*i*) the ranging phase, during which the distances between the tag node and individual anchors are measured, and (*ii*) the localization phase, during which the current position of the tag is calculated through a multilateration algorithm by using distances from the ranging phase.

The distance between two UWB nodes is commonly estimated by carrying out the alternative double-sided two-way ranging (AltDS-TWR) method [9]. As shown in Fig.1, AltDS-TWR is a time-of-flight based method, which requires exchanging of three messages (*Poll*, *Response*, and *Final*) between an initiator (node *A*) and a responder (node *B*). During the message exchange, nodes *A* and *B* take timestamps $(t_1, \dots t_6)$ of receive and send events on the physical layer using their respective local clocks. The timestamps are then used to calculate the time of flight (T_{tof}), and therefore the distance between nodes *A* and *B*.

[1] with the Faculty of Electronic Engineering at University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, e-mail: {milica.jovanovic, igor.stojanovic, sandra.djosic, goran.lj. djordjevic}@elfak.ni.ac.rs.
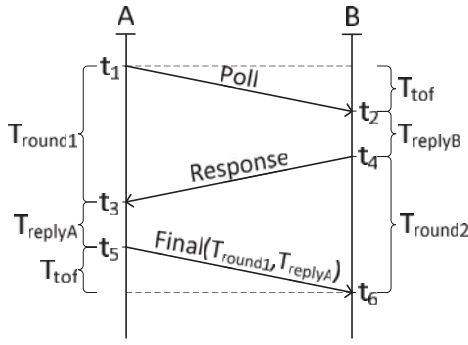
Fig. 1. Asymmetric double sided two-way ranging method

The time of flight is calculated by substituting measured round-trip times (T_{ro}, T_{ro}) and reply times (T_{replyA}, T_{replyB}) into the formula (1). Note that the distance is calculated by the responder node B after it receives T_{ro} and T_{replyA} from the initiator node A.

$$T_{tof} = \frac{T_{ro} \quad T_{ro} \quad -T_{replyA}T_{replyB}}{T_{round} +T_{ro} \quad +T_{replyA}+T_{replyB}} \tag{1}$$

Although UWB is not a new technology, its widespread adaptation has recently been accelerated by the commercialization of the IEEE 802.15.4-compliant UWB transceivers, such as DW1000 [10]. The DW1000 transceiver provides high precision UWB ranging and high data rate communications up to 6.8 Mbps. The DW1000 implementation of AltDS-TWR takes about $5\,ms$ with ranging precision of $10\,cm$ indoors.

III. PROPOSED UWB-BASED INDOOR LOCALIZATION SYSTEM

A. System Overview

The proposed UWB localization system aims to enable localization and tracking of a walking person in a complex multi-room indoor environment. The system is composed of multiple battery powered UWB nodes, including: a) a set of N static anchor nodes placed at fixed and known positions in the localization environment, and b) single mobile tag node carried by the person to be localized. Distances between the tag and anchor nodes are estimated through AltDS-TWR method, with anchor nodes acting as the initiators (node A in Fig. 1), and tag node as a responder (node B). The estimated distances are collected by the tag, and then sent to the location server (LS), which executes the localization algorithm to obtain the estimated location of the tag.

The choice of using TOF localization approach requires a specific MAC protocol design. First, in order to save the energy, the protocol should organize ranging operations in a way to minimize the idle listening of UWB nodes. Second, the protocol has to provide a mechanism for delivering the ranging data through the anchor nodes in cases when the location server is out of range of the tag node. Finally, the protocol needs to allow system installation with minimal setup and effort.

B. Network Architecture

The logical organization of the UWB-based localization system is shown in Fig. 2. In the proposed localization system, one of anchor nodes plays a role of network coordinator (C). In addition to participating in UWB ranging with the tag (T), like any other anchor node, the coordinator also serves as a gateway between the UWB network and the location server (LS), and provides the synchronization service for the entire UWB network. The remaining anchors are referred to as peripheral anchors (P). Each anchor is preassigned a unique identifier (ID) in range 0 to $N - 1$. The anchors are organized in a time-synchronized multi-hop network of sink-tree topology rooted at the coordinator node. Each peripheral anchor has its parent node in the tree. The level number (L) of a peripheral anchor is one greater than the level number of its parent. The level number of the coordinator node is $L = 0$. The depth of a sink-tree, L_{max}, is defined as the maximum level number in the network.
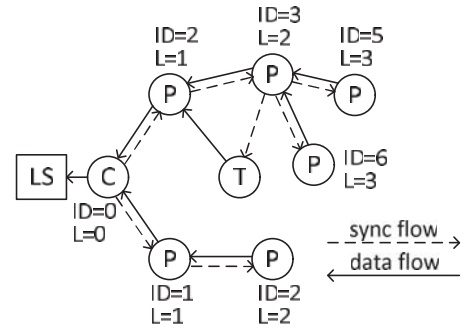
Fig.2. Sink-tree network of depth $L_{max} = 3$. **Notice**: C – coordinator node, P – peripheral anchor, T – tag, and LS – location server.

The parent-child relationship between nodes is used for both time synchronization and routing of ranging data from the tag to the coordinator node. A network-wide synchronisation is achieved through distribution of sync beacons. The coordinator node periodically broadcast a sync beacon to all its children. After receiving the sync beacon from its parent node, a peripheral anchor adjusts its local clock and rebroadcasts the beacon to its children. Due to its mobility, the tag node does not have a permanent parent node. Instead, it chooses one of anchors in the radio range as its temporal synchronization parent. After it stops receiving the sync beacons, the tag chooses a new synchronization parent.

Periodically, the tag initiates a ranging process, which includes performing ranging operations with all anchors in succession. For each successfully completed ranging operation, the tag calculates the distance to the corresponding anchor. At the end of the anging process, the tag formulates a report message containing the ranging data and sends it to one of neighboring anchors. Note that the tag is the only data source, while the coordinator node is the only data sink in the network. The peripheral anchors never generate their own data, but they only serve as relay nodes for forwarding ranging data. A peripheral anchor can only receive data from the tag or from its children. After data is received, the

peripheral anchor is obligated to send the data to its parent. In this way, by forwarding data from upper level to lower level nodes, the report message finally reaches the coordinator node.

IV. MAC PROTOCOL DESIGN

In order to organize UWB nodes in the sink-tree network, the time is divided into fixed-length periodic frames, and each frame is composed of N time slots of equal duration. The time slots are numbered, and each anchor owns one slot in the frame according to its ID. A peripheral anchor or tag is considered to be a part of the network only if it is in SYNC state, i.e., it is time synchronized with its parent node. In the SYNC state, anchor is active in its own time slots, and in the slots owned by its parent, only. At the beginning of its own time slot, anchor sends the *Poll* message containing its ID and level number. The *Poll* messages play role of sync beacons. In the parent's slot, peripheral anchor or tag receives a *Poll* message from its parent (Fig 3(a)). Peripheral anchor or tag uses parent's *Poll* message to adjust is local clock, and to set its level number to one greater than the level number contained in the message. After sending the *Poll* message in its own slot, the anchor waits for a possible response. If the *Response* message is received from the tag, the anchor completes the ranging procedure by responding with the *Final* message (Fig. 3(b)). If a data message (*Data*) is received from a child node, the anchor temporary buffers the received message (Fig. 3(c)). The buffered message will be resent by the anchor in the first next slot owned by its parent.
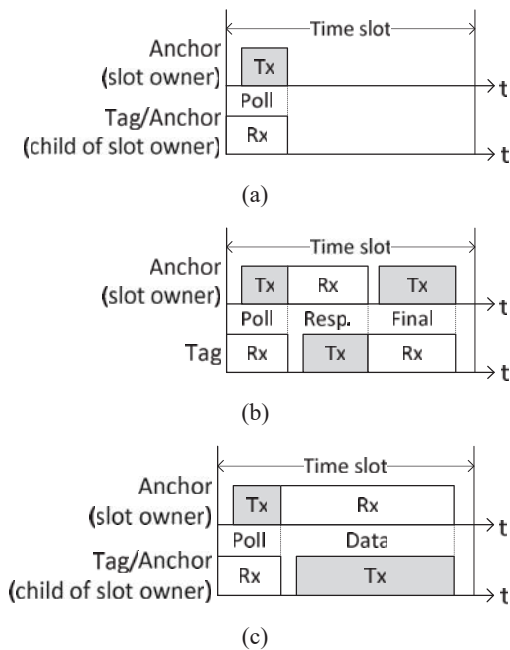
Fig. 3. Activities in a time slot: (a) synchronization only, (b) ranging, and (c) synchronization and data forwarding.

State diagrams of coordinator node, peripheral anchor, and tag are shown in Fig. 4. Being the only source of sync beacons, the coordinator node is considered to be permanently in SYNC state (Fig. 4(a)). Any other node (i.e., peripheral

anchor, or tag) has to follow a specific procedure to synchronize with the coordinator, i.e. to enter the SYNC state.

The peripheral anchor begins its lifetime in NO_SYNC state (Fig. 4(b)). In this state it keeps its UWB transceiver in the receive mode. If no *Poll* message has been received for the duration of an entire frame, the peripheral anchor turns off UWB transceiver, makes a pre-specified pause of T_d, and then it tries again. After receiving a *Poll* message, the peripheral anchor adjusts its local clock and moves to SCANNING state. In the SCANNING state, the peripheral anchor wakes up in every time slot during an entire frame. Among all the anchors from which the *Poll* message was received, the peripheral anchor chooses the one with the smallest level number as its parent, and then sets its own level number accordingly. By selecting the lowest-level neighbouring anchors for the parent nodes, the protocol tries to construct a sink-tree of as small depth as possible. In SYNC state, the peripheral anchors continues to receive *Poll* messages from its parent, and resynchronises its local clock with each message received. In the case of missing *Poll* message, the anchor returns to SCANNING state in order to select a new parent.
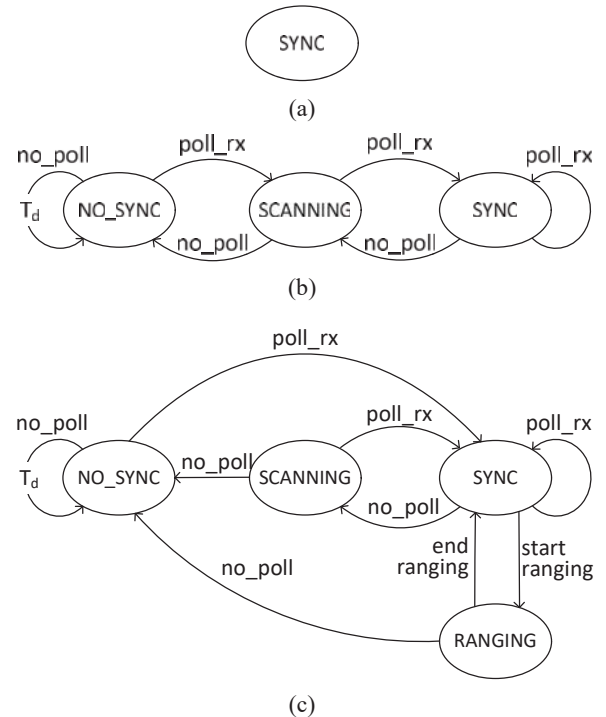
Fig. 4. State diagram: (a) coordinator, (b) peripheral anchor, and (c) tag

The synchronization procedure for tag node is somewhat simpler. Because there is no need to choose the lowest level neighboring anchor as the synchronization parent, the tag enters SYNC state as soon as a *Poll* message is received in NO_SYNC state. Also, after the synchronization is lost in SYNC state, the search for a new synchronization parent ends once the first *Poll* message is received in SCANNING state. The ranging process is implemented by RANGING state, and can only be started if the tag is in SYNC state. At the end of the ranging process, the tag sends the report message to the anchor at the lowest level of all the anchors with which it

performed the successful ranging. Also, before returning to SYNC state, the tag selects the anchor at the smallest measured distance as its new synchronization parent. To prevent interference between ranging operations and forwarding of the report message, the tag should not initiate the new ranging process for at least L_{max} frame periods.

V. Discussion and Conclusion

Location update period, T_L is the most critical parameter of the proposed UWB localization system because it determines the frame period and hence the maximum system size (i.e., the number of anchors). As already pointed out, the tag is allowed to initiate a new ranging process only after the report message from the previous ranging process is delivered to the location server. The report message is forwarded in $L + 1$ hops, where L is the level number of anchor to which the tag sent the report message. The time needed for one hop depends on the relative positions of time slots of transmitting and receiving anchors within the frame, and it ranges from one time slot to the entire frame period, T_F. Therefore, in the worst case, the report message forwarding time equals $(L_{max} + 1)T_F$. Also, one entire frame period is needed for the ranging process. Hence, $T_L \geq (L_{max} + 2)T_F$. For example, in system with the depth of the sink-tree of $L_{max} = 3$, which is set to operate with the location update period of $T_L = 1s$, the frame period must be shorter then $T_F = 200ms$. Assuming the time slot duration of $5\,ms$, the system can comprise at most 40 anchors, which is sufficient for most practical use cases.

In conclusion, the single-tag restriction, which enables the adaptation of pre-determined time slot allocation within the frame, and sink-tree topology for data forwarding and time synchronization, considerably simplifies the MAC protocol design and allows obtaining significant power savings without performance loss in terms of network scalability and location update rate. One possible direction of further research would be to explore opportunities to increase the location update rate through scheduling transmissions of individual UWB nodes. Of particular importance is also the generalization of the protocol to the case of multi-tag localization.

Acknowledgement

References

[1] Y. Liu and Z. Yang, Location, Localization, and Localizability. Location-awareness Technology for Wireless Networks, Springer, 2010.

[2] G. Cheng, "Accurate TOA-based UWB localization system in coal mine based on WSN," *Physics Procedia*, vol. 24, pp. 534–540, 2012.

[3] M. Ghavami, L. B. Michael, and R. Kohno, Eds., *Front Matter Ultra Wideband Signals and Systems in Communication Engineering*, John Wiley & Sons, Ltd, February 2006.

[4] Z. Sahinoglu, S. Gezici, and I. Gvenc, *Ultra Wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols*, Cambridge University Press, New York, NY, USA, 2011.

[5] F. Zafari, A. Gkelias, and K. Leung. (Sep. 2017). ''A survey of indoor localization systems and technologies.'' [Online]. Available: https://arxiv.org/abs/1709.01015

[6] J. Zhang, P. V. Orlik, Z. Sahinoglu, A. F. Molisch, and P. Kinney, "UWB Systems for Wireless Sensor Networks," Proceedings of the IEEE, vol. 97, no. 2, Feb. 2009.

[7] M. Ridolfi, S. Van de Velde, H. Steendam, E. De Poorter, "Analysis of the Scalability of UWB Indoor Localization Solutions for High User Densities", *Sensors*, vol. 18, no.6, 2018.

[8] M. Ridolfi, S. Van de Velde, H. Steendam and E. De Poorter, "WiFi ad-hoc mesh network and MAC protocol solution for UWB indoor localization systems," 2016 Symposium on Communications and Vehicular Technologies (SCVT), Mons, pp. 1-6, 2016.

[9] D. Neirynck, E. Luk, M. McLaughlin, "An alternative double-sided two-way ranging method", Proceedings of the 13th Workshop on Positioning, Navigation and Communications (WPNC), Bremen, Germany, pp. 1–4, 19–20 October 2016.

[10] Decawave Ltd., DW1000 Datasheet. Version 2.12, 2016.

Method and Algorithm for Automatically Targeting of Unmanned Aerial Vehicle with Vertical Landing on Mobile Landing Site

Ilia Iliev[1], Georgi Stanchev[2] and Krume Andreev[3]

Abstract – **The article provides a method and algorithm for automatically targeting of unmanned aerial vehicles with a vertical landing on a mobile landing site commensurate with its dimensions, using GPS for civilian free use and a modified radio navigation system for pseudo-conical scanning.**

Keywords – **Automatic landing of UAV; pseudo-conical scanning; radio-navigation and radiolocation.**

I. Introduction

The landing of unmanned aerial vehicles (UAVs), and in particular those of a lower category, is carried out under the supervision of a qualified operator. The landing site is usually commensurate with the dimensions of the UAVs and GPS accuracy of the GPS for civilian purposes is not sufficient to target it directly. This is done using a differential GPS, the cost of which is significant [1]. In patents [2,3], the landing is conducted by a image recognition of the mobile landing site. After the landing, a protective enclosure is erected at the end of the landing site, which is a very precarious means of stably restraining UAVs. A solution is also available which is based on the automatic targeting of UAVs to the center of the stationary landing site through a modified pseudo-scanning method [4].

The functional scheme and operation of a vertical landing gearless landing system on a mobile landing site in which the present method can be implemented is described in [5]. This article presents the method and algorithms for its realization.

II. Staged implementation of the automatic landing method on a mobile object

The landing method on a mobile object is realized by a series of operations, whose interrelationship is also presented in the form of algorithms (Figure 1).

[1]Ph.D. Ilia Iliev is a Professor at the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: igiliev@tu-sofia.bg .

[2]Ph.D. Georgi Stanchev is an Assoc. Professor at Faculty of Mechanical Engineering, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gstanchev@tu-sofia.bg .

[3]Krume Andreev is Ph.D. student at the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: andreev.k@abv.bg .

A. STAGE 1 - Positioning of the UAS over the pseudo-conical scan zone

Block 1. Sending the UAVs command to the moving landing site by an operator or automatically after the task has been performed and transmitting the coordinates of the erroneously positioning cylinder as defined in [4,5].

Block 2. The on-board communication-information module directs the UAVs to it, driven by continually updated GPS coordinates of the erroneously positioning cylinder.

Block 3. Intrusion of the UAVs into an area for close communication and connection between the UAVs and the landing site through their wireless interfaces for close communication. If there is no communication - it cyclically searches and continues to fly to the coordinates.

Block 4. Switch on the microwave receiver on the UAVs and the microwave transmitter at the landing mobile site. The transmitter, via an electronically pseudo-conical scanning antenna mounted at the center of the landing site, radiates continuously sequentially the four modulated signals in four beams. The modulated signals carry the landing site identification number, the identification code of each beam (left, right, forward, backward), the current GPS coordinates of the erroneously positioning cylinder, the speed and the change of motion direction vector, based on the gyro. The latter contributes to a faster response to the movement of the landfill site. After this data, each beam emits a constant signal to measure its level.

Block 5. There is a check. Are these signals accepted by UASs? If NO continues cyclical searching + emergency procedure. If YES, follows:

B. STAGE 2 - Signal processing and positioning of the UAS at the center of the pseudo-conical scanning zone

Block 6. Processing of the level of signals received by pseudo-conical scanning rays (described in Block 4) from an information microcontroller by algorithm from [6].

The position of the UAV is determined by processing the measured power of the received signal from the four beams of the antenna by pseudo-conical scanning. For this purpose, different algorithms can be used, as in the [6], the classical conical scan method, the least mean square (LMS) estimator, the Kalman filter (KF) based on the classical conical scan method and others. They are adapted to pseudo-conical scanning, the specificity of the controled object, the landing conditions and the instrumental error in measuring the received power.

Block 7. Targeting the UAVs to the central axis 0Z above the landing site as a result of the data and processing in Block

Fig. 1. Algorithm for realization of the method

accordance with Block 6 and the altimeter. Communication is routed through proximity communication interfaces.

Block 10. Upon reaching a few centimeters above the landing site, command is sent to turn the electro-magnetic or electro-mechanical gripping device on the landing site, depending on its inertia. A landing sensing sensor can also be set up to automatically engage the landing. Successful landing message is sent. With the UAVs landing, it must be firmly engaged on the site, especially at higher speeds on the mobile object. Electro-mechanical clamping can be done with two side-clips relative to the landing site driven by micro-electric motors.

III. ADDITIONAL REMARKS

The algorithms to perform the individual stages are autonomous and run sequentially, each ending with an emergency signal to the landing site if problems arise. In this case, its movement must be stopped to ensure a safe landing and, if necessary, stopping the movement of the object and moving to manual control.

IV. CONCLUSION

The proposed method, based on pseudo-radio scanning in radio navigation, is suitable for targeting and controlling the landing of plumbing aircraft on a mobile landing site commensurate with their gauges using civil GPS systems.

REFERENCES

[1] https://www.u-blox.com/en/product/c94-m8p#tab-documentation-resources

[2] Fumio Ohtomo, Kazuki Osaragi, Tetsuji Anai, Hitoshi Otani, US Grant Patent number US20120277934A1, "Taking-Off And Landing Target Instrument And Automatic Taking-Off And Landing System", 2011.

[3] Jingxuan Sun, Boyang Li, Yifan Jiang and Chih-yung Wen, "A Camera-Based Target Detection and Positioning UAV System for Search and Rescue (SAR) Purposes", Sensors 2016, 16, 1778; doi:10.3390/s16111778.

[4] Andreev K., Stanchev G., "Flight Safety Sensor and Auto-Landing System of Unmanned Aerial System", Int. J. of Reasoning-based Intelligent Systems (IJRIS), Japan, 2019.

[5] Arnaudov. R, Stanchev G., Andreev K., "System for automatically targeting of unmanned aerial vehicle with vertical landing on mobile landing site", 54th International Scient. Conference on Information, Communication and Energy Systems and Technologies, Ohrid, Republic of North Macedonia, June 27-29, 2019.

[6] Souza A.L.G., Ishihara J.Y., Ferreira H.C., Borges R.A., Borges G.A., Antenna pointing system for satellite tracking based on Kalman filtering and model predictive control techniques, Advances in Space Research, 58,2016, pp. 2328–2340

[7] Official U.S. government information about the Global Positioning System (GPS) and related topics, https://www.gps.gov/systems/gps/performance/accuracy/

[8] EASA Navipedia, https://gssc.esa.int/navipedia/index.php/Galileo_General_Introduction

5, the result of which is fed to the onboard dashboard communication-information module. The on-board communication-information module receives information in which direction to move. It directs and positions just above the center of the landing site, where the signal levels of all four beams are equal. The gyroscope signal helps for a very fast flight direction correction to follow the wrong positioning cylinder and correspondingly the pseudo-conical scan range.

Block 8. Aligning speeds, turning on the altimeter and sending a landing permission request (automatically from the mobile station of landing site or from the operator). If no permission is given, action is made on the plan and flies to a landing or landing site of the mobile object or eventual manual landing control.

C. STAGE 3 - Downgrading the UAS and landing on the site

Block 9. After receiving permission, the UAVs shall descend downward to the center of the landing site, driven by the data and processed pseudo-conical scan signals, in

Antenna array designing with an application for navigation landing of unmanned flying vehicle

Ivaylo Nachev[1] and Ilia Iliev[2]

Abstract – **This paper shows the examining of the design on the four-element antenna for pseudo-conical scanning with the application for navigation and inclusion to the landing of the unmanned aerial vehicle.**

Keywords – **phased-array antenna, electronic scanning antenna, pseudo–conical scanning, the orientation of unmanned aerial vehicles for automatic landing.**

I. INTRODUCTION

With the entry of the unmanned aerial vehicles (UAV's) for various applications in the recent years, simulations and research results have emerged in scientific journals related to the automation of the process in the exploitation of the UAV. There are published different navigation and landing developments. Most of them are based on - differential GPS systems [1], pseudo-conical scanning [2], laser optical curtain [3], and landing site image recognition [4]. Some of these methods are designed with a big difference in landing area or they are expensive for realization.

This paper will consider the possibility to navigate and land of the UAV via a patch antenna array. As the main disadvantage of the pseudo-conical scanning [5] which is a problem can be the realization of the antenna and its high cost. The advantage of patch antenna arrays is their relatively small size, weight, and low cost. For this reason, this development can solve this problem.

II. ANTENNA ARRAY DESIGN METHOD

An important step in designing a patch antenna is to choose a dielectric substrate on which patches are made. The employed substrate for the proposed design is Rogers RO4003 [6]. To increase the bandwidth and efficiency of the antenna "suspended substrate" is used [7], where the thickness of the gap between the ground plane and the antenna is Δ=1mm. In this method, we use the equivalent Dielectric constant (7). The operating frequency is chosen to be fc = 10.525GHz, then wavelength λ = 28.50mm, and λg = 22.6150mm – the wavelength in the substrate. For calculation of the antenna dimensions, we are using the basic algorithm like in [8].

[1]Ivaylo Nachev is a student at the Faculty of Telecommunications at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ivaylonachev@yahoo.com
[2]Ilia Iliev, Professor at the Faculty of Telecommunications at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: igiliev@tu-sofia.bg

Step 1: Width (W) calculation: Width of the Microstrip antenna is given by the equation:

$$W = \frac{c}{2fc}\left(\frac{\varepsilon_r+1}{2}\right)^{-\frac{1}{2}} = 12.2mm, \frac{W}{h} > 1 \qquad (1)$$

Where c is the speed of the light c=$3*10^8$, ε_r is a dielectric constant of the substrate, fc is operation frequency and h is the substrate thickness.

Step 2: Effective Length (L_{eff}) calculation: Length of the microstrip radiator is given by:

$$L_{eff} = \frac{c}{2fc\sqrt{\varepsilon_r}} = 11.3mm \qquad (2)$$

Where ε_r is an effective dielectric constant of the substrate.

Step 3. Length extension (L_{exp}) calcualation: The actual length is obtained by the equation (accounting for the fringing fields):

$$L_{exp} = 0.412h\frac{(\varepsilon_{eff}+0.3)\left(\frac{W}{h}+0.264\right)}{(\varepsilon_{eff}-0.258)\left(\frac{W}{h}+8\right)} = 0.88mm \qquad (3)$$

Step 4: The actual length of the Patch (L) calculation: The actual length of the patch is given by:

$$L = L_{eff} + 2 L_{exp} = 12.96mm \qquad (4)$$

Determination of effective dielectric permittivity ε_{eff}:

$$\varepsilon_{eff} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2}\left(1 + 10\frac{h}{W}\right)^{0.5} = 1.58 \qquad (5)$$

The length of the guided wave:

$$\lambda g = \frac{\lambda}{\sqrt{\varepsilon_{eff}}} = 22.62 \qquad (6)$$

The influence of the gap is taken into account:

$$\varepsilon_{req} = \frac{\varepsilon_r(h+\Delta)}{(\varepsilon_r+h)\Delta} = 1.72 \qquad (7)$$

Where Δ is the thickness of the gap. The new dielectric constant in technology with the suspended substrate is ε_{req}=1.72, where the characteristics of the air are also taken into account.

In Table 1 are presented the values of one element from the antenna array. The simulation model of the antenna array structure is shown in Figure 1. Corporate feed network is used, thereby scanning is available in

azimuth and elevations. The rounded edges of the square model of patch antenna provide circular polarization. It is important in navigation because there is interference between the antenna and the object sought. The antenna will not be able to detect the subject with ordinary linear polarization.

TABLE I
THE VALUES OBTAINED FOR THE DIMENSION OF THE ANTENNA COMPONENTS

	W [mm]	L_{eff} [mm]
Calculated	12.2	23.96
After simulation	19.5	10.5
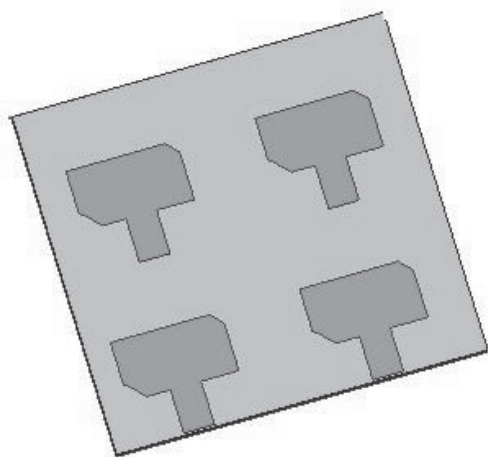
Fig. 1. Patch array 2x2 – simulation model

III. TRANSMISSION LINES PHASE SHIFTER

The antenna feeder uses systems of connected microstrip lines (MS), delay lines (DL), T-junctions and quarter wave transformers in planar realization. The designed antenna is for realization with the pseudo-conical scanning. Pseudo-conical scanning is a method in which the antenna beam is shifted from its central axis. The beam has several states. Every state moves the beam away from the beam position from the previous one. The beam makes a whole circuit around the central axis. In our case - by sequentially changing these states, a flying device measure beams in different antennas states. When the signal levels of the four beams are equal, it is assumed that, that the device is in the center.

The designed antenna has five states – two states in azimuth, and two states in elevation, and one main state - the beam is the same as the center axis of the radiation. These are accomplished by increasing the electrical length by 90º of two patch elements using a delay line. Each of different beam has width 18º. Different beams are at a distance of 40º. The center of each beam is offset at 20º from the central axis.

Figure 2 shows the block diagram of the feeder network. The antenna is supplied on feed point, then through the system made up of microwave elements, divided into four equal parts - 1/4 in each antenna. The control between diverse states of the antenna beams is performed by the switches represented in the block diagram. For this purpose, it is appropriate to use it is a PIN switch – instead of one diode, two diodes sequentially connected. This increase isolation. The high isolation is necessary, because of the high operating frequency is used. This way avoids unwanted reflections and drops in the effect of the antenna. [9]
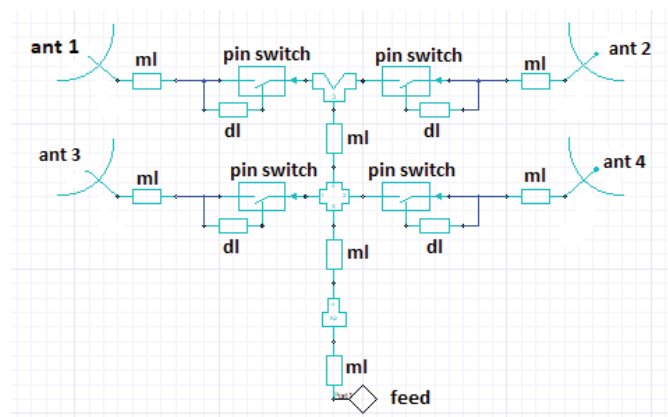
Fig. 2. Block diagram of feeder antenna network

IV. SIMULATIONS RESULT

The next figures show the results of the simulations. In Fig. 3 is displayed the return loss of each segment of the antenna. The figure shown that each patch has a level S11 > - 15dB, for the operation frequency. From this value follows that the VSWR of each antenna element meets the requirement for high antenna efficiency. Table 3 shows the S11 and VSWR of each antenna element in the array. Figure 4 shows the return loss on antenna array inputs, after connecting the array and the feeder network. The figure shows that S11 = - 27.5dB. By using the formula VSWR = (1+S11) / (1-S11) [10], we get the value VSWR = 1.07, from which it follows that the losses in the antenna will not be bigger than Mismatch loss = 0.004dB and return power <0.10% [11].
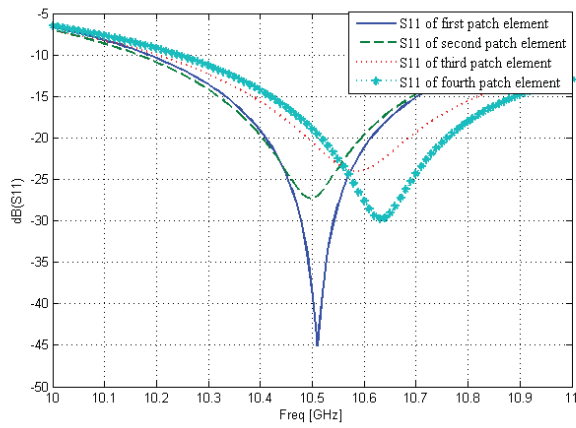
Fig. 3. Antenna bandwidth (S11)
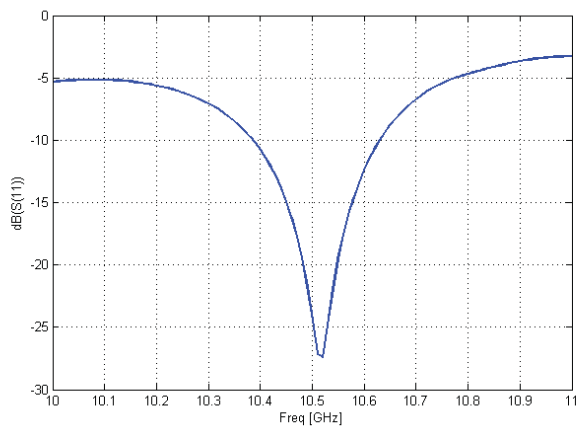
Fig. 4. Antenna bandwidth on feeder network inputs (S11)

TABLE II
VSWR IN DIFFERENT ANTENNA STATES

	S11 [dB]	VSWR
First antenna element	-45.13	1.04
Second antenna element	-28.32	1.07
Third antenna element	-20.69	1.10
Fourth antenna element	-19.55	1.11

Figure 5 shows the antenna radiation pattern in case 1 whit Gain = 12dB. To calculate the antenna coverage distance, the gain and the width of the beam plot are

generally -3dB from the main peak [12]. In this case gain $GAcase_1= 12 – 3 = 9dB$, whit the width of the beam = 20º. Taking the gain in each beam in scanning mode, shown in figure 6 and 7, respectively, are shown the two options of the antenna beam in azimuth and two in elevation. Each beam of the antenna has the same gain as case 1. The width of each the beam = 20º, the distance between the centers of the individual beams is 40º. The center of each beam is offset at 20º from the central axis.
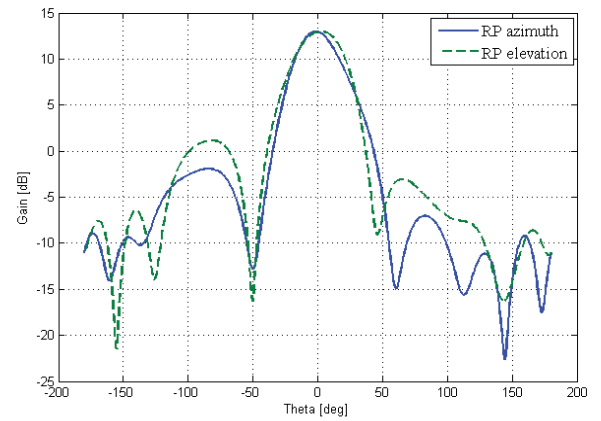
Fig. 5. Radiation pattern in case 1
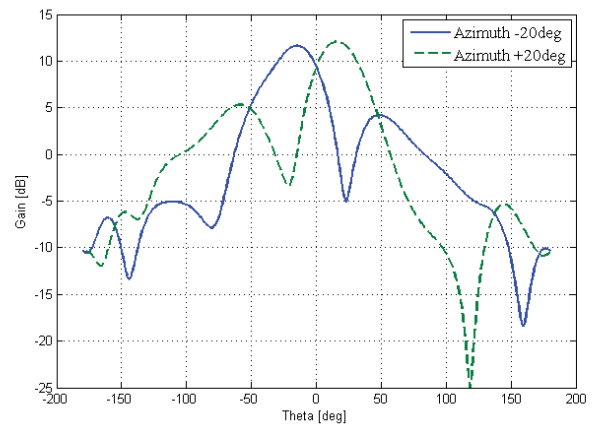
Fig. 6. Radiation pattern - two cases in azimuth
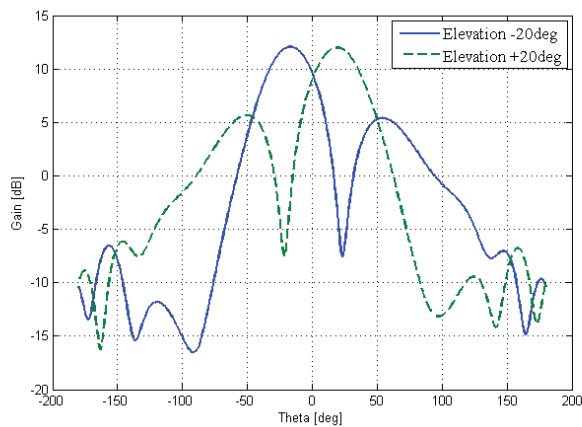
Fig. 7. Radiation pattern - two cases in elevations

V. CONCLUSION

The main advantages of this proposed antenna are:
- easy control with a microcontroller;
- small sizes;
- low cost.

This antenna can be used for various applications for automatic landing and control of the landing of unmanned aerial vehicles on a ground or mobile landing site. Also, this type of antennas can be used with various applications in radio navigation, robotics, and automation of production processes.

The designed antenna is suitable for integration into a landing site of the UAV, whit a modified pseudo-conical scanning method. The targeting of UAV to the landing site can be performed on its GPS-coordinates for civilian purposes. Because the GPS position error is very large, the beams of the diagram of the antenna array cover the area in which the drone is localized (a cylinder of wrong positioning). When UAV get position in the center of the landing site, by measuring the levels of the different antenna beams in scanning mode, the UAV can start landing. Thus, the landing pad may have dimensions commensurate with the gauge of the UAV.

REFERENCES

[1] https://www.u-blox.com/en/product/c94-m8p#tab-documentation-resources

[2] Andreev K., Stanchev G., "Flight Safety Sensor and Auto-Landing System of Unmanned Aerial System", Int. J. of Reasoning-based Intelligent Systems (IJRIS), Japan, 2019.

[3] Arnaudov R., Botusharov V., Stanchev G., Andreev K., "System for Automatic Targeting of Unmanned Aerial Vehicles in Vertical Landing ", Patent Application No. 112876, Bulgarian Patent Office, February 2019, Sofia.

[4] Fumio Ohtomo, Kazuki Osaragi, Tetsuji Anai, Hitoshi Otani, US Grant Patent number US20120277934A1, "Taking-Off And Landing Target Instrument And Automatic Taking-Off And Landing System", 2011.

[5] Andreev K., "Improvement of functional parameters on control systems of unmanned aerial vehicles", Ph.D. Thesis Dissertation, TU-Sofia, Bulgaria, 2019.

[6] Prabhakaran. M, Veeramani. R, 2015, "Comparative Analysis of Microstrip Phased Array Antenna Design", International Journal of Advanced Information Science and Technology (IJAIST), Vol.4, No.2, February 2015, ISSN: 2319:2682, DOI:10.15693/ijaist/2015.v4i2.306-310

[7] Rogers RO4003 specification: https://www.rogerscorp.com/documents/726/acm/RO4000-Laminates---Data-sheet.pdf

[8] Chang, Kai; "RF and Microwave Wireless Systems", 2000, ISBN: 0-471-35199-7.

[9] Arnaudov. R, Stanchev G., Andreev K., " System for automatically targeting of unmanned aerial vehicle with vertical landing on mobile landing site", 54th International Scient. Conference on Information, Communication and Energy Systems and Technologies, Ohrid, Republic of North Macedonia, June 27-29 2019.

[10] Balanis A. Constantine, Antenna Theory: Analysis and Design, 3rd Edition, ISBN-13: 978-0471667827

[11] https://www.markimicrowave.com/blog/wp-content/uploads/2016/11/return-loss-to-vswr.pdf

[12] http://www.rfwireless-world.com/calculators/Antenna-Range-Calculator.html

Strip Horn Array for X–band Operation

Nenad Popovic[1], Predrag Manojlovic[1], Ivana Radnovic[1] and Bojan Virijevic[2]

Abstract – **The paper presents the computer simulation and the measurement results of the realized strip horn antenna array intended for operation in the X-band (9.5-10.5 GHz). The proposed antenna is realized by placing the array of four quarter-wavelength monopole antennas into the wide short-circuited stripline whose strips are linearly expanding toward the antenna aperture at an angle of 60°.**

The array is fabricated with 20 mm wide and 1 mm thick aluminum strips. The arms of the horn are 60 mm long whereas the aperture angle is 60°. The excitation is performed through the SMA connector pins which are placed at the quarter wavelength ($\lambda_g/4$) distance from the short-circuited strip.

Keywords – **Microwaves, microwave antenna arrays, horn antennas, strip horn.**

I. INTRODUCTION

Horn antennas [1-3] represent the oldest forms of microwave antennas (1897 - Jagadish Chandra Bose). Owing to their simple design, low-cost fabrication and wide operating range (typically 10:1) they are often used in various applications: microwave telecommunications, radars, radar weapons, automatic door systems, distance measurements, alarm sensors, etc.

There are three basic types of horn antennas: sectoral, pyramidal and conical. All of them are characterized by the gradual widening of the waveguide aperture creating the gradual transition from the characteristic impedance of the waveguide to the impedance of the free space (around 377 Ω), thus reducing the reflection at the horn input. Excitation of sectoral and pyramidal horns is usually realized with rectangular waveguide with dominant wave type TE_{10} [1] (and only one component of the electric field $E_y=E_0\cos(\pi x/a)$, [2]). The horn antenna aperture size varies from around 1.5 wavelengths for small-sized to around 6 wavelengths for medium-sized horns [2].

According to [2], an array with N identical elements with identical amplitudes but each succeeding element has a β progressive phase lead current excitation relative to the preceding element is referred to as a uniform array.

The array factor (AF) of an N-element linear array of isotropic sources is [4]:

$$AF = 1 + e^{j(kdcos\Theta+\beta)} + e^{j2(kdcos\Theta+\beta)} + \cdots + e^{j(N-1)(kdcos\Theta+\beta)} \quad (1)$$

Equation (1) can be written as:

$$AF = \sum_{n=1}^{N} e^{j(n-1)\psi} \quad (2)$$

where $\psi=kdcos\Theta+\beta$

Multiplying both sides of the expression (2) by $e^{j\psi}$, and after some manipulations, the array factor can be rewritten as:

$$AF = \left[\frac{e^{jN\psi}-1}{e^{j\psi}-1}\right] = e^{j[(N-1)/2]\psi}\left[\frac{\sin\left(\frac{N}{2}\psi\right)}{\sin\left(\frac{\psi}{2}\right)}\right] \quad (3)$$

When the reference point is in the physical center of the antenna array, (3) can be reduced to:

$$AF = \frac{\sin\left(\frac{N}{2}\psi\right)}{\sin\left(\frac{\psi}{2}\right)} \quad (4)$$

To obtain the maximum of the AF of a uniform linear array in the broadside direction, ψ has to be zero, i.e. β=0 – namely, all the elements must have the same phase excitation.

The normalized array factor for the four element array (N=4) is

$$AF = \frac{\sin[2\psi]}{\sin[\frac{\psi}{2}]}$$

and is shown in Fig. 1.
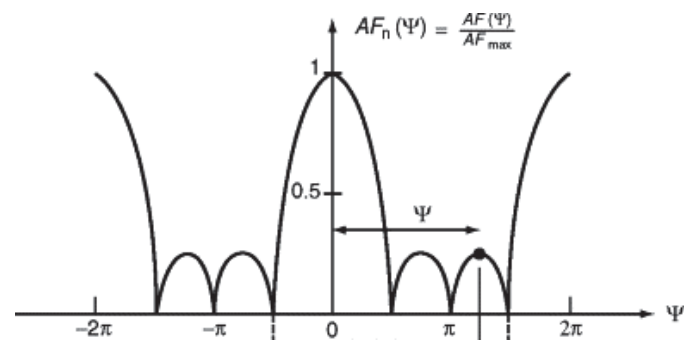
Fig. 1. Graphical representation of the normalized array factor of the 4-element array as a function of ψ.

[1]Nenad Popovic, Predrag Manojlovic and Ivana Radnovic are with the Research and Development Department, IMTEL Komunikacije, Blvd Mihajla Pupina 165b, Belgrade, Serbia (e-mails: nenad@insimtel.com, pedja@insimtel.com, ivana@insimtel.com).
[2]Bojan Virijevic is with the Military Technical Institute (VTI), Ratka Resanovića 1, 11030 Belgrade, Serbia.

II. DESIGN OF THE STRIP HORN ANTENNA ARRAY

The linear array [1-2] of strip horns that are fed in-phase can be realized in two ways – as individual strip horns [5-11] positioned along the straight line at the mutual distance of $\lambda_0/4$ or with the common reflector system with radiating elements spaced $\lambda_0/4$ apart, where λ_0 is the free space wavelength at the center frequency (f_c=10 GHz) of the frequency range of interest. Figures 2a and 2b display both concepts of the design.

(a)

(b)
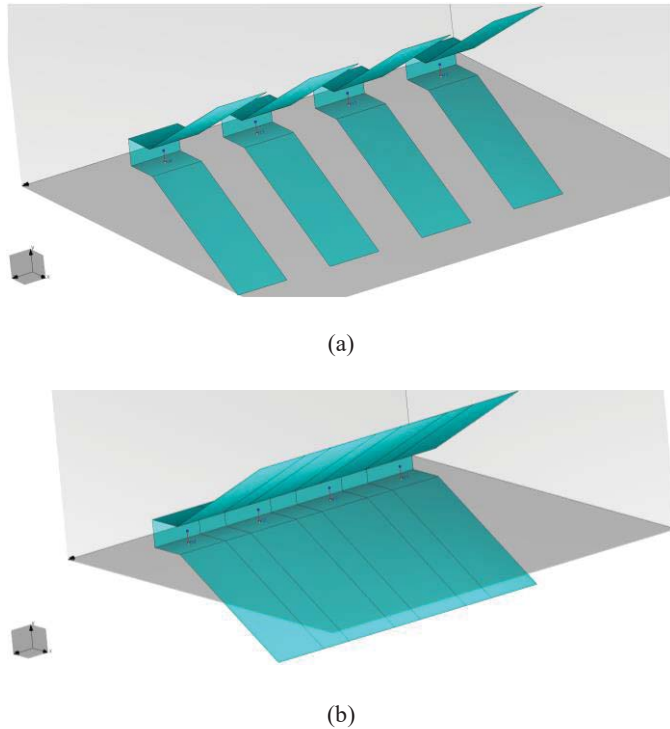
Fig. 2. Linear array of strip horn antennas: (a) Linear array of 4 individual horn antennas, (b) Linear array of 4 radiating elements in the common reflector system.

The excitation of the individual radiating elements in the array is performed through the feed network [3, 6-7] realized in microstrip technology on the RO4003C dielectric substrate with permittivity ε_r=3.38, and thickness h=0.2 mm. The microstrip lines' widths and lengths are calculated at the center frequency (f_c=10 GHz) of the X-band. The radiating elements – monopole antennas – are realized with the SMA connectors' pins. Figures 3 and 5 display the photographs of the separate array feeding network and its integration with the linear array of strip horn antennas, respectively.
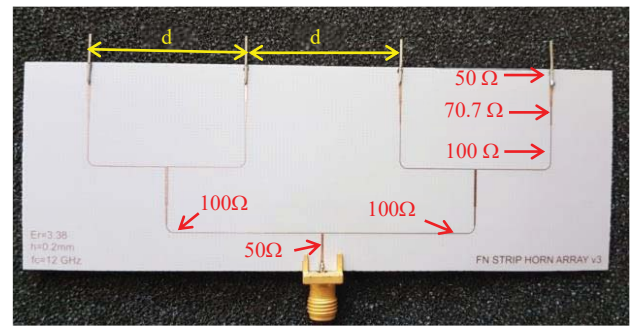
Fig. 3. Photograph of the microstrip feed network for the linear array of strip horn antennas (with indicated impedances).

III. ANALYSIS AND THE SIMULATED RESULTS

In the first step, an array composed of four individual strip horns is simulated and analyzed using the program package WIPL-D [12]. The distance between two neighboring horns (d) is equal to the odd number of quarter wavelengths at the center frequency f_c. In our case, this distance is chosen to be $5\lambda_0/4$. Next, the same array of radiating elements is placed into the common reflector and simulated.

Simulated radiation patterns in $x0y$- plane of both models – with individual strip horns and with the common reflector – at the center frequency, are shown in Fig. 4. It can be seen that the compact array model has a slightly lower gain (~17.2 dBi) than the model with separated strip horn antennas. The 3D radiation pattern of the compact array is displayed in Fig. 5.
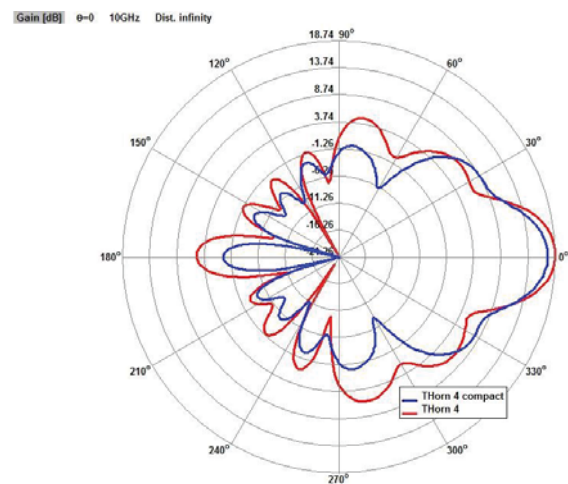
Fig. 4. Radiation patterns of the linear array of strip horn antennas in $x0y$-plane: individual horns (red line) and with the common reflector (blue line).
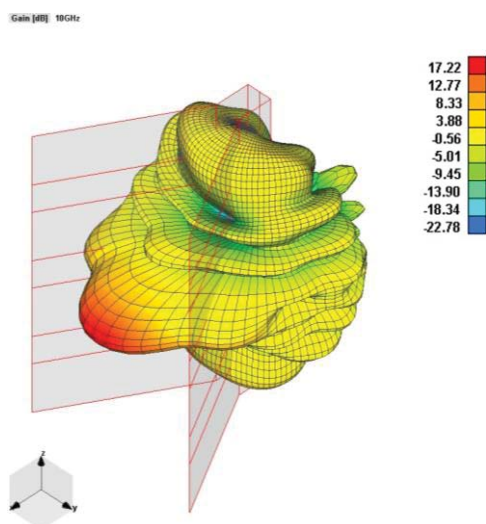
Fig. 5. 3D radiation pattern of the proposed linear array of strip horn antennas.

The measured reflection coefficient S_{11} of the realized antenna array, Fig. 7, is less than −10 dB in about 9% of the bandwidth around the center frequency.
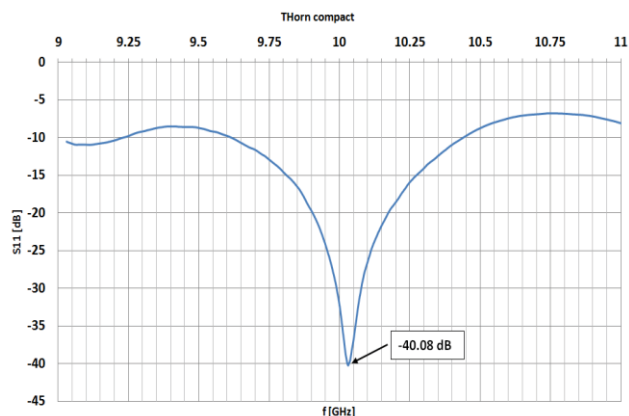
Fig. 7. Measured reflection coefficient S_{11} of the realized strip horn antenna array.

IV. REALIZATION AND THE MEASURED RESULTS

The realization of the complete antenna system which consists of the metallic strip horn with the array of four quarter-wavelength monopole antennas and the feed network is shown in Figs. 6 (a,b).

(a)

(b)

Fig. 6. Photograph of the realized linear array of strip horn antennas: (a) rear view, (b) front view.

The antenna gain measurement is performed using the gain-comparison technique which requires two sets of measurements. The standard gain horn antenna for operation in the range (8.2-12.4) GHz is used as a transmitting antenna. In the first measurement, the test antenna is used as a receiving antenna and the received power is measured. In the second set, the test antenna is replaced by the standard gain horn antenna. The difference between these two sets of measured results gives us the gain of the antenna under test, shown in Fig. 8.
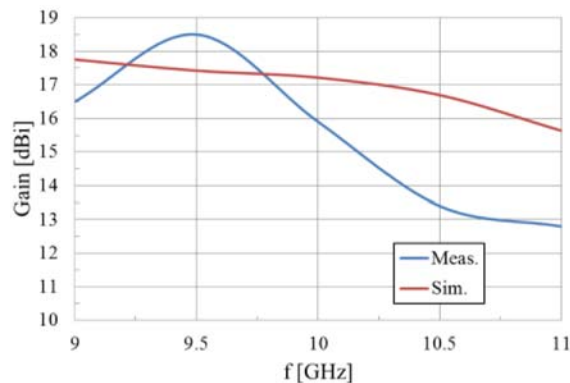
Fig. 8. Measured and simulated gains of the realized strip horn compact antenna array.

The discrepancies between the measured and the simulated array gains, especially at higher frequencies of the range, can be attributed to the insertion loss of the microstrip feed network.

Acknowledgement

References

[1] Momčilo Dragović, "Antene i prostiranje radio talasa," Akademska misao, Beograd, 2008, pp. 96-102.

[2] Constantine A. Balanis, "Antenna theory analysis and design", John Wiley & sons Inc., Hoboken, New Jersey, 2005.

[3] A.D. Grosvenor, P.J.B Clarricoasts, A.A. Kishk and Shafai, "Microwave Horns and Feeds", IEEE press, New York, 1994, pp.149-180.

[4] http://www.waves.utoronto.ca/prof/svhum/ece422/notes/15-arrays2.pdf.

[5] N. Popović and P. Manojlović., "Strip Horn for Operation at 9-15GHz Range," TELSIKS 2017, Niš, 2017, Poster Section.

[6] B.C. Wadell, "Transmission line design handbook," Artech House Inc, 1991, pp. 71-72.

[7] V.V. Petrović, D.V. Tošić i A.N. Đorđević, "Mikrotalasna pasivna kola", Univerzitet u Beogradu - Elektrotehnički fakultet, Beograd, 2010, p.36.

[8] C.A. Grosvenor, R.T. John, S. Canales, B. Davis and J. Veneman, "TEM Horn Antenna design principles", NIST Technical note 1544, Boulder, USA, 2007.

[9] A.R. Mallanzader and F. Karshenas, "Modified TEM Horn Antenna for broadband applications", Progress in Electromagnetics Research, PIER 90, 2009, p.p.105-119.

[10] Y. Wang, G. Fang, H. Su, Y. Ji and X. Zang, "A novel UWB TEM Horn antenna with microstrip-type feed", International Journal of Antennas and Propagation, Vol. 2015, article ID182140, 6 pages.

[11] S. Bassam and J.P. Mohassel, "A Chebyshev tapered TEM Horn antenna", Piers Online, Vol.2, No.6, 2006, pp. 706-709.

[12] www.wipl-d.com

System for Automatically Targeting of Unmanned Aerial Vehicle with Vertical Landing on Mobile Landing Site

Rumen Arnaudov1, Georgi Stanchev2 and Krume Andreev3

Abstract – **The article provides a system for automatically targeting of unmanned aerial vehicles with a vertical landing on a mobile landing site commensurate with its dimensions, based on a pseudo-conical scanning modification used to accompany radar targets.**

Keywords – **Automatic landing of UAV; pseudo-conical scanning; radio-navigation and radiolocation.**

I. INTRODUCTION

The landing of unmanned aerial vehicles (UAVs), and in particular those of a lower category, is carried out under the supervision of a qualified operator. The landing site is usually commensurate with the dimensions of the UAVs and GPS accuracy of the GPS for civilian purposes is not sufficient to target it directly. This is done using a differential GPS, the cost of which is significant [1]. All this complicates their expected massive application for various services in the future. For these reasons scientists and engineers are working to automate this process by using civil GPS data. The following are generally accepted:

- Image processing and recognition of the landing site combined with light signals [2];
- Applying the radiolocation principle of pseudo-conical scanning to accompany objectives [3];
- Applying the principles of lidar optic systems [4].

The solution to automatic landing problems also leads to the next step - landing on a moving landing site. Such landing can be on a vessel, a car and with a lot of reserves and conventions on a train, because over it there is a power supply line, and there are obstacles (trees, poles). At a landing site on a car moving on a relatively straight section at approximately uniform and low speed there would be no problems as long as there were no big trees above the road, the crowns of which are over it. At a landing site on a vessel, a single problem can occur if there is a

[1]Ph.D. Rumen Arnaudov is a Professor at the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ra@tu-sofia.bg .
[2]Ph.D. Georgi Stanchev is an Assoc. Professor at Faculty of Mechanical Engineering, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gstanchev@tu-sofia.bg .
[3]Krume Andreev is a Ph.D. student at the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: andreev.k@abv.bg .

strong water turbulence that requires much more serious software to compensate for it. Therefore, this article proposes a system of landing on a vessel floating in calm waters (lakes, dams, floating rivers and a relatively calm sea).

A solution for automatic landing on a moving vehicle is available in [5], assuming that the UAV has taken off from a landing site that is mounted on a car, has completed a task or mission and is back to the same landing site. It is proposed to use a landing site that is recognizable by a video camera whose image is processed by the droning equipment. There is no mention of what happens if there are crowns of trees above the road, what is the influence of the unevenness of the road, the change of direction of movement.

Video-image processing is a perspective area of development but with some drawbacks to this application. Image quality is dependent on weather conditions, and software is much more complex. If it is also applied to mobile landing sites, the problems are much more due to the dynamics of the incoming image. In this case, it is good to look for an alternative solution.

II. FORMULATION OF THE TASK AND SCHEMATIC SOLUTION

The most appropriate solution for a method and an automatic landing system on a mobile area is the use of the radiolocation principle to accompany targets with conical or pseudo-conical scanning.

This is not the same situation like in the case of radar accompanying targets. Here the object of accompaniment is "own" and not "alien". For this reason, the principle of radiolocation, which works with reflected signals from the monitored "foreign object", disappears. This makes it possible to significantly simplify the automatic landing system, all the more so that the whole process can be digitized. Here, the scanning area is stationary upwards above the landing site. The UAV moves and searches for this area on its assigned GPS coordinates, aiming to position itself above it, level the speed of the vehicle, and then turn on the automatic landing system just above the landing.

With this set-up, signal processing software and hardware implementation are made much easier and more economical.

In Fig. 1 shows the functional scheme of the system and the setting of the individual stages of the landing.
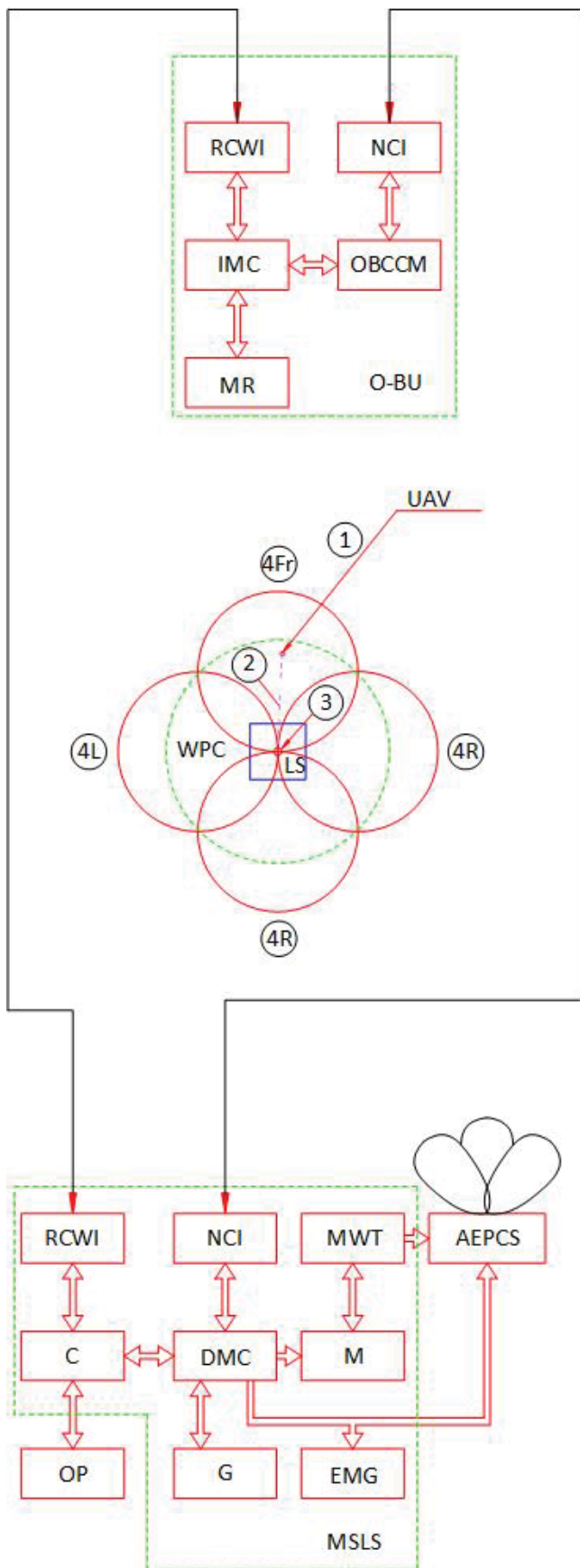
Fig. 1. Figure example

On board of the UAV is an on-board unit (O-BU), which consists of an on-board control-command module (OBCCM), which includes all flight control and control components [6].

Control components are controllers, GPS receivers, altimeters, gyroscopes, barometers, electric motor drivers and others. It is also connected to the Wireless Interface for Remote Communication 1 (RCWI) with the Mobile Station of the landing site (MSLS). Additionally, an information microcontroller (IMC) is installed. The information microcontroller receives information from a microwave receiver (MR), and through a nearby communication interface 1 (NCI) (Wi-Fi, RF or other), it is possible to connect to another.

The pilot-in-command of the UAV guides the flights through the mobile station at the landing site which is located on the mobile object. He is accessing the system via a personal computer (C). The computer (C) connects to the UAV via a Wireless Interface for Remote Communication 2 (RCWI) (GSM, RF or other) and continuously transmits information about its movement and position. The directional microcontroller (DMC) manages the positioning and landing processes. Through a nearby communication interface 2 (NCI) can be established a close communication with the UAV, and by the gyroscope (G) the change in the direction of motion of the mobile object is monitored, via the modulator (M) transmitting data to the microwave transmitter (MWT), the signals from which are emitted by the antenna with Electronic Pseudo-Conical Scanning (AEPCS). The antenna is controlled by the Directional Microcontroller (DMC). After landing, the system also includes electromechanical or electro-magnetic grip (EMG) for the UAV to stay stationary on the landing site (LS).

The landing process consists of three stages: positioning over the pseudo-conical scanning area, signal processing, and positioning of the UAV at the center of pseudo-conical scanning, downhill and landing on the landing site [9].

The landing method on a mobile object is realized by a series of operations, whose interrelationship is also presented in the form of algorithms (Figure 1).

III. A CONDITION FOR SUCCESSFUL POSITIONING OF THE UAV OVER THE LANDING SITE

Figure 2 shows a schematic diagram of the positioning of the UAS over the pseudo-conical scanning antenna. In order not to complicate the figure, only two beams ("left-right") facing each other are pointed upwards. The other two beams ("back and forth") are in the perpendicular plane of these two. The geographic directions are not used here because the object is mobile but the antenna is fixed to it and to the direction of movement as well, which facilitates the movement of the drones that move in the same direction and at the same speed.

The center axis for the entire setting 0Z is in a perpendicular upward direction and pseudo-conical scanning is performed against it. The angle of deflection depends on the accuracy of the GPS module used in the drones. Trough the given maximum permissible errors for civilian purposes of GPS and Galileo in the three coordinates ($\Delta X = 2 - 4$ m, $\Delta Y = 4$ m and $\Delta Z = 8$m) [7, 8], the UAV can be expected to be a part of the air space, formed by a cylinder with a radius of 8 m and a height of 16 m. This area can be called a "wrong positioning cylinder" [3]. To prevent a collision between the drones and the site it is good to have the base of this zone (cylinder) a few meters above it (for example, $\Delta Z = 8$m). This determines the system's maximum operating height H = 3, $\Delta Z = 24$ m and the coordinate Z for positioning of the drones before landing should be 16 m above the landing site. Depending on it, the UAV must fall somewhere in the space inside the wrong positioning cylinder. This cylinder must be into the scanning cone to ensure the successful capture and landing. The cone is formed by the 0M rays.

With this configuration, it is necessary to determine at what angle the four beams pointing up should be dissolved. Figure 2 shows the angle of light dissolution α and the angle of the tip of the inverted cone β, formed by the lower base of the wrong positioning cylinder and the center of the landing site. It can be seen that:

$$\frac{\Delta X}{\Delta Z} = tg \ \frac{\alpha}{2} \ or \ \alpha = 2 \ arctg \ \frac{\Delta X}{\Delta Z} \qquad (1)$$

After taking into account the errors of GPS and Galileo:

$$\alpha = 2 \ arctg \ \frac{4}{8}$$

$$\tag{2}$$

angle β must be less than 60 ° and angle α = 60 °

An additional condition is to reliably accept the signals of the rays in the upper base of the wrong positioning cylinder, which depends on the sensitivity of the microwave receiver (MWT). The level of all signals is increased in the landing process and centering direction. The angle of dissolution of the rays α is better to be larger than the angle β, but unlike the analog conical scan, even if the UAV is from the outside of the beam, it will still orient itself, where it is to fly because it receives coded digital targeting, which is not affected by its position relative to the axis of the beam.

The algorithms to perform the individual stages are autonomous and run sequentially, each ending with an emergency signal to the landing site if problems arise. In this case, its movement must be stopped to ensure a safe landing and, if necessary, stopping the movement of the object and moving to manual control.
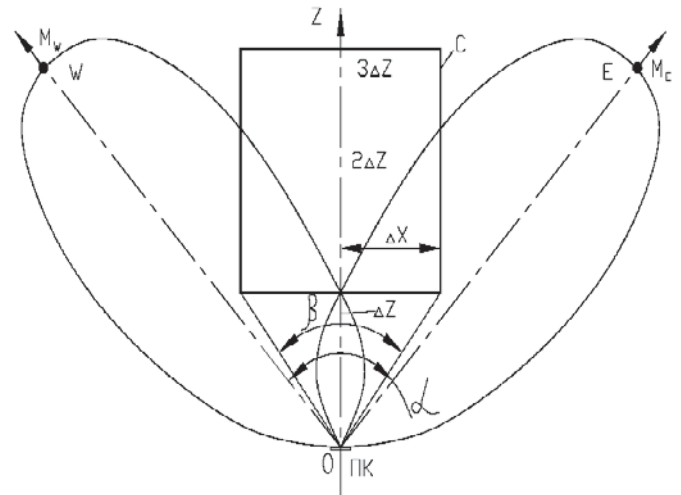
Fig. 2. Diffusion angle α and tip angle of the inverted cone β of the pseudo-conical scanning setup

V. CONCLUSION

The article provides a system for automatically targeting unmanned aerial vehicles with a vertical landing on a mobile site, based on a pseudo-scanning modification used to accompany targets with radar. By pointing up the four rays of pseudo-scanning, a gripping and guiding zone for the unmanned aircraft is formed. It is positioned over GPS coordinates above this area, and landing accuracy is achieved as a result of its pseudo-conical direction by processing the signals emitted from the center of the site.

REFERENCES

[1] https://www.u-blox.com/en/product/c94-m8p#tab-documentation-resources
[2] Fumio Ohtomo, Kazuki Osaragi, Tetsuji Anai, Hitoshi Otani, US Grant Patent number 8630755B2, "Automatic taking-off and landing system", 2014.

IV. ADDITIONAL REMARKS

[3] Andreev K., Stanchev G., "Flight Safety Sensor and Auto-Landing System of Unmanned Aerial System", Int. J. of Reasoning-based Intelligent Systems (IJRIS), Japan, 2019.

[4] Arnaudov R., Botusharov V., Stanchev G., Andreev K., "System for Automatic Targeting of Unmanned Aerial Vehicles in Vertical Landing ", Patent Application No. 112876, Bulgarian Patent Office, February 2019, Sofia.

[5] Fumio Ohtomo, Kazuki Osaragi, Tetsuji Anai, Hitoshi Otani, US Grant Patent number US20120277934A1, "Taking-Off And Landing Target Instrument And Automatic Taking-Off And Landing System", 2011.

[6] Jingxuan Sun, Boyang Li, Yifan Jiang and Chih-yung Wen, "A Camera-Based Target Detection and Positioning UAV System for Search and Rescue (SAR) Purposes", Sensors 2016, 16, 1778; doi:10.3390/s16111778.

[7] Official U.S. government information about the Global Positioning System (GPS) and related topics, https://www.gps.gov/systems/gps/performance/accuracy/

[8] EASA Navipedia, https://gssc.esa.int/navipedia/index.php/Galileo_General_Introduction

[9] Iliev. I, Stanchev G., Andreev. K, "Method and Algorithm for Automatically Targeting of Unmanned Aerial Vehicle with Vertical Landing on Mobile Landing Site", ICEST 2019, Ohrid, Macedonia.

Application of Vilenkin's Additional Theorem in the Calculations of Mutual Coupling Between Circular Apertures on Conducting Sphere

Slavko Rupčić[1], Vanja Mandrić-Radivojević[1] and Nataša Nešić[2]

Abstract – **The paper deal with an application of Vilenkin's additional theorem to the mutual coupling calculations within Moment Method (MoM) between circular apertures on conducting sphere. When calculating the mutual coupling between the apertures one needs to calculate the vector-Legendre transforms of basis and test functions with the domain on the aperture located at an arbitrary position on the sphere. Clasical approach is to numerical calculate the need terms using Euler's formulas for coordinate system rotation. This approach is very time consuming. Instead, significantly accelerating the calculation is achieved by rotation in theta direction in closed form by using of Vilekin' additional theorem for associated Legendre functions.**

Keywords – **spherical antenna, circular aperture, Vilenkin's additional theorem, method of moments .**

I. INTRODUCTION

An array of aperture antennas on the surface of a sphere is of importance because such an array provides wide hemispherical scan coverage with low grating lobe levels.

The mutula coupling of aperture antenna arry is determined in this article by using the Vilenkin's theorem instead of the classic approach to calculations such attachment significantly shortens the calculation time. The computerized routine used is verified by using data from the available literature as well as by using measurements on a derived laboratory model.

II. METHOD OF ANALYSIS

The problem of determining the mutual coupling using the moment method (MoM) in spectral domain for an array circular apertures placed at a spherical ground plane is the problem of calculation equivalent magnetic current placed at different apertures (openings of waveguide Fig.1.). The mutual admittance is given by:

[1]Slavko Rupčić and Vanja Mandrić-Radivojević are with Department of Communications Faculty of Electrical Engineering Osijek 31000, Croatia, E-mails: vanja.mandric@etfos.hr, rupcic@etfos.hr

[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

$$Y_{ij} = -\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} 2\pi S(n,m) r_s^2 \widetilde{\mathbf{M}}_i^T(\mathbf{r},n,-m)\overline{\overline{\widetilde{\mathbf{G}}}}(n,m,r_s|r_s)\widetilde{\mathbf{M}}_j(\mathbf{r},n,m), \quad (1)$$

where $\overline{\overline{\widetilde{\mathbf{G}}}}(n,m,r_s|r_s)$ is a spectral domain dyadic Green's function for grounded spherical surface and $\overline{\overline{\mathbf{L}}}(n,m,\theta)$ is the kernel of the vector-Legendre transformation. $\widetilde{\mathbf{M}}_j(r,n,m)$ and $\widetilde{\mathbf{M}}_i^T(r,n,m)$ are the equivalent magnetic current and the transposed equivalent magnetic current both placed at the i-th and j-th opening of waveguide. This basic/test functions are located at different apertures.

The appropriate spectral-domain Green's function of a multilayer spherical structure is calculated using the G1DMULT algorithm [2].
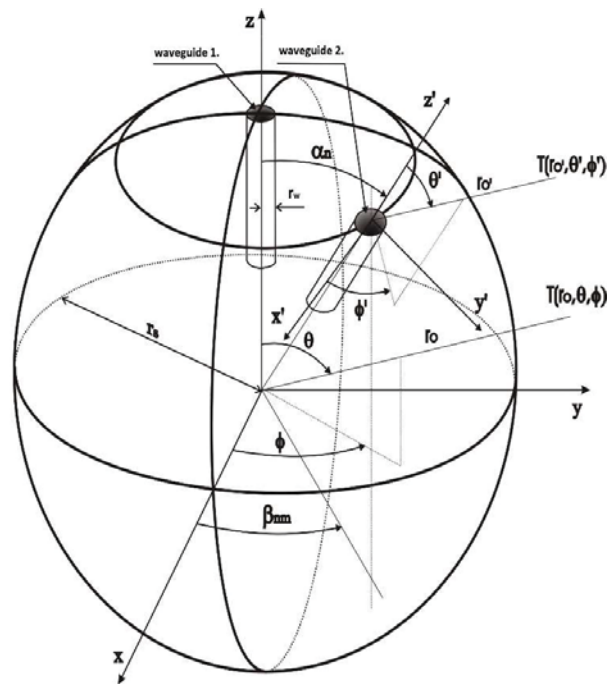
Fig. 1. Circular waveguides (aperture antennas) on a spherical surface – spherical geometry

It is easy to calculate the vector-Legendre transformation of the equivalent current of the waveguide opening located on the

north pole ($\alpha_n = 0$, $\beta_{nm} = 0$). However, when calculating the mutual coupling between the apertures, it is necessary to determine the vector-Legendre transformation of the base and test functions with the domain on the wavegudie aperture of the arbitrary position in the sphere. One way is to numerically determine the required expressions using formulas that link global and local coordinates. The transformation process is based on the vector-Legendre transformation of base and test functions located on the displaced waveguide:

$$\tilde{\mathbf{M}}_\mathbf{j}(r,n,m) = \begin{bmatrix} 0 \\ \tilde{M}_\theta \\ \tilde{M}_\phi \end{bmatrix} = \frac{1}{2\pi\, S(n,m)} \iint\limits_{wav} \tilde{\overline{\mathbf{L}}}(n,m,\theta) \cdot \mathbf{M}_\mathbf{j}(r,\theta',\phi') \cdot \sin\theta' \; e^{-jm\phi'} \; d\theta' \, d\phi' \tag{2}$$

where θ and ϕ are coordinates in the global coordinate system, and θ' and φ' in the local coordinate system. It is important to know that both matrices, $\mathbf{L}$-matrices and basis functions $\mathbf{M}_i$ are written in relation to the basis $\left(\hat{e}_\theta,\hat{e}_\phi,\hat{e}_r\right)$, which is the basis of the global coordinate system. The connection between local and global coordinates is given by the following equations:

$$\cos\theta = -\sin\alpha_n \sin\theta' \cos\phi + \cos\alpha_n \cos\theta' \tag{3a}$$

$$\cot\phi = \frac{\cos\alpha_n \sin\theta' \cos\phi' + \sin\alpha_n \cos\theta'}{\sin\theta' \sin\phi'} \tag{3b}$$

α_n and β_{nm} are the su θ- and ϕ-coordinates of the center of each waveguide in the global system.

This approach is very time consuming since for each basis/tes function one needs to calculate a double-integral of rapidly varying function.

A much more efficient and faster algorithm is the one in which ([6] and [7]) we have the following relationship between the vector Legendre transformations of the basis/test function with domain on the central apertures ($\theta=0$) and on the aperture whose center has coordinate $\theta=\alpha_n$, $\phi=\beta_n=0$. Equivalent magnetic density of the current at the aperture in the spatial coordinates is defined by the equation:

$$\mathbf{M}_k(r_{wav.aperture},\theta,\phi) = \nabla\times(\hat{r}\, r_{wav.aperture}\, M_k^A(r_{wav.aperture},\theta,\phi))$$
$$+ \nabla\cdot(r_{wav.aperture}\, M_k^\psi(r_{wav.aperture},\theta,\phi)) \tag{4}$$

In the spherical coordinate system of equation (4) we can write:

$$\mathbf{M}_k(r_{wav.aperture},\theta,\phi) = \hat{\theta}\frac{1}{\sin\theta}\frac{\partial M_k^A}{\partial\phi} - \hat{\phi}\frac{\partial M_k^A}{\partial\theta} + \hat{\theta}\frac{\partial M_k^\psi}{\partial\theta} + \hat{\phi}\frac{1}{\sin\theta}\frac{\partial M_k^\psi}{\partial\phi} \tag{5}$$

Equations (4) and (5) are described with two functions M_k^A and M_k^ψ.

In addition, base and test functions can be written by using direct inverse vector Legendre transformation as:

$$\mathbf{M}_k(r,\theta,\phi) = \begin{bmatrix} 0 \\ M_\theta \\ M_\phi \end{bmatrix} = \sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} e^{j\omega t}\Bigg\{\hat{\theta}\Bigg[\frac{\partial P_n^{|m|}(\cos\theta)}{\partial\theta}M_{k\theta}(r_{wav.aperture},n,m) - \frac{jm P_n^{|m|}(\cos\theta)}{\sin\theta}M_{k\phi}(r_{wav.aperture},n,m)\Bigg]$$
$$+ \hat{\phi}\Bigg[\frac{jm P_n^{|m|}(\cos\theta)}{\sin\theta}M_{k\theta}(r_{wav.aperture},n,m) + \frac{\partial P_n^{|m|}(\cos\theta)}{\partial\theta}M_{k\phi}(r_{wav.aperture},n,m)\Bigg]\Bigg\} \tag{6}$$

By comparing the relation (5) and (6), it is apparent that the functions M_k^A and M_k^ψ have the following form:

$$M_k^A(r_{wav.aperture},n,m) = -M_{k\phi}(r_{wav.aperture},n,m)\, P_n^{|m|}(\cos\theta)\, e^{jm\phi},$$
$$M_k^\psi(r_{wav.aperture},n,m) = M_{k\theta}(r_{wav.aperture},n,m)\, P_n^{|m|}(\cos\theta)\, e^{jm\phi} \tag{7}$$

Partial derivations are:

$$\frac{\partial M_k^A}{\partial\phi} = -jm\, M_{k\phi}(r_{wav.aperture},n,m)\, P_n^{|m|}(\cos\theta)\, e^{jm\phi} \tag{8a}$$

$$\frac{\partial M_k^A}{\partial\theta} = -M_{k\phi}(r_{wav.aperture},n,m)\frac{\partial P_n^{|m|}(\cos\theta)}{\partial\theta}e^{jm\phi} \tag{8b}$$

$$\frac{\partial M_k^\psi}{\partial\phi} = jm\, M_{k\theta}(r_{wav.aperture},n,m)\, P_n^{|m|}(\cos\theta)\, e^{jm\phi} \tag{8c}$$

$$\frac{\partial M_k^\psi}{\partial\theta} = M_{k\theta}(r_{wav.aperture},n,m)\frac{\partial P_n^{|m|}(\cos\theta)}{\partial\theta}e^{jm\phi} \tag{8d}$$

Furthermore, theta and phi components of equivalent currents can be written by the equations:

$$M_\theta = \sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} e^{j\omega t}\Bigg\{\Bigg[\frac{\partial P_n^{|m|}(\cos\theta)}{\partial\theta}M_{k\theta}(r_{wav.aperture},n,m) - \frac{jm P_n^{|m|}(\cos\theta)}{\sin\theta}M_{k\phi}(r_{wav.aperture},n,m)\Bigg]\Bigg\}e^{jm\phi} \tag{9}$$

$$M_\phi = \sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} e^{j\omega t}\Bigg\{\Bigg[\frac{jm P_n^{|m|}(\cos\theta)}{\sin\theta}M_{k\theta}(r_{wav.aperture},n,m) + \frac{\partial P_n^{|m|}(\cos\theta)}{\partial\theta}M_{k\phi}(r_{wav.aperture},n,m)\Bigg]\Bigg\}e^{jm\phi} \tag{10}$$

Or shorter written:

$$M_\theta = \sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} e^{j\omega t}\left\{\frac{1}{\sin\theta}\frac{\partial M_k^A}{\partial\phi} + \frac{\partial M_k^\psi}{\partial\theta}\right\} \tag{11a}$$

$$M_\phi = \sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} e^{j\omega t}\left\{-\frac{\partial M_k^A}{\partial\theta} + \frac{1}{\sin\theta}\frac{\partial M_k^\psi}{\partial\phi}\right\} \tag{11b}$$

The selected view is similar to the display of the electric field via vector and scalar potentials. It is also important to note that

these relationships do not depend on the coordinate system, ie they are valid in the global and local coordinate system. It is possible to connect the equivalent currents M_k^A and M_k^ψ and in different coordinate systems using the following rule of Legendre functions:

$$e^{-jm\phi}P_n^{|m|}(\cos\theta) = j^m\left[\frac{(n+m)!}{(n-m)!}\right]^{1/2}\sum_{k=-n}^{n}j^{-k}\left[\frac{(n-k)!}{(n+k)!}\right]^{1/2}P_{m,k}^n(\cos\theta_{12})P_n^{|k|}(\cos\theta')e^{-jk\phi'}$$

(13)

or

$$e^{jm\phi}P_n^{|m|}(\cos\theta) = \sum_{m=-n}^{n}(-j)^{|m|-|k|}\sqrt{\frac{(n+|m|)!}{(n-|m|)!}}\sqrt{\frac{(n-|k|)!}{(n+|k|)!}}P_{|m|,sign(m)\cdot k}^n(\cos\theta_{12})P_n^{|k|}(\cos\theta')e^{jk\phi'}$$

(14)

The function $P_{m,k}^n(\cos\theta_{12})$ is defined in [1], and θ_{12} is the angle between the global and the local coordinate system (it is important to note the following: local coordinates have a dash mark on the exponent's site, and the ϕ coordinate of the center of the shifted aperture is equal to zero).

By connecting the relation (7) and (14), the terms for functions M_k^A and M_k^ψ with the domain on the displaced wavelength can be determined:

$$M_k^A(r_{wav.aperture},n,k)$$
$$= -P_n^{|k|}(\cos\theta')\,e^{jk\phi'}\sum_{m=-n}^{n}(-j)^{|m|-|k|}\sqrt{\frac{(n+|m|)!}{(n-|m|)!}}\sqrt{\frac{(n-|k|)!}{(n+|k|)!}}P_{|m|,sign(m)\cdot k}^n(\cos\theta_{12})M_{k\phi}(r_{wav.aperture},n,m),$$

$$M_k^\psi(r_{wav.aperture},n,k)$$
$$= P_n^{|k|}(\cos\theta')\,e^{jk\phi'}\sum_{m=-n}^{n}(-j)^{|m|-|k|}\sqrt{\frac{(n+|m|)!}{(n-|m|)!}}\sqrt{\frac{(n-|k|)!}{(n+|k|)!}}P_{|m|,sign(m)\cdot k}^n(\cos\theta_{12})M_{k\theta}(r_{wav.aperture},n,m).$$

(15) and (16)

Furthermore, by comparing the relation (15) and (16) with the reaction (7), the following is obtained:

$$M_k^A(r_{wav.aperture},n,k) = -P_n^{|k|}(\cos\theta')\,e^{jk\phi'}M_{k\phi_S}(r_{wav.aperture},n,k),$$

(17)

$$M_k^\psi(r_{wav.aperture},n,k) = P_n^{|k|}(\cos\theta')\,e^{jk\phi'}M_{k\theta_S}(r_{wav.aperture},n,k).$$

(18)

Since we are interested in the theta and phi components of the equivalent currents in the spectral domain when we write routine for the computation of programs (programs) we define them in terms of:

$$M_{k\phi_S}(r_{wav.aperture},n,k) = \sum_{m=-n}^{n}(-j)^{|m|-|k|}\sqrt{\frac{(n+|m|)!}{(n-|m|)!}}\sqrt{\frac{(n-|k|)!}{(n+|k|)!}}P_{|m|,sign(m)\cdot k}^n(\cos\theta_{12})M_{k\phi}(r_{wav.aperture},n,m),$$

(19)

$$M_{k\theta_S}(r_{wav.aperture},n,k) = \sum_{m=-n}^{n}(-j)^{|m|-|k|}\sqrt{\frac{(n+|m|)!}{(n-|m|)!}}\sqrt{\frac{(n-|k|)!}{(n+|k|)!}}P_{|m|,sign(m)\cdot k}^n(\cos\theta_{12})M_{k\theta}(r_{wav.aperture},n,m).$$

(20)

The base and test functions on the displaced waveguide are easily determined by using the relation (5), (11) and (12).

If the wavelength center has a φ-coordinate different from zero, it is easy to express the vector-Legendre transformation of base / test functions using the following Fourier series rule $\tilde{M}_{i1}(n,m) = \tilde{M}_{i2}(n,m)e^{jm(\beta_1-\beta_2)}$, ie. it is possible to connect the vector-Legendre transformation of base / test function of different waveguides with the same θ- coordinates. koordinatama.

It is equally possible to determine the equivalent magnetic currents at the opening of each waveguide and calculate the mutual coupling of the aperture antenna on the spherical surface using equation (1).

III. DISCUSSION OF NUMERICAL AND EXPERIMENTAL RESULTS

By using the above-described procedure, the calculation of the mutual admittance of the aperture antennas on the spherical surface of three different radii was made. The figures (Figures 2, 3, 4 and 5) also show the values of the mutual admittance from the literature ([3] and [4]). It is noticeable to match the results from the literature and the calculated results.
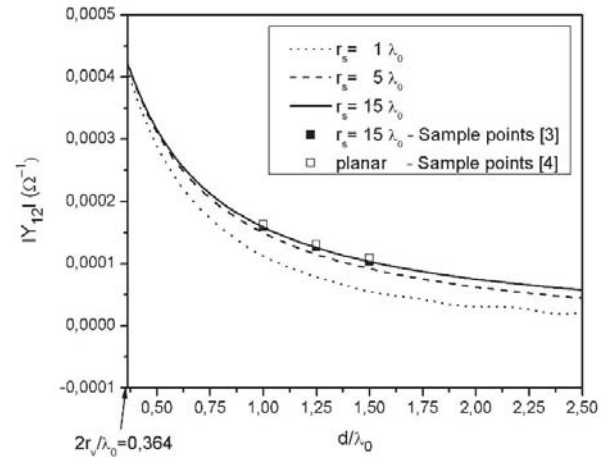
Fig. 2. The mutual admittance of circular apertures on spherical surfaces for different radius of spherical surface – E plane
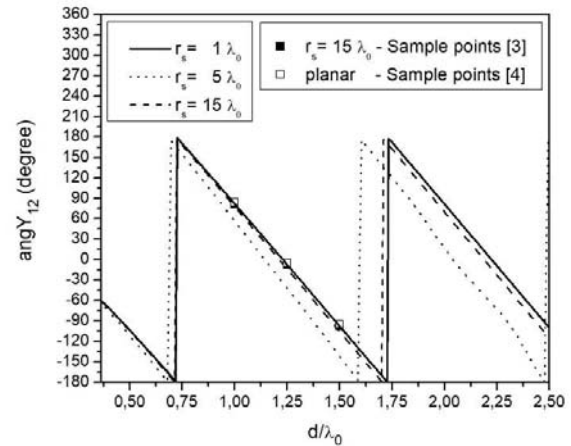
Fig. 3. The mutual admittance angle of circular apertures on spherical surfaces for different radius of spherical surface – E plane
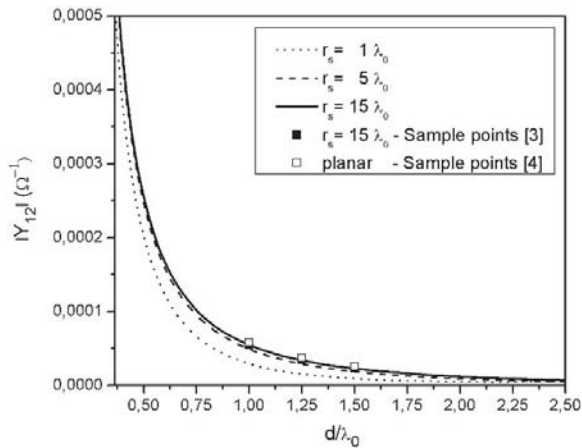
Fig. 4. The mutual admittance of circular apertures on spherical surfaces versus distance between apertures centars for different radius of spherical surface – H plane
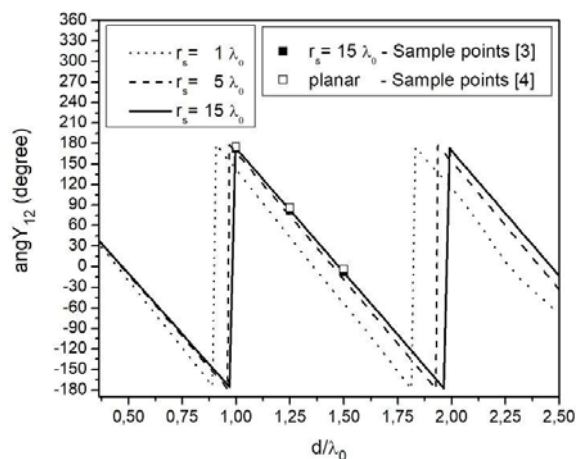
Fig. 5. The mutual admittance angle of circular apertures on spherical surfaces for different radius of spherical surface – H plane

Verification of the procedure and its application to the radiation problem of the antenna aperture on the spherical surface is made with another comparison as shown in Fig. 6. The radius of the circular apertures is 1.905 cm, and the distance between them is 6.35 cm.

As the reference curves, the curves for the planar case were used according to Lit 5. It can be seen that by increasing the spherical radius the results of the mutual coupling approach the planar results to Lit 5.

Furthermore, a model antenna array model was created on a aluminium spherical surface with radius of 30 cm radius. Radius of the circular waveguides are 6 cm (Fig 7.). The mutual coupling of the openings were measured, the first being positioned in the north pole ($\alpha_1=0$ deg and $\beta_1 = 0$ deg) while

the second in the position of $\alpha_2=56$ deg and $\beta_2 = 180$ deg. Figure 8. shows an excellent match between measurement results and those obtained by simulation.
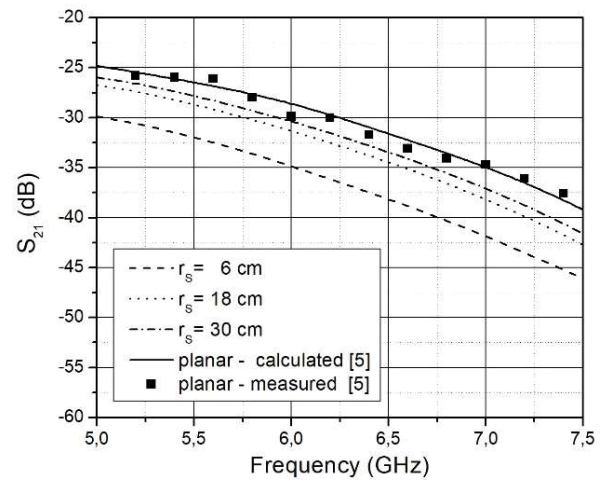
Fig. 6. Calculated magnitude of S_{21} parameter of two circular apertures as a function of frequency for different radius of spherical surface and also calculated and measured magnitude of S_{21} parameter of two circular apertures on the planar surface – H plane [6]

Fig. 7. Photo of the developed laboratory model of aperture antenna array: W1 - $\alpha_1=0$ deg and $\beta_1 = 0$ deg; W2 - $\alpha_2=56$ deg and $\beta_2 = 180$ deg; $r_w= 6cm$; $r_s=30$ cm.
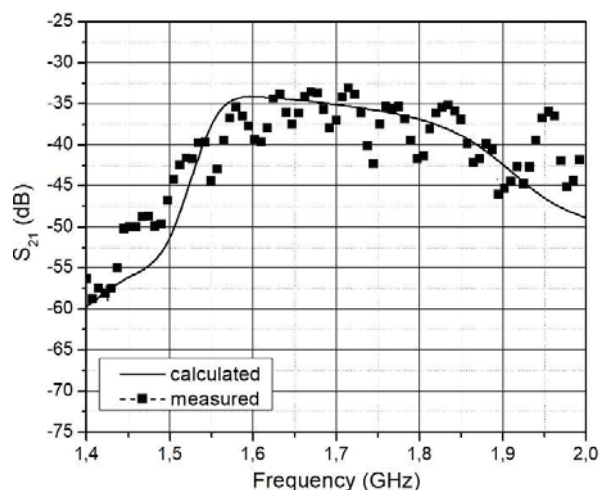
Fig. 8. Calculated and measured magnitude of S_{21} parameter of two circular apretures (radius=6cm) on spherical surfaces with radius=30 cm - E plane [6]

IV. CONCLUSION

The calculation of the mutual coupling (admittance) of the circular aperture of the antenna located on the spherical substrate using the moment method is considerably accelerated by using the Additional theorem for spherical harmonics. This Vilenkin's theorem was applied when calculating equivalent magnetic currents at the analyzed openings. Conversion of the coordinates and integration for returning to the spatial domain has been omitted, which is why the acceleration of this calculation procedure is compared with the classical procedure. The results of this calculation are compared with the results from available literature, and compared to the performed laboratory model. The results show excellent agreement with comparative results.

REFERENCES

[1] N.Ja.Vilenkin, "Special Functions and the Theory of Group Representation", American Mathematical Society, Providence, 1968.

[2] Z.Sipus, P.S.Kildal, R.Leijon, M.Johansson, "An Algorithm for Calculating Green's functions for planar, circular czlindrical and spherical multilazer substrates@, Applied Computational Electromagnetics Societz Journal, Vol. 13, pp. 243/245, 1998.

[3] A.Hesel, Y.L.Liu, and J.Shmoys, "Mutual admittance between circular apertures on a large conducting sphere!, Radio Science, Vol.14, pp. 35 -41, January, 1979.

[4] M. C. Bailey, "Analysis of finite-size phased arrays of circular waveguide elements", NASA TR R-408, National Aerunautics and Space Administration, Washington, DC.

[5] M. C. Bailey, C.W.Bostian, "Mutual Coupling in a Finite Planar Array of Circular Apertures ", IEEE Transactions on Antennas and Propagation, Vol. AP-22, No.2, pp. 178-184, March. 1974.

[6] S.Rupcic, "Circular Waveguide Antenna Arrays on Spherical Structures" PhD Thesis, University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, 2009.

[7] Z.Sipus, S.Skokic, "Applications of Vilenkin's additional theorem in the analysis of spherical antennas and periodic structures", First European Conference on Antennas and Propagation EUCAP, Proceedinsg of EUCAP, 6-10 November, 2006, Nice, France

[8] Z.Sipus, M.Bosiljevac, S.Skokic, "Analysis of Conformal Waveguide Antennas Using Hybrid Spectral Domain UTD-Method, Proceedings of JINA 2004 International Symposium on Antennas / Guiraud, Jean-Louis, editor(s). Nica : JINA, pp. 92-93, 2004..

Use of infrared radiometry in temperature measurement of plant leaf

Hristo Hristov[1], Kalin Dimitrov[2] and Stanyo Kolev[3]

Abstract – **Through our present work we will show the importance of infrared radiometry in conducting various plant studies. We will look at the factors that affect temperature measurements and their significance. We will draw conclusions about the significance of the distance between the thermal camera and the object of study.**

Keywords – **infrared radiometry, infrared thermography, agriculture, solid angle**

I. INTRODUCTION

Infrared thermography (IRT) plays an important role in some methods of assessing the condition of different vegetative tissues. It is a key tool in developing a non-invasive analysis of plant metabolism. It also helps diagnose and monitor various phytopathogenic processes. The use of modern sensor matrices allows the thermographic images to be produced at high resolution, allowing for early diagnosis of plant diseases when the affected tissue is a small percentage and tissue contamination levels are low. Infrared thermography is used in methods to test the viability of seedlings [1-8].

Remote sensing and thermal imaging analysis are methods of collecting, processing and interpreting data without physical contact between the measuring device and the object so that it can be analyzed repeatedly and harmlessly [9-12]. All surrounding objects, thanks to their own temperature, emit infrared (IR) radiation. Thermal imaging is performed with detectors sensitive to a wavelength of 8 to 12µm. The image we see represents variations in the temperature of the leaves. IR thermography is used to plan irrigation, to evaluate plant-pathogen interactions through models of surface temperature monitoring of the leaves and to fully monitor their interaction with the environment [13-18]. The lack of need for contact with the object of research makes it particularly suitable for remote monitoring and data collection. It is also preferred as it is as harmless as possible to plants and the environment.

IRT is used in studying plant temperature stress processes, their adaptation, their ability to acclimate, their endurance and survival in cold weather conditions.

Various pests and diseases hinder growth and reduce crop

[1]Hristo Hristov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: hristoveu@gmail.com

[2]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.

[3]Stanyo Kolev, is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: skolev@tu-sofia.bg.

yields. They can change the temperature and water balance of the plants. The thermophagus could be used to monitor and create patterns of various diseases or pest infestations and detect them before the occurrence of visible symptoms. It works well in field conditions to locate places where culture is more affected and therefore requires more urgent intervention [19,20]. It also allows for the monitoring of the content of different nutrients in crops. In this way, the processes of fertilization and soil incorporation of the ingredients necessary for the development of the plants are also correct and conducted at the right time. Reducing or increasing the water content of the soil causes a change in the leaf temperature. Thermal images can be used to improve the irrigation mode. It is known that the over-irrigation of different fruit crops can reduce the content of sugar and other substances, hence the quality of production. Lower soil moisture content than required may result in a decrease in yield quality and quantity. In conclusion, proper management of the limited water resource is of utmost importance. This contributes to the development of various specific irrigation methods.

When designing any radiometric system, detection is essential. Important parameters are sensitivity, signal noise ratio and visualization based on temperature differences. In order to achieve high accuracy in measurements and visualization, special techniques for recognition, processing, various types of corrections and optimizations are used.

II. THEORY

The aim is to establish, record and visualize changes in the surface temperature of the examined bodies. This information is contained in the energy characteristics of their radiation, in the energy characteristics of other sources found in the system and in the characteristics of the distribution medium.

The radiometric approach is suitable for the study of the spread of thermal radiation and its interaction with objects in the distribution area when the preliminary specification of its complete coherence and the absence of the effects of interference and diffraction are not essential for the systems under study.

When we conduct measurements in real situations from the standpoint of physics and mathematics, it is suitable to apply the radiometric approach. The propagation of the thermal radiation is carried out in scattering, absorption and emitting environment. During the establishment of the propagation equation, we have to take into account all influencing factors [7,10,12].

Due to the fact that the point of view is generally not perpendicular to the emitting surface, we will use a dimension representing the surface and angle density of the emitted flux, that is, its radiance [11,12]

$$L = \left(d^4\Phi\right)/\left(d\,\Omega dA\cos\theta\right).\qquad(1)$$

Since we will perform measurements in a precisely defined frequency range, we will use the spectral density of the same magnitude

$$L_\lambda = dL/d\lambda.\qquad(2)$$

To define the process, we will use the radiation energy output of the elementary flux at a given point from the radiating surface in the half space 2π

$$M = \left(d^2\Phi\right)_{2\pi}/dA.\qquad(3)$$

As we are interested in a certain part of the wave spectrum, we will use the spectral density of the radiation energy output

$$M_\lambda = \left(d^3\Phi\right)_{2\pi}/(dAd\lambda).\qquad(4)$$

When we need to do research in the infrared spectrum, we always base Planck's law on the emission of a black body.

In nature, the bodies differ in their radiance from that of the black body. This requires the introduction of a correction factor called emissivity coefficient

$$M_\lambda(\lambda,T) = \varepsilon(\lambda,T)M_{blackb.}(\lambda,T).\qquad(5)$$

Let us consider a real radiometric system for temperature measurements. We will monitor the interactions and energy transformations of the radiations with atmosphere located in the volume between the studied object and the radiometer. The set-up is shown in principle in Fig. 1.

In the volume enclosed by the viewing angle towards the radiometer, a radiation from the studied object enters. With
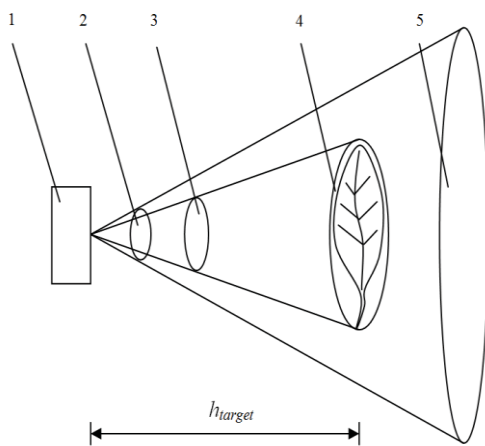
Fig. 1. General set-up of a plant leaf radiometric investigation 1-radiometric device; 2-solid view angle of the radiometric device; 3-solid view angle of the plant leaf investigated; 4-essential area of the plant leaf surface which produces radiation (target); 5-radiation from the atmosphere, h_{target} - distance to the target

such studies, the scattering and absorption in the atmospheric channel are of utmost importance. Also, the thermal radiation of the atmosphere itself is important, as part of this radiation is aimed at the camera and added up to the radiation of the studied object, because their directions coincide. In the volume between the object and the camera, it is also possible for other radiations to enter. These are distributed in different directions. Some of them can come from independent external sources, others could be the result of multiple scattering of the aerosol particles and the molecules of the different gases. Generally, when we describe the energy interactions within a system, we should take into account all influencing factors and the expression will be the following

$$dL_\lambda/dz = -\left(\alpha^{(s)}+\alpha^{(a)}\right)L_\lambda + \alpha^{(a)}L_{\lambda,blackb.} +$$
$$+\left(\alpha^{(s)}/4\pi\right)\int_{4\pi} L'_\lambda P_A d\,\Omega' \quad,\qquad(7)$$

where $\alpha^{(s)}$ is a scattering coefficient of atmosphere, $\alpha^{(a)}$ is an absorption coefficient of atmosphere, and the last addend of the formula

$$\left(\alpha^{(s)}/4\pi\right)\int_{4\pi} L'_\lambda P_A d\,\Omega'; ,\qquad(8)$$

shows the scattering by other sources. In our case, the distance from the object to the camera is ignorably small and the neighbouring volumes have the same temperature, therefore this ingredient does not exert any practical impact and can be removed [11,12]. Due to the earlier clarification that the distance from the studied object to the radiometer is small enough, which creates homogeneity of the atmospheric channel, apart from using one particular wavelength within the spectrum (8-12µm), we also assume that the source of radiation is isotropic, as a result of all these conditions, we can assume that the coefficients of atmospheric scattering and absorption are constant in relation to the space and spectrum. In order to perform a numerical calculation of the flux entering the radiometer aperture, we need to solve the following expression

$$\Phi_r = A_r\Omega_r\tau_r\left(\varepsilon_t \exp\left(-\left(\alpha^{(s)}+\alpha^{(a)}\right)Z\right)\right)F(T_t,\lambda_1,\lambda_2)+$$
$$+ A_r\Omega_r\tau_r\left(\alpha^{(a)}/\left(\alpha^{(s)}+\alpha^{(a)}\right)\right)\left(1-\exp\left(-\left(\alpha^{(s)}+\alpha^{(a)}\right)Z\right)\right)F(T_t,\lambda_1,\lambda_2)$$
$$(9)$$

where A_r is the area of the receiving part of the antenna, Ω_r is the spatial angle of view. The last multiplier in the two addends is the value obtained after the integration of the spectral radiance of a black body at a given temperature in the range of λ_1 to λ_2.

We can see that the flow is formed by two main components, the radiation from the object and the radiation from the atmosphere.

III. Numerical investigation

The aim is to establish, record and visualize changes in the surface temperature of the examined bodies. This information is contained in the energy characteristics of their radiation, in the energy characteristics of other sources found in the system and in the characteristics of the distribution medium [10,14].

Let us look at a specific example of radiometric investigation for plant leaf, using the theory written above. Spatial viewing angle of the radiometric sensor is fixed. We will see how changing the distance from the radiometric device to the investigated object will affect the heat flows in the system.

For the simulation process we choose the following parameters:

$Ta=295K$, $Tt=293K$, $\lambda1=8\mu m$, $\lambda2=12\mu m$, $\varepsilon t=0,98$, $\varepsilon_{atm}=0,96$,

$\alpha^{(s)} = 0,03 km^{-1}$, $\alpha^{(a)} = 0,4 km^{-1}$, $\Omega_r = 6.10^{-3} sr$,

$A_r = 8.10^{-5} m^2$, $\tau_r = 1$

Using Scilab we calculate the values of (9) for different distances [17]. Part of the results are shown in the following Table I.

TABLE I
PART OF SIMULATION DATA

No	h_{target}[m]	Φt [W],T=293K	Φa [W],T=295K
1	0.25	1.40752E-05	1.43632E-09
2	0.5	1.40737E-05	2.87249E-09
3	0.75	1.40722E-05	4.30851E-09
4	1	1.40707E-05	5.74436E-09
5	1.25	1.40692E-05	7.18007E-09
6	1.5	1.40676E-05	8.61562E-09
7	1.75	1.40661E-05	1.0051E-08
8	2	1.40646E-05	1.14863E-08
9	2.25	1.40631E-05	1.29213E-08
10	2.5	1.40616E-05	1.43563E-08

It is very important to compare the flow coming from the target with the flow coming from the atmosphere and how they change with the change of the distance from the camera to the investigated object.

IV. Conclusion

Clearly, as the distance in these near-boundaries increases, the flow of the site slowly decreases and the flow from the atmosphere increases smoothly. This is explained by the increase in the radiant volume and shows in practice the increase in the atmospheric channel losses. As the distance increases, the flow of the atmosphere increases much faster than the decrease in the flow from the site but is thousands of times less than it. The linearity in the change in flow values indicates that atmospheric scatter and absorption coefficients are space and spectral constants for the given conditions. At

these distances from the radiometric apparatus to the object, the horizontal profiles of the atmospheric scattering and absorption coefficients remain unchanged.

Tracking the proper development of crops requires research to be carried out in their natural environment, in the field or in the greenhouse. This will allow for any deviations that may indicate a problem, to seek the cause immediately and to react as quickly as possible. Recently used infrared thermography remote sensing techniques provide more accurate data for the characterization of plant parameters. They can cover huge areas over a long period of time. The reflectivity of plants depends on the properties of their leaves and their orientation and structure. The amount of energy reflected for a particular wavelength depends on the color of the leaves, their geometry, the composition of the cells and the amount of water in them. These several factors determine the infrared characteristics of the plants. There are big differences and sharp boundaries in these properties. Their study and knowledge makes it possible to improve the methods for remote monitoring and, respectively, to improve the management of plant processes. Technological advances in the manufacture of radiometric detectors give additional impetus to the use of thermography as an indispensable tool in the monitoring, diagnostics and management of plant cultures.

References

[1] I. Kranner, G. Kastberger, M. Hartbauer, and H. W. Pritchard, „Noninvasive diagnosis of seed viability using infrared thermography", Proc Natl Acad Sci U S A. 2010 February 23; 107(8): 3912–3917, Published online 2010 February 3

[2] M. Lindenthal, U. Steiner, H.-W. Dehne, and E.-C. Oerke, „Effect of Downy Mildew Development on Transpiration of Cucumber Leaves Visualized by Digital Infrared Thermography", Phytopathology 95:233-240, Institute for Plant Diseases, University of Bonn, Nussallee 9, D-53115 Bonn, Germany, Accepted for publication 19 October 2004.

[3] R. Ishimwe, K. Abutaleb, F. Ahmed, „Aplications of Thermal Imaging in Agriculture – A Review", Scientific Research, Advances in Remote Sensing, 2014, 3, 128-140

[4] M. Wisniewskil and M. Fuller, „Ice nucleation and deep supercooling in plants: new insights using infrared thermography",

[5] J. M. Costa, O. M. Grant and M. M. Chaves, „Thermography to explore plant–environment interactions", Journal of Experimental Botany, Vol. 64, No. 13, pp. 3937–3949, 2013

[6] M. Schlessinger, *Infrared Technology Fundamentals*, CRC Press, 1994.

[7] J. Miller, *Principles of Infrared Technology: A Practical Guide to the State of the Art*, Springer Science & Business Media, 2012.

[8] M. Diakides, J. Bronzino, D. Peterson, *Medical Infrared Imaging: Principles and Practices*, CRC Press, 2012.

[9] G. Zibordi, C. Donlon, A. Parr, *Optical Radiometry for Ocean Climate Measurements*, Academic Press, 2014.

[10] Z. Zhang, B. Tsai, G. Machin, *Radiometric Temperature Measurements: II. Applications*, Academic Press, 2009.

[11]   W. Wolfe, *Introduction to Radiometry*, SPIE Press, 1998.

[12]   E.Ferdinandov, Basics of optoelectronics, Technika, 1993 (in Bulgarian).

[13]   M. Schlessinger, *Infrared Technology Fundamentals*, CRC Press, 1994.

[14]   J. Miller, *Principles of Infrared Technology: A Practical Guide to the State of the Art*, Springer Science & Business Media, 2012.

[15]   E. Dereniak, G. Boreman, *Infrared detectors and systems*, Wiley, 1996.

[16]   A. López, F. D. Molina-Aiz, D. L. Valera, A. Pena, „Determining the emissivity of the leaves of nine horticultural crops by means of infrared thermography", Elsevier, Scientia Horticulturae 137 (2012) 49–58

[17]   C. Gomez, *Engineering and Scientific Computing with Scilab*, Springer Science & Business Media, 2012.

[18]   O. M. Grant, Ł. Tronina, H. G. Jones and M. M. Chaves, „Exploring thermal imaging variables for the detection of stress responses in grapevine under different irrigation regimes", Journal of Experimental Botany, Vol. 58, No. 4, pp. 815–825, 2007

[19]   H. G. Jones, „Infrared estimations of leaf or canopy temperature", Prometheus Wiki, 24.05.2011

[20]   I. Romm and B. Cukurel, "Quantitative image fusion in infrared radiometry", 5 April 2018, IOP, Measurement Science and Technology, Volume 29, Number 5

VHF Chebyshev Low Pass Filter with Lumped Elements

Veljko Crnadak[1] and Siniša Tasić[2]

Abstract – **In this paper, the practical design of the VHF Chebyshev low pass filter with lumped elements is presented. The magnitude response of the filter is approximated with the Chebyshev polynomial of the 9th order, resulting in a ladder network, consisting of five series inductors and four shunt capacitors. The passband ripple of the filter is 0.02 dB and the cutoff frequency is 220 MHz. The ladder network, with the ideal reactive elements, is then converted to the real low pass filter with the real lossy reactive elements, which is then simulated in 3D EM simulator. The values of the S_{11} and S_{21} parameters of the simulated filter are then being compared, to the values of the same S-parameters of the produced filter, in order to test the design process. The low pass filter is designed to have minimal insertion loss and maximal return loss, for the frequency range from 150 MHz to 200 MHz. Purpose of the filter is to reduce or eliminate harmonics, at the output of a high-power amplifier**

Keywords – **capacitor, Chebyshev, coil, filter, inductor, low pass, lumped, VHF.**

I. INTRODUCTION

The magnitude square of the Chebyshev low pass filter transfer function of the Nth order is,

$$|H(j\omega)|^2 = \frac{H_0}{1+\varepsilon^2 T_N^2(\frac{\omega}{\omega_c})}, \tag{1}$$

where H_0 is the dc attenuation, ε is the ripple magnitude, ω_c is the cutoff frequency and $T_N(\frac{\omega}{\omega_c})$ is the Chebyshev polynomial of the Nth order,

$$T_N\left(\frac{\omega}{\omega_c}\right) = \begin{cases} cos[N\,cos^{-1}\left(\frac{\omega}{\omega_c}\right)], 0 \le \omega \le \omega_c \\ cosh[N\,cosh^{-1}\left(\frac{\omega}{\omega_c}\right)], \omega > \omega_c \end{cases}. \tag{2}$$

The Chebyshev response, shown in Fig. 1, oscillates in the passband between the two values, H_0 and $\frac{H_0}{1+\varepsilon^2}$, while in the stopband it approaches zero at infinity. The Chebyshev low pass filter is also known as all pole filter, because the transfer function has zeros only at infinity [1]. The poles of the

[1] Veljko Crnadak is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia, and with the Company for Microwave and Millimeter-Wave Techniques and Electronics IMTEL- Komunikacije Joint-Stock Company Belgrade, Bulevar Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, E-mail: veljko@insimtel.com.
[2] Siniša Tasić is with the Company for Microwave and Millimeter-Wave Techniques and Electronics IMTEL- Komunikacije Joint-Stock Company Belgrade, Bulevar Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, E-mail: tasa@insimtel.com.

transfer function are all on the left half of the s-plane, and are given with the following expression,

$$s_k = \omega_c(\sigma_k + j\omega_k)\,, k = 1, 2, ..., N, \tag{3}$$

where

$$\sigma_k = -\sinh[\frac{1}{N}\sinh^{-1}\left(\frac{1}{\varepsilon}\right)]\sin\frac{(2k-1)\pi}{2N}, \tag{4}$$

and
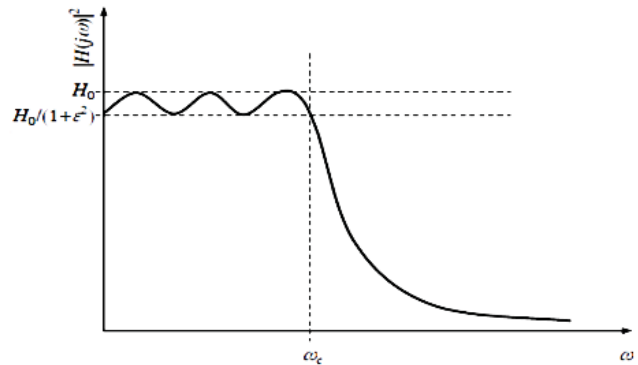
Fig. 1. The Chebyshev low pass response of the 6th order [1].

$$\omega_k = cosh[\frac{1}{N}sinh^{-1}\left(\frac{1}{\varepsilon}\right)]cos\frac{(2k-1)\pi}{2N}. \tag{5}$$

II. DESIGN THEORY

Our objective is to design a low pass Chebyshev filter that has a passband ripple of $A = 10\log(1 + \varepsilon^2) = 0.02$ dB, the cutoff frequency of $f_c = 220$ MHz and an insertion loss at $f = 300$ MHz of at least $L = 35$ dB. Number of reactive elements in the filter is determined from the following inequality,

$$N \ge \frac{cosh^{-1}\sqrt{\frac{10^{0.1L}-1}{10^{0.1A}-1}}}{cosh^{-1}\left(\frac{f}{f_c}\right)}. \tag{6}$$

From the inequality (6), we learn that the filter has to be of the 9th order at least. Because the source and the load resistance are of 50 Ω each, Fig. 2, the dc attenuation of the filter is $H_0 = 1$. Before determining the values of the ideal reactive elements of the circuit in Fig. 2, we must obtain the values of the elements of the Chebyshev low pass prototype. The values of the prototype inductances are, $g_1 = g_9 = 0.9021, g_3 = g_7 = 1.88$ and $g_5 = 1.972$, the values of the

prototype capacitances are, $g_2 = g_8 = 1.4518$ and $g_4 = g_6 = 1.7053$, and the values of the prototype resistances are, $g_0 = g_{10} = 1$. The values of the prototype elements were obtained using the recursion formulas [2].

The actual values of the filter inductances are,

$$L_k = \frac{R_0 g_{(2k-1)}}{\omega_c}, k = 1, 2, 3, \tag{7}$$

and the actual values of the filter capacitances are,

$$C_k = \frac{g_{2k}}{R_0 \omega_c}, k = 1, 2, \tag{8}$$

where $R_0 = 50 \ \Omega$ and $\omega_c = 2\pi f_c$.

We see from the formula (7), that there are three different inductances in the filter, as it is shown in Fig. 2, $L_1 = 32.63$ nH, $L_2 = 68.0026$ nH and $L_3 = 71.3123$ nH.

We see from the expression (8), that there are two different capacitances in the filter, as it is shown in Fig. 2, $C_1 = 21.0056$ pF and $C_2 = 24.6733$ pF.
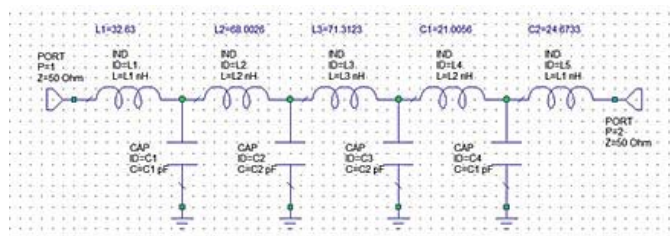
Fig. 2. The schematic of the low pass filter circuit.

During the filter realization, all the inductors were designed with the help of formula [3],

$$L = \frac{0.394 r^2 N^2}{9r + 10l}, \tag{9}$$

where r is the coil radius in cm, l is the coil length in cm, N is the number of turns and L is the inductance in µH.
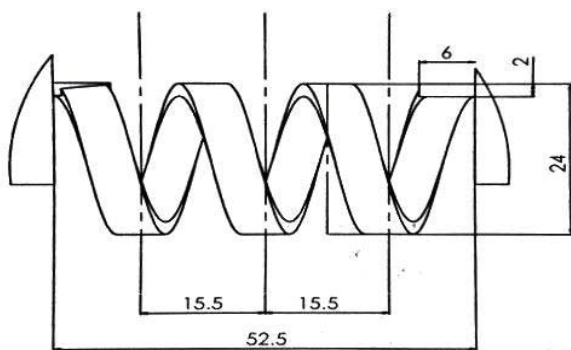
Fig. 3. The schematic of a coil, with the inductance L_A.

Every ideal inductor in Fig. 2, with the inductance L_2, was substituted with the coil in Fig. 3, with the inductance L_A, in order to realize the filter. The coil in Fig. 3, has three turns and was designed as a copper strip of thickness $t = 2$ mm, of radius $r = 10$ mm, of length $l = 52.5$ mm and of width $w = 6$ mm. The inductance of a coil in Fig. 3, according to formula (9), is $L_A = 57.6585$ nH.

The ideal inductor in Fig. 2, with the inductance L_3, was substituted with the coil in Fig. 4, with the inductance L_B, in order to realize the filter. The coil in Fig. 4, has three turns and was designed as a copper strip of thickness $t = 2$ mm, of radius $r = 10$ mm, of length $l = 44.3$ mm and of width $w = 5$ mm. The inductance of a coil in Fig. 4, according to formula (9), is $L_B = 66.5291$ nH.
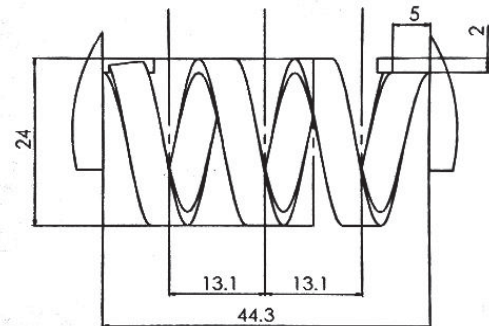
Fig. 4. The schematic of a coil, with the inductance L_B.

Every ideal inductor in Fig. 2, with the inductance L_1, was substituted with the coil in Fig. 5, with the inductance L_C, in order to realize the filter. The coil in Fig. 5, has two turns and was designed as a copper strip of thickness $t = 2$ mm, of radius $r = 9$ mm, of length $l = 30.45$ mm and of width $w = 6$ mm. The inductance of a coil in Fig. 5, according to formula (9), is $L_C = 33.1144$ nH. All the dimensions in Figs. 3, 4 and 5 are in mm.
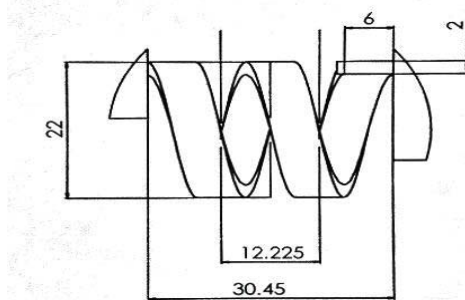
Fig. 5. The schematic of a coil, with the inductance L_C.

In Fig. 6, looking from left to right, we can see the substitutions for the ideal capacitors C_1 and C_2 from Fig. 2. Every ideal capacitor C_1, in Fig. 2, was substituted with the parallel connection of two capacitors, Fig. 6, with the equivalent capacitance of $C_1' = 19.2281$ pF. Every ideal capacitor C_2, in Fig. 2, was substituted with the parallel connection of two capacitors, Fig. 6, with the equivalent capacitance of $C_2' = 20.9374$ pF.
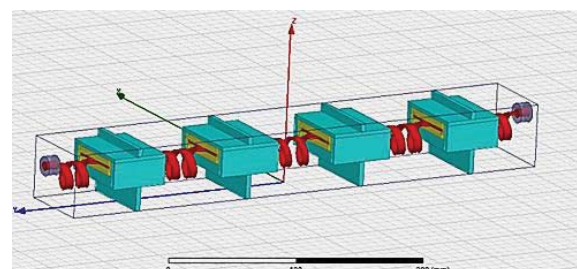
Fig. 6. The isometric view of the model of the VHF low pass filter in 3D EM simulator

III. Simulation And Measurement Results

A. Simulation

Being aware of the criteria that our filter has to meet, the return loss, $-20\log|S_{11}|$, larger than 20 dB, and the insertion loss, $-20\log|S_{21}|$, smaller than 0.1 dB, for the frequency range from 150 MHz to 200 MHz, we were able to direct the process of 3D EM simulation towards satisfying the above-mentioned conditions and obtaining the final dimensions of the filter, shown in Fig. 6. Frequencies below 150 MHz are of no interest to us because they were reduced or eliminated by previous filtering stages of the system. The inductors dimensions were firstly determined through the application of formula (9) and were later corrected through 3D EM simulation. The capacitors dimensions were firstly obtained through the application of a well-known formula for the capacitance of parallel-plate capacitors:

$$C = \varepsilon_0 \varepsilon_r \frac{S}{d} \qquad (10)$$

where ε_0 is the vacuum permittivity, ε_r is the relative permittivity of the dielectric, S is the area of the plates, and d is the distance between the plates. In our case dielectric is Teflon, so $\varepsilon_r = 2.1$. The capacitors dimensions were later corrected through 3D EM simulation. As we can see from Fig. 7, capacitors in a parallel connection that form equivalent capacitor C_1', share the same copper plate (colored in red), of the dimensions 45 mm × 46.2 mm × 2 mm, that is surrounded by Teflon in an aluminum housing. In Fig. 7, Teflon is colored in yellow and aluminum is colored in blue.
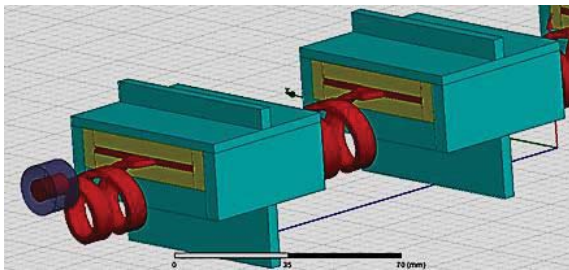
Fig. 7. The model of the capacitors C_1' and C_2' in 3D EM simulator.

As we can see from Fig. 7, capacitors in a parallel connection that form equivalent capacitor C_2', share the same

Fig. 8. The produced VHF low pass filter.

copper plate (colored in red), of the dimensions 45 mm × 50.8 mm × 2 mm, that is surrounded by Teflon in an aluminum housing. The distance d between the plates, for capacitors in a parallel connection that form both C_1' and C_2', is the same and equal to 5.45 mm. The photograph of the produced filter is given in Fig. 8.

Because of the edge effect, the capacitances C_1' and C_2' were calculated using the Method of Moments. The inductors and the capacitors are placed in the aluminum casing, filled with air. The length of the casing is 404.2 mm, the height of the casing is 51.4 mm and the width of the casing is 65 mm. All the dimensions in Fig. 9 are in mm.
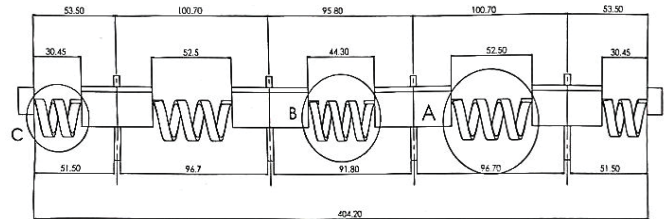
Fig. 9. The schematic of the lateral dimensions of the VHF low pass filter.

The filter uses two 7/16-connectors.

B. Simulated and measured results

As we can see from Fig. 10, the return loss is larger than 21 dB in the frequency range from 150 MHz to 200 MHz, which is satisfactory. Unfortunately, the insertion loss exceeds 0.1 dB, in the same frequency range, due to losses in conductors and dielectric. The highest insertion loss in the frequency range of interest is 0.16 dB.
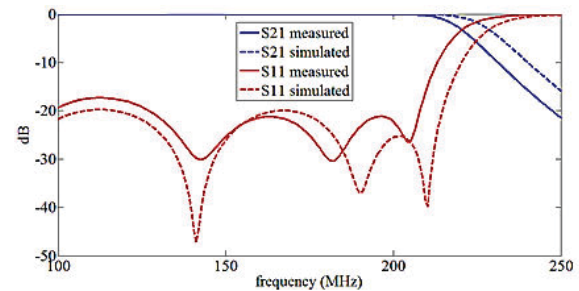
Fig. 10. Comparison of simulated and measured magnitudes of S_{21} and S_{11}.
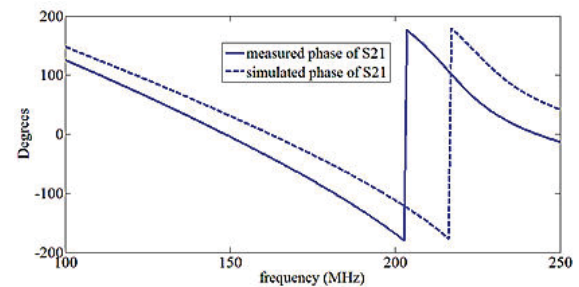
Fig. 11. Comparison of simulated and measured phase angle of S_{21}.
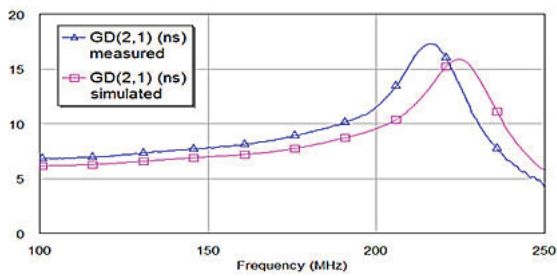
Fig. 12. Comparison of simulated and measured group delay.
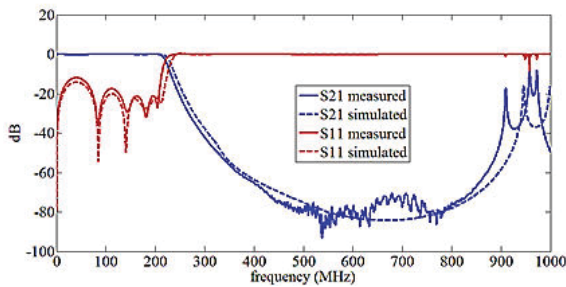
Fig. 13. Comparison of simulated and measured magnitudes of S_{21} and S_{11}.
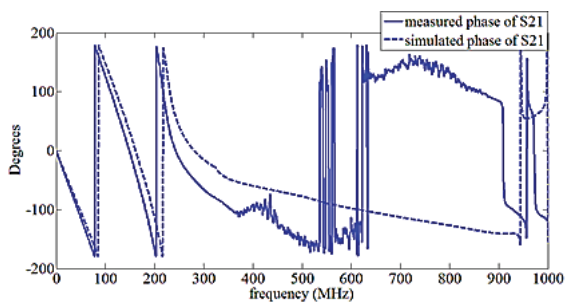
Fig. 14. Comparison of simulated and measured phase angle of S_{21}
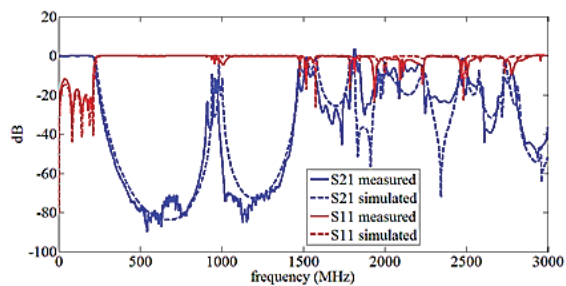
Fig. 15. Comparison of simulated and measured magnitudes of S_{21} and S_{11}.

It can be seen from Fig. 11, that the phase response of the filter is nonlinear, which has for a consequence a nonconstant group delay, as it is shown in Fig. 12.

These are all the typical traits of a Chebyshev low pass filter. Ideally, the filter is a lossless two-port network, which means that the zeros of the reflection coefficient are located at the frequencies, $f_k = f_c \cos\frac{(2k-1)\pi}{2N}$, that is $f_k = 0$, 75.24 MHz, 141.41 MHz, 190.52 MHz, and 216.66 MHz. We can see from Fig. 13, that there are five zeros of the reflection
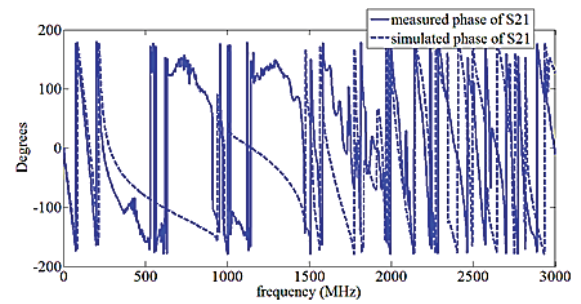
Fig. 16. Comparison of simulated and measured phase angle of S_{21}.

coefficient and they are located around above-mentioned frequencies, which is the sign of good filter design. The insertion loss at 300 MHz is 42.01 dB, as it is shown in Fig. 13. Nonlinearity of the filter phase response only increases with frequency, as can be seen in Figs. 14 and 16. Real filters have re-entry modes that limit the high-frequency capability of the filter [4]. It can be seen in Figs. 13 and 15, that higher frequency signals can appear at the output of the filter.

IV. CONCLUSION

In this paper, we have thoroughly explained the theory behind and the design of the VHF Chebyshev low pass filter with lumped elements. Our primary objective was to design the low pass filter that can be used for high power applications, more precisely at the output of a high power amplifier, where it would reduce or eliminate harmonics. In such an application, the nonlinearity of the phase response is not important, the only thing that matters is that the transition between passband and stopband be sharp enough, so for that reason, the Chebyshev low pass response was chosen. The bulky dimensions of the filter are a consequence of the purpose for which the filter was built. The goals regarding the level of the return loss in the passband and the level of the insertion loss in the stopband were met. Only the level of the insertion loss in the passband has exceeded the projected value by a small margin, at the first place because of the losses introduced by real inductors, and at the second place because of the losses introduced by Teflon.

ACKNOWLEDGEMENT

REFERENCES

[1] N. Kinayman and M. I. Aksun, "Modern Microwave Circuits", 1st ed., ARTECH HOUSE, 2005, pp. 415–437.
[2] D. K. Misra, "Radio-Frequency and Microwave Communication Circuits", 1st ed., John Wiley & Sons, 2001, pp. 322.
[3] C. Bowick, "RF Circuit Design", Newnes, 1997, pp. 17.
[4] https://www.markimicrowave.com/blog/engineering/resources/rf-and-microwave-knowledge-base/#filters.

Throughput Performance of MU-MIMO-OFDM with Optimal Pair-wise Algorithm under Imperfect CSI

Aleksandra Panajotović[1], Nikola Sekulović[2] and Daniela Milović[3]

Abstract – **In this paper, a multiuser multiple-input multiple-output orthogonal frequency division multiplexing (MU-MIMO-OFDM) system with zero-forcing beamforming (ZFBF) precoding applying optimal pair-wise semi-orthogonal user selection (SUS) algorithm is considered. The knowledge of perfect channel state information at transmitter (CSIT) is required to exploit full benefit of that system, but it is the ideal case. We analyse real scenario in which an imperfect CSIT affects throughput performance of system compliant with IEEE 802.11ac. Extensive simulation results are presented to support our analysis.**

Keywords – **CSI, IEEE 802.11ac, MU-MIMO-OFDM, User scheduling algorithm, ZFBF.**

I. Introduction

To satisfy demands of broadband wireless communication market for channel capacity, higher and higher from day to day, Wi-Fi standard IEEE 802.11ac suggests multiple-input multiple-output orthogonal frequency division multiplexing (MU-MIMO-OFDM) as technique enabling speeds ranging from 500 Mbps up to several Gbps [1]. In order to achieve those speeds it is necessary to tackle with beamforming (precoding), user selection and power allocation.

Beamforming increases performance of wireless network focusing signal towards selected user and reduces in that way multi-access interferences. Capacity achieving precoding technique is dirty paper coding (DPC) [2]. Although optimal, DPC is impractical because of a tremendous computational complexity at both side (transmit and receive) even for moderate number of users. Therefore, more practical linear (zero-forcing beamforming (ZFBF) and block diagonalization (BD)) or nonlinear (Tomlinson-Harashima) precoding techniques should be applied [3, 4].

To materialize the huge potential that MU-MIMO brings, in addition to precoding, access point (AP) has to select a group of users which should be served in that time slot. 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, design of suboptimal user selection algorithm is very important issue. Semi-orthogonal user selection (SUS) algorithm in combination with ZFBF gives performance reasonably close to that of DPC under practical value of N_u [5]. Modification of that algorithm to be applicable in IEEE 802.11ac system is presented in [6-8] as generalized multicarrier SUS (GMSUS) algorithm. In order to realize better performance than one achieved with GMSUS algorithm, in [9] authors propose the optimal pair-wise SUS algorithm.

To achieve benefit arising from precoding and user selection it is required to know channel state information at transmitter (CSIT). Unfortunately, in practice CSIT is imperfect. Its quality depends on quantization effects and/or delays. Imperfect CSIT, besides precoder design and selection of user group, also influences on link adaptation (transmission mode selection) causing throughput and error performance drops. In [10], authors present general framework to evaluate the performance of various linear multicarrier MU-MIMO schemas taking into account the accuracy of the channel information feedback to the AP.

In this paper we investigate how imperfect CSIT, expressed in number of bits using in quantization process of channel information, influences on MU-MIMO system performance. We suppose that ZFBF precoding technique is applied, since it cancels both inter- and intra-user interferences. According to results presented in [9], we choose the optimal pair-wise SUS as algorithm for user selection.

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, I_L is $L \times L$ identity matrix, $\|a\|$ represents the Euclidian norm of a vector 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. In downlink scenario with large number of users, the AP serves users with favourable channel conditions. At the given scheduling period AP conveys information to a subset

[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.

$Q = \{u_1, \ldots, u_{|Q|}\}$, $Q \subset N_u$, of selected physical users with receiving L_{u_i} spatial streams. Following inequality should be satisfied $L_Q \triangleq \sum_{i=1}^{|Q|} L_{u_i} \le N_T$. Let $\mathbf{H}_{u_i}[q] \in C^{N_R \times N_T}$, $i = \overline{1, N_u}$, be MIMO propagation channel between AP and u_i-th user over q-th subcarrier. Singular value decomposition (SVD) applied on $\mathbf{H}_{u_i}[q]$ results into

$$\mathbf{H}_{u_i}[q] = \mathbf{U}_{u_i}[q]\mathbf{\Sigma}_{u_i}[q]\mathbf{V}_{u_i}^H[q], \tag{1}$$

where $\mathbf{U}_{u_i}[q] = \left[\mathbf{u}_{u_i,1}[q], \ldots, \mathbf{u}_{u_i,N_R}[q]\right] \in C^{N_R \times N_R}$ and $\mathbf{V}_{u_i}[q] = \left[\mathbf{v}_{u_i,1}[q], \ldots, \mathbf{v}_{u_i,N_T}[q]\right] \in C^{N_T \times N_T}$ are unitary matrices containing the left and the right singular vectors of $\mathbf{H}_{u_i}[q]$ and $\mathbf{\Sigma}_{u_i}[q] = D\left(\sigma_{u_i,1}[q], \ldots, \sigma_{u_i,N_R}[q]\right) \in R^{N_R \times N_T}$ is a diagonal matrix which elements on the main diagonal are singular values of $\mathbf{H}_{u_i}[q]$. In order to eliminate inter-user interference, ZFBF post-processing is applied at receiver, so now the equivalent channel matrix corresponding to selected spatial streams of u_i-th user on q-th subcarrier is defined as

$$\tilde{\mathbf{H}}_{u_i}[q] = \tilde{\mathbf{U}}_{u_i}^H[q]\tilde{\mathbf{U}}_{u_i}[q]\tilde{\mathbf{\Sigma}}_{u_i}[q]\tilde{\mathbf{V}}_{u_i}^H[q] = \tilde{\mathbf{\Sigma}}_{u_i}[q]\tilde{\mathbf{V}}_{u_i}^H[q]. \tag{2}$$

where $\tilde{\mathbf{U}}_{u_i}[q] \in C^{N_T \times L_{u_i}}$ contains left singular vectors associated to L_{u_i} spatial streams. Similar, $\tilde{\mathbf{\Sigma}}_{u_i}[q] \in C^{N_T \times L_{u_i}}$ contains singular values, while $\tilde{\mathbf{V}}_{u_i}[q] \in C^{N_T \times L_{u_i}}$ contains right singular vectors associated to L_{u_i} spatial streams.

In order to eliminate intra-user interferences post-processing at transmitter have to be performed as

$$\mathbf{x}[q] = \mathbf{W}_Q[q]\mathbf{P}_Q^{1/2}[q]\mathbf{s}_Q[q], \tag{3}$$

where $\mathbf{W}_Q[q] = \left[\mathbf{w}_{u_1}[q] \ldots \mathbf{w}_{u_{|Q|}}[q]\right] \in C^{N_T \times L_Q}$ is ZFBF precoding matrix on subcarrier q, $\mathbf{P}_Q[q] = D\left(\mathbf{P}_{u_1}[q], \ldots, \mathbf{P}_{u_{|Q|}}[q]\right) \in R^{L_Q \times L_Q}$ is power allocation matrix and $\mathbf{s}_Q[q] = \left[\mathbf{s}_{u_1}^T[q] \ldots \mathbf{s}_{u_{|Q|}}^T[q]\right]^T \in C^{L_Q \times 1}$ being the vector which contains the information symbols sent to selected users.

Due to limited feedback, channel between AP and selected physical users can be modelled as $\hat{\mathbf{H}}_Q[q] = \tilde{\mathbf{H}}_Q[q] - \tilde{\mathbf{\Sigma}}_Q[q]\mathbf{E}_Q[q]$, where $\tilde{\mathbf{H}}_Q[q] \triangleq \left[\tilde{\mathbf{H}}_{u_1}^T[q] \cdots \tilde{\mathbf{H}}_{u_{|Q|}}^T[q]\right]^T$. Furthermore, $\hat{\mathbf{H}}_{u_i}[q] \triangleq \tilde{\mathbf{\Sigma}}_{u_i}[q]\hat{\mathbf{V}}_{u_i}^H[q]$, where $\hat{\mathbf{V}}_{u_i}^H[q]$ is quantized version of $\tilde{\mathbf{V}}_{u_i}[q]$, i.e. $\hat{\mathbf{V}}_{u_i}[q] = \tilde{\mathbf{V}}_{u_i}[q] - \mathbf{E}_{u_i}[q]$ and

$\mathbf{E}_Q[q] = \left[\mathbf{E}_{u_1}^T[q] \ldots \mathbf{E}_{u_{|Q|}}^T[q]\right]^T \in C^{L_Q \times N_T}$ is global quantisation noise matrix. Now, for ZFBF precoding, precoding matrix is defined as $\mathbf{W}_Q[q] = \hat{\mathbf{H}}_Q^H[q]\left(\hat{\mathbf{H}}_Q[q]\hat{\mathbf{H}}_Q^H[q]\right)^{-1}$. It yields to following equation for the signal at the output (after postprocessing) of selected user

$$\begin{aligned}\mathbf{y}_{u_i}[q] = &\ \mathbf{P}_{u_i}^{1/2}[q]\mathbf{s}_{u_i}[q] + \tilde{\mathbf{\Sigma}}_{u_i}[q]\mathbf{E}_{u_i}[q]\\ &\times \sum_{j=1}^{|Q|}\mathbf{W}_{u_i}[q]\mathbf{P}_{u_j}^{1/2}[q]\mathbf{s}_{u_j}[q] + \tilde{\mathbf{\eta}}_{u_i}[q],\end{aligned} \tag{4}$$

where $\tilde{\mathbf{\eta}}_{u_i}[q] \sim CN\left(0, \sigma_\eta^2 \mathbf{I}_{L_{u_i}}\right)$. The second term in Eq. (4) represent interference leakage due to imperfect CSI.

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 mutliantenna users should be served. Therefore, design of user scheduling algorithm is important issue in MU-MIMO systems. In [9], authors present the optimal pair-wise SUS algorithm and show its advantage over other SUS-based algorithms for small and medium SNR. The steps of the optimal pair-wise SUS algorithm are:

Step 1: Initialization

$$Q_0 = \{1, \ldots, K\}, \tag{5}$$

$$i = 1. \tag{6}$$

Step 2: Determine the degrees of orthogonality, $\beta_{l,p}$, between all spatial stream pairs $l \ne p$:

$$\beta_{l,p} = \sum_{q=1}^{N_d} \frac{\left|\tilde{\mathbf{h}}_l[q]\tilde{\mathbf{h}}_p^*[p]\right|}{\left\|\tilde{\mathbf{h}}_l[q]\right\|\left\|\tilde{\mathbf{h}}_p[q]\right\|}. \tag{7}$$

where $\tilde{\mathbf{h}}_l[q]$ represents lth row of matrix $\tilde{\mathbf{H}}_Q[q]$.

Step 3: Find the pair P_i from Q_{i-1} with the smallest degree of orthogonality

$$P_i = \{l, p\} = \arg\max_{l,p \in Q_{i-1}} \beta_{l,p}. \tag{8}$$

Step 4: Select i-th spatial strem to be eliminated as follows

$$\pi(i) = \arg\min_{r \in P_i} \sum_{q=1}^{N_d}\left\|\tilde{\mathbf{h}}_r[q]\right\| \tag{9}$$

$$Q_i = \{m \in Q_{i-1} | m \ne \pi(i)\}, \tag{10}$$

$$i \leftarrow i + 1. \tag{11}$$

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 spatial streams selected in Step 4 and find the combination providing maximal throughput.

$$Q_{opt} = \arg \max_{S \subset Q_{i-1}} T(S). \tag{12}$$

In IEEE 802.11ac system, only finite set of transmission modes, which represents a combination of modulation type and coding rate (MCS), is available [8]. Subsequent to selection of the group of users, fast link adaptation (FLA) carries out a procedure to allocate MCS to users in order to maximize system throughput satisfying in the same time 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 many parameters (allocated MCS, the received SNR, packet length and channel realization), derivation and evaluation of analytical expression for PER is almost impossible. Therefore, there are 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$_{eff}$.

Effective SNR is a function of received SNR of l-th spatial stream associated to physical user u_i which can be expressed as [10]

$$\gamma_{u_i,l}[q] = \frac{P_{u_i,l}[q]}{\left[\mathbf{R}_{u_i}[q]\right]_{l,l} + \left[\mathbf{R}_{\eta_{u_i}}[q]\right]_{l,l}}, \tag{13}$$

in environment with limited feedback, where

$$\mathbf{R}_{u_i}[q] = \tilde{\mathbf{\Sigma}}_{u_i}[q]\mathbf{E}_{u_i}[q]\left(\sum_{j=1}^{|Q|}\mathbf{W}_{u_j}[q]\mathbf{P}_{u_j}[q]\mathbf{W}_{u_j}^H[q]\right) \\ \times \mathbf{E}_{u_i}^H[q]\tilde{\mathbf{\Sigma}}_{u_i}^H[q] \tag{14}$$

and $\mathbf{R}_{\eta_{u_i}}[q] = \sigma_n^2[q]\mathbf{I}_{Q_{u_i}}$. It is obvious that incomplete CSIT influence on MCS selection reflecting that influence also on throughput and error.

IV. NUMERICAL RESULTS

In order to show influence of imperfect CSIT on throughput performance of MU-MIMO-OFDM system with optimal pair-wise SUS scheduling algorithm, this section presents the simulation results obtained using parameters from IEEE 802.11ac standard. System operates at 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 and

E 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]. Homogenous scenario in which all users sitting on circumference centered at AP and all experiencing the same average SNR is supposed.

Figures 1 and 2 present the system throughput as a function of average received SNR for MU-MIMO-OFDM in B channel characterizing environment with little-to-moderate frequency selectivity. System throughput is evaluated for different degree of CSIT, i.e. for different number of bits used in quantization process of channel information [13]. That quantized value of estimated channel is fed back to transmitter and further used in beamforming, user selection and FLA. If we analyse results from these two figures, we can conclude that using seven bits in quantization provides enough quality of CSIT, so difference in throughput realized for perfect CSIT and one achieved for seven bits is negligible, even for high SNR. In addition, it is obvious that throughput performance for low SNR is independent on quality of CSIT. So, in that region small number of bits should be used in quantization process of channel information. Even results for 6 bits are not presented here, it is useful to emphasize that evaluated throughput results for that case is not close enough to throughput curve for perfect CSIT. Therefore, quantization with 7 bits remains as the best possible solution. Comparison these results with ones presented in [10] show that somehow optimal pair-wise SUS algorithm is a bit more resistant to imperfect CSIT than other SUS-based algorithms.

In addition, It can clearly be seen from those two figures that having more users in the system leads to higher throughput provoked by multiuser diversity that a larger number of users brings along.
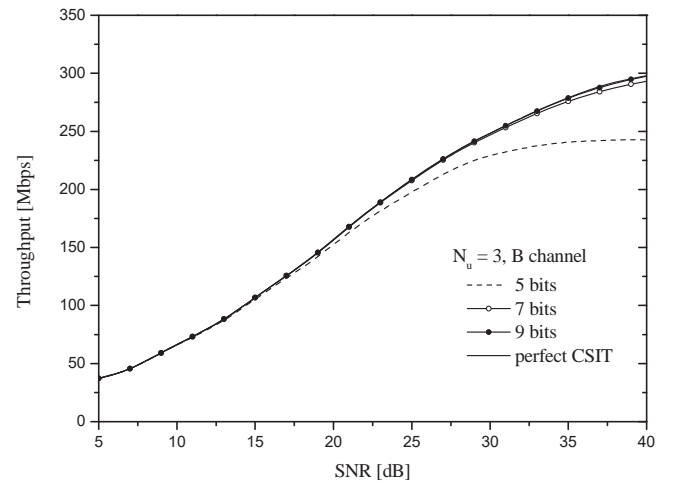
Fig. 1. Throughput for different degree of CSIT for $N_u = 3$ users and B channel

Figure 3 represents throughput of MU-MIMO-OFDM system with imperfect CSIT operating in E channel, i.e. environment with moderate-to-large frequency selectivity. System in such environment is sensitive to imperfect CSIT,

especially for high SNR, when throughput starts to decrease for small number of bits used in quantization process.

quantize channel information, especially for system operating in E channel.
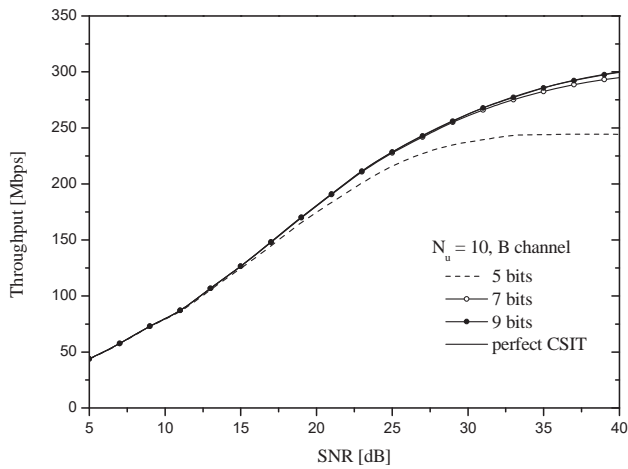
Fig. 2. Throughput for different degree of CSIT for N_u =10 users and B channel
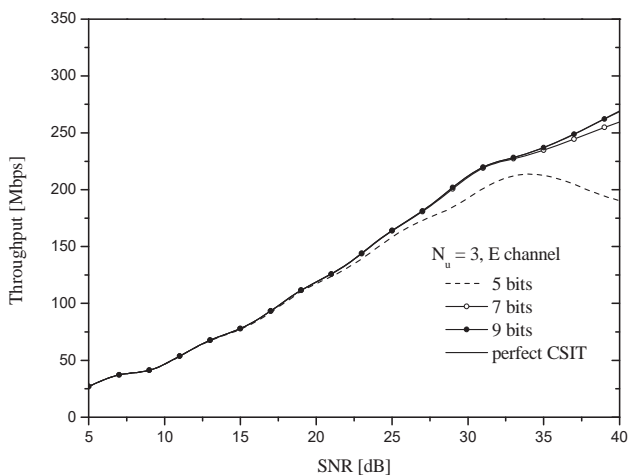
Fig. 3. Throughput for different degree of CSIT for N_u =3 users and E channel

V. CONCLUSION

This paper has presented study of the influence of limited feedback on throughput performance of MU-MIMO-OFDM system complaint with IEEE 802.11ac standard. According to the previously published results which pointed out advantage of optimal pair-wise SUS algorithm over other SUS-based counterparts, that scheduling algorithm has been used.

Numerical results have shown that the most accurate quantisation level which hardly affects system performance with respect to that achieved under a perfect CSIT is given in the form of using 7 bits for quantization. Noticeable throughput degradation has been evident for using 5 bits to

REFERENCES

[1] "802.11ac: The Fifth Generation of Wi-Fi" – Technical White Paper, Cisco public, 2018.
[2] H. Weingarten, Y. Steinberg and S. Shakai, "The Capacity Region of the Gaussian Multiple-input Multiple-output Broadcast Channel," IEEE Trans. Inf. Theory, vol. 52, no. 9, pp. 3936-3964, 2006.
[3] 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.
[4] A. A. D'Amico, "Tomlinson-Harashima Precoding in MIMO Systems: A Unified Approach to Transceiver Optimization Based on Multiplicative Schur-Convexity", IEEE Trans. Sig. Proc., vol. 56, no. 8, pp. 3662-3677, 2008.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] A. Panajotović, N. Sekulović and D. Milović, "Optimal Pair-Wise SUS Scheduling Algorithm for Multiuser MIMO," ICEST 2018, Conference Proceedings, pp. 179-182, Sozopol, Bulgaria, 2018.
[10] A. Panajotovic, F. Rierra-Palou and G. Femenias, "Limited Feedback MU-MIMO-OFDM Systems", WSA Conference Proceedings, 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.
[13] A. Rico-Alvariño and R. W. Heath, "Learning-based Adaptive Transmission for Limited Feedback Multiuser MIMO-OFDM," IEEE Trans. Wireless Commun., vol. 13, no. 7, pp. 3806–3820, 2014.

Techniques for eliminating fading for FSO systems

Vladimir Saso[1], Borivoje Milosevic[2], Srdjan Jovkovic[3]

Abstract – **Paper deals with the implementation of a new technology for the transmission of multimedia content FSO - Free Space Optics, optical wireless transmission. FSO – Optical wireless transmission has this feature to be implemented everywhere quickly, either in terms of flow, or in terms of realization speed. With the convergence of microwaves, optics and network technologies, the expansion of telecommunication networks over wireless links, mostly to remote locations, is becoming a very affordable option for many organizations, medium and large enterprises, as well as cable operators. Optics in a free space, also known as Optical Wireless or Lasercom (Laser Communication), is a technology that reappears using modulated optical rays for establishing short, medium or long range wireless transmission. Traditional linear combinational techniques will be used for fading elimination, that is the Selective Combining (SC) technique, which is based on the selection of a branch in which there is the largest ratio of the mean power of the signal and the noise power, assuming that the power of the noise in all branches is the same; a technique based on equalizing the phases in all branches of the receiver, equal to gain and signal combining from all branches (EGC - Equal Gain Combining) and the so called MRC (Maximum Ratio Combining) technique where phase equalization is performed as well as greater evaluation of stronger diversity signals with the subsequent addition of signals from all branches.**

Keywords – **FSO, Diversity, SC, EGC, MRC, SNR, BER.**

I. INTRODUCTION

Mobile communications have been developing rapidly in recent years, as well as wireless channel models that are used to describe different effects. Laser wave propagation is a very complex phenomenon. If a sufficiently small wavelength of laser waves is assumed, their propagation obtains the form of the expansion of optical beams. Geometric optics separates several basic phenomena of expansion such as diffraction, scattering, transmission, reflection, refraction, and absorption. Diffraction is the bending of waves around an obstacle whose dimensions are significantly larger than the wavelength, which allows the duplication of waves to the receiver even though there is no optical visibility with the transmitter. This effect is also known as the effect of shadow or shadowing. Scattering occurs when the laser wave encounters obstacles whose dimensions are comparable to the wavelength of the laser waves. This is a phenomenon similar to diffraction, except that the laser wave is dispersed in several directions. Therefore, this effect is difficult to predict. Transmission occurs when a laser wave hits an obstacle that is somewhat transparent to the laser wave. This mechanism allows the

existence of laser signals inside buildings. Reflection occurs when a laser wave hits an object that is significantly larger than the wavelength of the incident wave. The reflected wave can increase or decrease the signal at the mobile station. In the center where there are many reflected waves, the receiving signal at one point is usually variable.

These factors, in combination with the others atmospheric turbulences, are responsible for the difference between the transmission and the reception power of the signal. Since there is optical visibility between the receiver and the transmitter, then the component of the signal that crosses this line is far more intense than the components obtained by scattering and therefore it can be described by Rician's distribution. Rician fading occurs when several low-power signals (different reflections) on the receiving antenna, are accompanied by a strong signal (direct wave) -LOS (Line Of Sight propagation conditions). Signal distributions will be observed, also in cases when a signal that can be described by Gamma Gamma distribution and Lognormal distribution is present. The coefficients of signal weakening in the free space environment will be calculated, the SNR and BER will be expressed, and the methods will be proposed to improve the characteristics of the system.

Due to the different refractive index of the atmosphere, the paths of laser wave propagation are curved. As a result, the coverage area is usually higher. The signal strength changes due to the variable refractive index. As there is often no direct visibility between the transmitter and the mobile station, the received signal is the sum of the signals resulting from the above described phenomena. Because of this, the receiving signal is often time and spatial-varying.

II. DIVERSITY TECHNIQUES

Traditional linear combinational techniques will be used, that is the selective combining (SC) technique, which is based on the selection of a branch in which there is the largest ratio of the mean power of the signal and the noise power, assuming that the power of the noise in all branches is the same; a technique based on equalizing the phases in all branches of the receiver, equal to gain and signal combining from all branches (EGC - Equal Gain Combining) and the so called MRC (Maximum Ratio Combining) technique where phase equalization is performed as well as greater evaluation of stronger diversity signals with the subsequent addition of signals from all branches.

Statistical analysis of these processes, as can be seen in the paper, requires the development of mathematical models, their processing, and solution proposal by application of a very complicated mathematical apparatus, so that the basic goal of this paper, based on the obtained results of statistical analysis of the signals in the environment.

[1,3]Vladimir Saso and Srdjan Jovkovic is with the University ALFA BK Belgrade, Serbia, email: vladimir.saso@fepn.edu.rs; srdjansms11@gmail.com
[2]IBorivoje Milosevic is with the University UNION Nikola Tesla, Faculty of Business and Law, Belgrade, Serbia email: borivojemilosevic@yahoo.com

A. MRC - Maximum Ratio Combining

MRC combiners use a linear combination of signals from coherent branches to maximize the output: signal to noise ratio SNR. MRC is the optimum linear multi-system signal combining diversity technique, which provides statistically best results in limiting fading effects. The signal in each of the branching branches is multiplied by an appropriate weight factor, equalizing the phases of all the signals, and with the greater contribution, the branches with the more favorable SNR ratio are taken, Fig. 1. This results in a higher strength signal having a higher weight in, but it is therefore necessary to measure SNR ratio in all branches, which makes this technique a cost-effective set.
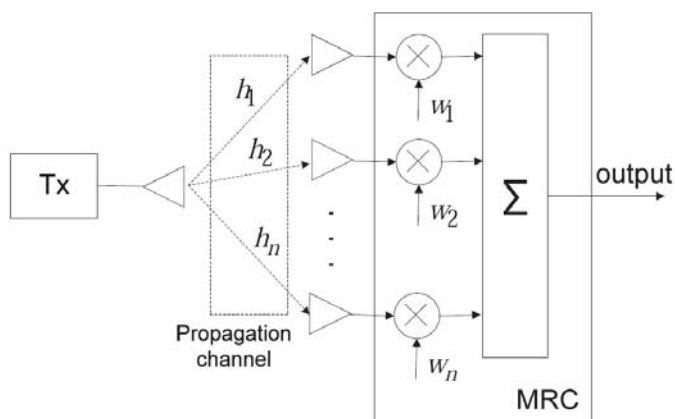
Fig. 1. MRC Combining

The following figure shows the SNR relationship depending on the number of receiving antennas, using the MRC technique, Fig. 2.
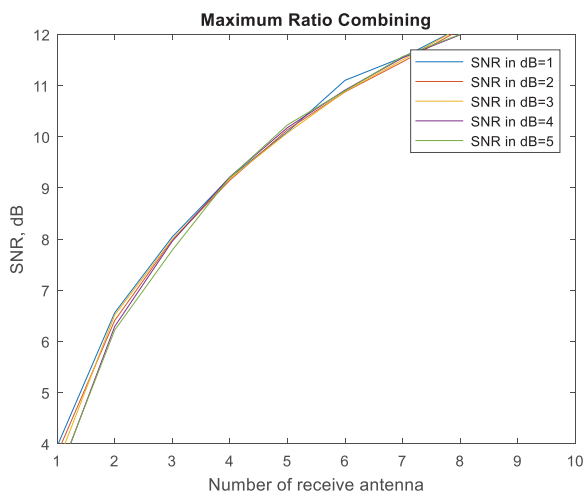
Fig. 2. MRC Combining

B. EGC - Equal Gain Combining

In practice, such a scheme is useful for modulation techniques that have the same symbols energy (e.g., M-PSK) because the output signal is a linear combination of all the diversity branches, where those are in phase and are taken with the same weight. This kind of combination reduces the complexity of the receiver. When applying the EGC signal combining technique, the signal phase change over compensation is performed in all the different branches, so the signals are summed up. Unlike MRC techniques, all addends have the same weight factor, so no SNR measurement and estimation in all the different branches is required, making this technique simpler and cheaper for practical implementation. Prices are somewhat worse in relation to the case of MRC technique. In the simulation model of the EGC receiver, the reception signal is determined as the sum of the signals from the receiving antennas, whereby the phase compensation of the signal at the diversity branches was performed, Fig. 3.
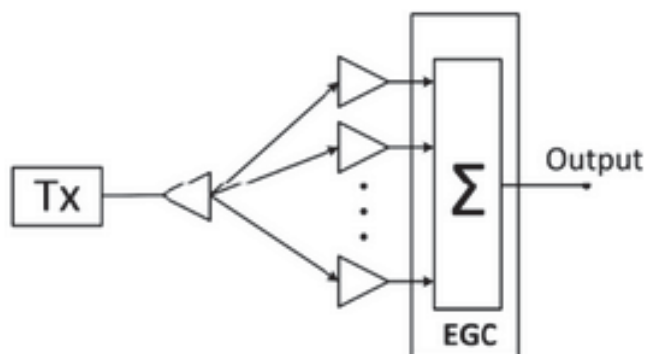
Fig. 3. EGC Combining

The following figure shows the SNR relationship depending on the number of receiving antennas, using the EGC technique, Fig. 4.
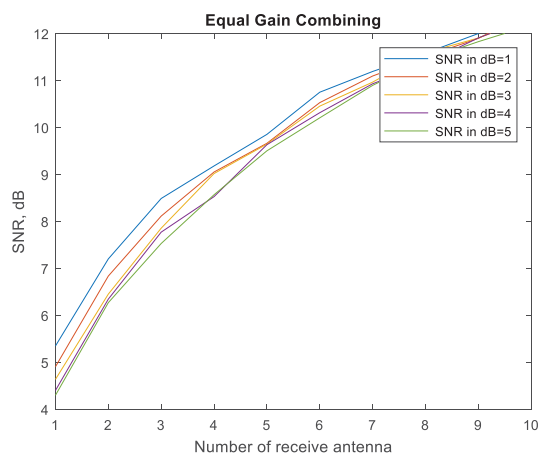
Fig. 4. EGC Combining

C. SC - Selection Combining

SC combining means that in each given time period the signal will be received from the branch to which the signal / noise ratio SNR is greatest, Fig. 5. SC is the simplest and most commonly used combination signaling technology in a diversity systems based on the choice of branch with currently the most favorable SNR ratio. The SC receiver estimates the current SNR value in all branches and choose the one with the best SNR relationship. In the simulation model of the SC receiver, after the equalization of the received signal with the known complex channel parameter, the receiver selects the branch with the best relation SNR.
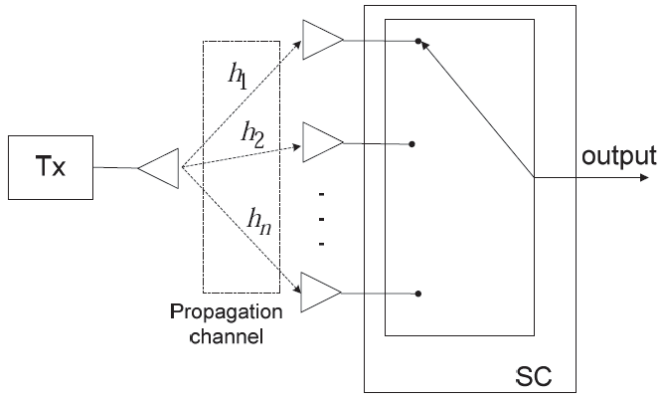
Fig. 5. SC Combining

The following figure shows the SNR relationship depending on the number of receiving antennas, using the SC technique, Fig. 6.
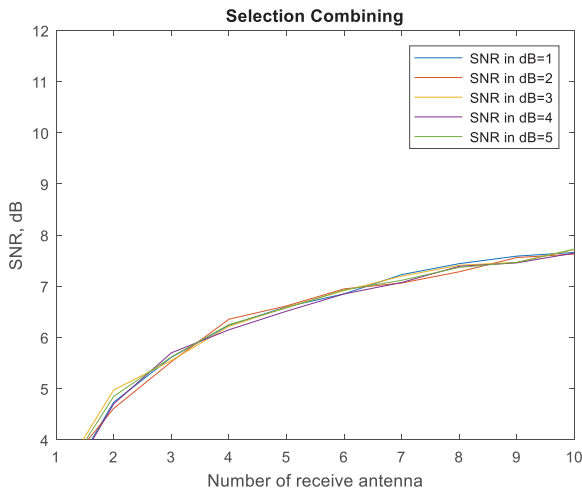
Fig. 6. SC Combining

The analysis of the aforementioned various techniques shows that the best results are achieved by the MRC combining technique which for even a smaller number of

receiving antennas offers a very favorable SNR ratio, Fig. 7.
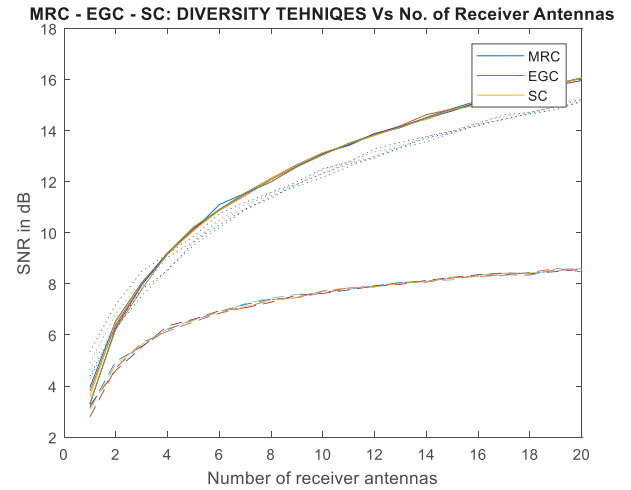
Fig. 7. MRC, EGC, SC Comparation

III. BER ANALYSIS

When there is no spatial diversity, the turbulence effect of the atmosphere results in a slow fading effecting for the symbol velocity through the channel. It is realistic to predict that the turbulent time of coherence is a great deal longer than two symbols time so that the demodulation of the DPSK of the programmed signal is possible. Direct detection of the wavefront on the receiver is used and sufficiently isolated so there is no correlation between them. BER for DPSK modulation can be written as:

$$P_{ec} = \frac{1}{2}\exp(-\frac{1}{2}SNR_e) ,$$

(1)

SNR can be calculated:

$$SNR_e = (RAI)^2 / 2\sigma^2$$

(2)

In the case of turbulence, the independent error value per bite is BER Pe = E [Pec] and is calculated using the Gauss-Hamilton quadrature integration:

$$Pe = \int_0^\infty \frac{1}{2}exp\left(-\frac{1}{2}SNRe\right)PI(I)dI =$$
$$\frac{1}{2\sqrt{\pi}}\sum_{i=1}^{n}\omega_i\, exp\left(-K^2 exp\ \left(x_i 2\sqrt{2}\sigma_i - \sigma_i^2\right)\right)$$

(3)

where
$$K = \frac{RAI}{2\sqrt{\sigma^2}}, \quad \{x_i\}_{i=1}^{n}$$

(4)

represents zeros of the n-order Hamilton polynomial combination and corresponding weight factor. The mean value for BER for multi-branch system can be expressed as:

$$BER = \frac{1}{M}\sum_{i=1}^{M} Pe_i \qquad (5)$$

We can now calculate and present graphically the error per bite. The figure represents the ratio of BER and normalized function SNR = (RE [I]) 2 / σ^2. For M = 1, n = 20, and when I_0, and R are normalized for σl = [5.0,2.0]. The results show that the BER error is 10^{-6} and turbulence levels σl = 0.2 and 0.5, DPSK modulation requires an additional ~ 1dB and ~ 1.5dB signal / noise ratio SNR respectively, comparing with the use of BPSK modulation. However, the complexity contained in adjusting the absolute phase using BPSK modulation has to be estimated in the narrow band of SNR amplification.With the figures in Figure 8, when a greater number of subcarriers M = [1, 5, 10] and σl = 0.5 were taken into account it can be concluded for the specified turbulence level that an additional signal / noise SNR ratio is needed, of course, to maintain the required bit error performance level. For example, in order to maintain a BER in the range of 10^{-6}, an additional SNR ratio of ~ 6 dB to ~ 14 dB and 20 dB is required only if M is increased from one to two.

A reasonable and very good approach to calculating BER in FSO systems is to take into account only the weakness of the signal that it suffers during propagation through the atmosphere (not taking into account signal waves and thermal processes that affect the signal, and taking into account the movement of the optical air through the atmosphere). Then, for BER we can write:

$$BER = \frac{1}{2} erfc\left(\frac{RPr}{2\sqrt{2\sigma^2}}\right) = \frac{1}{2} erfc\left(\frac{\sqrt{SNR}}{2\sqrt{2}}\right) \qquad (6)$$

where R is the aperture of the detector, Pr is the optical power on the detector and σ is the thermal noise occurring on the receiver. Typical configuration on the receiver is R = 1 A / W and the receiver's diameter of the receiver (optic) is 13 cm, the optical signal power on the transmitter is 10 mW for the 1 km link spacing. The graph based on these results is shown in Fig. 8.
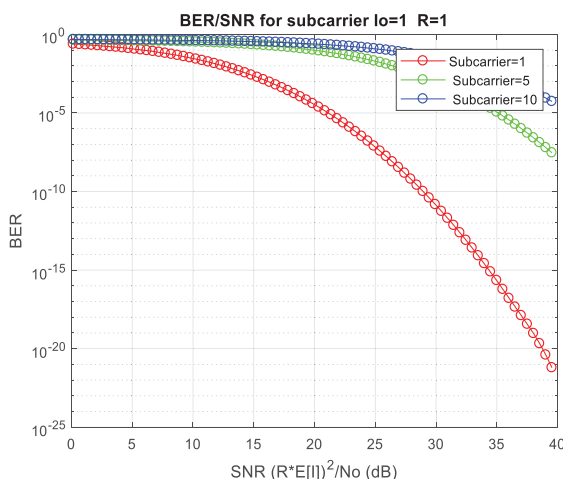
Fig. 8. BER/SNR

IV. CONCLUSION

We have presented an expression for evaluating the SNR and BER using DPSK for FSO system with in weak atmospheric turbulence. We also looked into the performance of SC, EGC and MRC spatial diversity as a possible means of circumventing the effects of scintillation. From our results, we found that the use of MRC spatial diversity in very weak turbulence resulted in reduction of link margin by up to 20 dB compared with case when no spatial diversity is used. However, as turbulence increases, MRC start to pay off, resulting in ~30 dB link margin with two photodetectors. But due to randomly varying characteristic of turbulence we do not suggest the use of SC spatial diversity with DPSK modulation.

REFERENCES

[1] H. Willebrand and B. S. Ghuman, Free Space Optics: Enabling optical Connectivity in today's network. Indiana.: SAMS publishing, 2002.

[2] M. Uysal, J. T. Li, and M. Yu, "Error rate performance analysis of coded Free-Space Optical Links over Gamma-Gamma atmospheric turbulence channels," IEEE Transactions on wireless communications, vol. 5, pp. 1229-1233, June 2006.

[3] D. Kedar and S. Arnon, "Optical wireless communication through fog in the presence of pointing errors," Applied Optics, vol. 42, pp. 4946-4954, Aug., 2003.

[4] W. K. Pratt, Laser Communication Systems, 1st ed. New York: John Wiley & Sons, Inc., 1969.

[5] R. M. Gagliardi and S. Karp, Optical Communications, 2nd Edition ed. New York: John Wiley, 1995.

[6] S. Bloom, E. Korevaar, J. Schuster, and H. Willebrand, "Understanding the performance of free-space optics," Journal of optical Networking, vol. 2, pp. 178-200, June 2003.

[7] G. R. Osche, Optical Detection Theory for Laser Applications. New Jersey: Wiley, 2002.

[8] M. K. Simon and v. A. Vilnrotter, "Alamouti-Type space-time coding for free space optical communication with direct detection," IEEE Transaction on communications, vol. 4, pp. 35-39, Jan., 2005.

[9] X. Zhu and J. M. Kahn, "Free-Space Optical Communication Through Atmospheric Turbulence Channels," IEEE Transactions on Communications, vol. 50, pp. 1293-1300, August 2002.

[10] E. J. Lee and V. W. S. Chan, "Optical communications over the clear turbulent channel using diversity," IEEE Journal on Selected Areas in Communications, vol. 22, pp. 1896-1906, 2004.

[11] S. M. Navidpour, M. Uysal, and L. Jing, "BER performance of MIMO free-space optical links," in 60th IEEE Vehicular Technology Conference, 2004. VTC2004, 2004, pp. 3378-3382

[12] P. Watson, K. C. Gupta, "EM-ANN Models for Microstrip Vias and Interconnects", IEEE Trans., Microwave Theory Tech., vol. 44, no. 12, pp. 2395-2503, 1996.

[13] B. Milovanovic, Z. Stankovic, S. Ivkovic and V. Stankovic, "Loaded Cylindrical Metallic Cavities Modeling using Neural Networks", TELSIKS'99, Conference Proceedings, pp.214-217, Nis, Yugoslavia, 1999.

[14] S. Haykin, Neural Networks, New York, IEEE Press, 1994.

Analysis of intercept events in hybrid satellite-terrestrial relay network in the presence of an eavesdropper

Aleksandra Cvetković, Jelena Anastasov, Dejan Milić, Daniela Milović, and Goran Đorđević

Abstract – **In this paper, we analyse the physical layer security of hybrid satellite-terrestrial relay networks (HSTRN). The HSTRN interconnects a satellite and destination node over decode-and-forward relay in the presence of an eavesdropper which tries to overhear satellite-relay communication. The satellite-relay link as well as satellite-eavesdropper link is subjected to shadowed-Rician fading while the terrestrial relay-destination link is corrupted by Rician fading. According to the existence of the line of sight component between relay and destination, eavesdropper is ineffective in intercepting of relay-destination transmission. Under given system scenario, we evaluate the probability of intercept and analyse the impact of various fading/shadowing channel conditions and other system parameters on secrecy performance. Monte Carlo simulations are also presented to confirm derived analytical expression.**

Keywords– **Decode-and-forward relay, Eavesdropper, Hybrid satellite-terrestrial relay network, Intercept probability.**

I. Introduction

The application of land-mobile satellite (LMS) communication is widely spread for broadcasting, navigation and rescue, in both the civil and military areas [1]. The nature and large area of LMS systems coverage makes it vulnerable for eavesdropping attack by illegitimate terrestrial nodes [2]. Due to the fact that the line-of-sight (LoS) links between the satellite and terrestrial destination nodes are often blocked by bad weather conditions or surrounding obstacles, the usage of relay between satellites and intended users is required. Thus, the advances in enhancing LMS communication are recently investigated in the context of hybrid satellite-terrestrial relay network (HSTRN) security issues.

Traditionally, cryptographic technology provides a certain level of security transmission imposing additional system complexity. This is beneficial when eavesdropper is computationally limited. Alternatively, an eavesdropper with unlimited computing power can easily decrypt cryptogram due to the *brute force attack* [3] and thus increase the system's security risk.

Apart from upper layer cryptographic techniques, physical layer security (PLS) can strengthen the secure wireless transmission by exploring dynamic nature of propagation channels [4]. The PLS of satellite-terrestrial communication has been widely explored in literature. In [5], secrecy outage probability of typical wiretapped satellite communication network is analysed. The LMS secrecy performance, where satellite employs the spot beam technique, and both the terrestrial user and eavesdropper [6] or eavesdroppers [7], are equipped with multiple antennas and utilize maximal ratio combining to receive the confidential message, is investigated.

In [8-12], a relay or even multiple relays are employed to enhance the PLS of satellite-terrestrial communication. A multi-antenna amplify-and-forward (AF) relay has been adopted in [8] to increase the secrecy capacity of HSTRN in the presence of an eavesdropper. System model in [8] refers to the scenario where user and eavesdropper are out of exclusion region. Authors in [9] employed both the AF and decode and forward (DF) relay protocols to evaluate ergodic secrecy rate in the presence of multiple eavesdropping attack. Typically, the satellite-relay channel is modeled by shadowed-Rician fading model [6-12], while the relay-destination link was treated as Rayleigh fading model [8], [9], or even as Gamma-gamma turbulence model in hybrid satellite-FSO system [10]. In addition, the PLS of HSTRN was also investigated in [11], [12] by employing different relay selection methods.

In this paper, we derive a novel expression for intercept probability of HSTRN. The information is transmitted to the destination via a DF relay and an eavesdropper tries to intercept satellite-relay signal transmission. The impact of channel condition parameters, the relay and eavesdropper antennas surface area ratio, and the average SNR values on the overall system secrecy performance, is shown. Independent Monte Carlo simulations confirm accuracy of presented analytics.

II. Problem Formulation

The system model we analysed in this paper is shown in the Fig. 1. A communication between the satellite (S) and the destination user (D) can not be performed directly but only via DF relay (R). This refers to a real military scenario for special security issues. For this purpose, a relay is equipped by a large dimensional antenna and relay-destination transmission is highly directional. An eavesdropper tries to intercept transmitted data from the satellite to intended user. Due to the fact that R-D communication is directed, eavesdropper (E) is incapable of overhearing R-D transmission phase but only S-R transmission phase. Typically, the eavesdropping attack should be hidden and therefore we assume that eavesdropper antenna's dimensions are noticeable smaller.

The S-R and S-E links are subjected to shadowed-Rician fading while the terrestrial R-D link is corrupted by Rician fading due to the existence of LoS component.

[1] Aleksandra M. Cvetković, Jelena A. Anastasov, 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: {aleksandra.cvetkovic, jelena.anastasov, daniela.milovic, dejan.milic, goran.t.djordjevic}@elfak.ni.ac.rs.
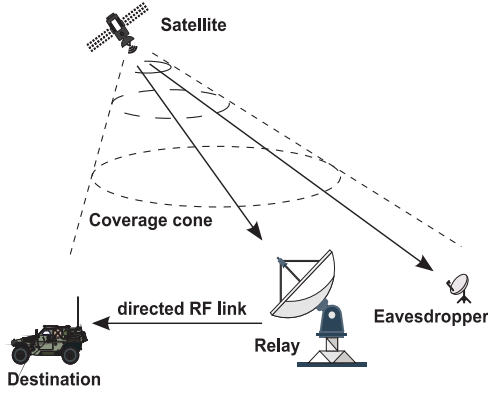
Fig. 1. System model

A. Satellite-terrestrial channels

The shadowed-Rician distribution was proposed in [13] for modeling signal amplitude over LMS channel and the probability density function (PDF) of the instantaneous signal-to-noise ratio (SNR) at the relay/eavesdropper node, $\gamma_* = E_s |h_*|^2 / \sigma^2$, is

$$p_{\gamma_*}(x) = \frac{\alpha_*}{E_s/\sigma^2} \exp\left(-\frac{\beta_* x}{E_s/\sigma^2}\right) {}_1F_1\left(m_*, 1, \frac{\delta_* x}{E_s/\sigma^2}\right). \quad (1)$$

Where $* = \{SR, SE\}$ denotes parameters of S-R and S-E communication. In Eq. (1), ${}_1F_1(.;.;.)$ is the confluent hypergeometric function [14, Eq. (9.210.1)], E_s denotes the satellite average energy and σ^2 is variance of additive white Gaussian noise (AWGN). The parameters α_*, β_* and δ_* can be calculated as

$$\alpha_* = \frac{1}{2b_*}\left(\frac{2b_* m_*}{2b_* m_* + \Omega_*}\right)^{m_*}, \quad \beta_* = \frac{1}{2b_*}, \quad \delta_* = \frac{\Omega_*}{2b_*(2b_* m_* + \Omega_*)}, \quad (2)$$

wherein $2b_*$ is the average power of the multipath components, Ω_* is average power of the LoS component and m_* is Nakagami-m parameter in shadowed-Rician links. For integer values of the fading parameter m_*, ${}_1F_1(.;.;.)$ can be simplified using [15, Eq. (07.20.03.0009.01)] and [15, Eq. (05.02.06.0005.01)] and thus (1) can be rewritten as

$$p_{\gamma_*}(\gamma) = \frac{\alpha_*}{E_s/\sigma^2} e^{-\frac{\beta_* - \delta_*}{E_s/\sigma^2}\gamma} \sum_{k=0}^{m_*-1} \frac{1}{k!}\binom{m_*-1}{m_*-1-k}\left(\frac{\delta_*}{E_s/\sigma^2}\right)^k \gamma^k. \quad (3)$$

Based on Eq. (3), the complementary cumulative density function (CCDF) can be determined using [14, Eq. (3.351.2)]

$$\bar{F}_*(\gamma) = \frac{\alpha_*}{\beta_* - \delta_*} \sum_{k=0}^{m_*-1} \sum_{l=0}^{k} \frac{1}{l!}\binom{m_*-1}{m_*-1-k}$$
$$\times \left(\frac{\delta_*}{\beta_* - \delta_*}\right)^k \left(\frac{\beta_* - \delta_*}{E_s/\sigma^2}\right)^l e^{-\frac{\beta_* - \delta_*}{E_s/\sigma^2}\gamma} \gamma^l. \quad (4)$$

B. Relay-user channel

We assume that relay position enables the LoS signal component between R and D. Also, it is assumed that fading over R-D channel follows Rician distribution. The PDF of the instantaneous SNR at the destination node is

$$p_{\gamma_{RD}}(\gamma) = \frac{(K+1)e^{-K}}{\bar{\gamma}_{RD}} e^{-\frac{(K+1)\gamma}{\bar{\gamma}_{RD}}} I_0\left(2\sqrt{\frac{K(K+1)\gamma}{\bar{\gamma}_{RD}}}\right), \quad (5)$$

where K denotes the Rician K factor (K is the ratio of the power of the LoS component to the average power of the scattered component), $\bar{\gamma}_{RD} = E\left[\gamma_{RD}^2\right]$ is the average SNR at the D and $I_0(x)$ is the zero-th order modified Bessel function of the first kind [14, Eq. (8.431.1)]. By using the infinite-series representation of $I_0(.)$ [14, Eq. (8.447.1)], the CCDF of the instantaneous SNR can be written in the following form

$$\bar{F}_{RD}(\gamma) = e^{-K} e^{-\frac{K+1}{\bar{\gamma}_{RD}}\gamma} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{K^i}{i!j!}\left(\frac{K+1}{\bar{\gamma}_{RD}}\right)^j \gamma^j. \quad (6)$$

C. Satellite -user subsystem

Under the DF relay strategy, for non-identically distributed S-R and R-D link, we can determine the CDF of γ_d in the following form

$$F_{\gamma_d}(\gamma) = 1 - \left(1 - F_{SR}(\gamma)\right)\left(1 - F_{RD}(\gamma)\right),$$
$$= 1 - \bar{F}_{SR}(\gamma)\bar{F}_{RD}(\gamma) \quad (7)$$

where $\gamma_d = \min\{\gamma_{SR}, \gamma_{RD}\}$. By substituting Eq. (4) for $* = SR$ and Eq. (6) into Eq. (7) we get

$$F_{\gamma_d}(\gamma) = 1 - \frac{\alpha_{SR} e^{-K}}{\beta_{SR} - \delta_{SR}} \sum_{k=0}^{m_{SR}-1} \sum_{l=0}^{k} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{K^i}{i!j!l!}\binom{m_{SR}-1}{m_{SR}-1-k}$$
$$\times \left(\frac{\delta_{SR}}{\beta_{SR} - \delta_{SR}}\right)^k \left(\frac{\beta_{SR} - \delta_{SR}}{E_s/\sigma^2}\right)^l \left(\frac{K+1}{\bar{\gamma}_{RD}}\right)^j e^{-\left(\frac{\beta_{SR} - \delta_{SR}}{E_s/\sigma^2} + \frac{K+1}{\bar{\gamma}_{RD}}\right)\gamma} \gamma^{l+j}. \quad (8)$$

III. PROBABILITY OF INTERCEPT

According to the Shannon capacity formula, we can evaluate the instantaneous channel capacity of the satellite-user subsystem as

$$R_d = \log_2(1 + \gamma_d). \quad (9)$$

We have already assumed a possible presence of an eavesdropper that attempts to intercept S-R transmission. Thus, the wiretap S-E channel capacity can be calculated as

$$R_e = \log_2(1 + \gamma_{SE}). \quad (10)$$

The occurrence of interception event i.e. intercept probability is defined as a probability that the transmission

rate of the main link falls below the rate on the wiretap link (secrecy rate becomes non-positive), in the following way [16]

$$P_{\text{int}} = \Pr\left[R_d - R_e < 0\right] = \Pr\left[\gamma_d < \gamma_{SE}\right]. \qquad (11)$$

After some mathematical manipulations, the previous formula becomes

$$
\begin{aligned}
P_{\text{int}} &= \int_0^\infty \int_0^{\gamma_e} p_d\left(\gamma_d\right) p_{SE}\left(\gamma_e\right) d\gamma_d d\gamma_e \\
&= \int_0^\infty F_{\gamma_d}\left(\gamma_e\right) p_{SE}\left(\gamma_e\right) d\gamma_e.
\end{aligned}
\qquad (12)
$$

Further, by substituting Eq. (8) and Eq. (3) for $* = SE$ in Eq. (12) and assuming that relay-eavesdropper antenna's surface area ratios is denoted by λ, we get

$$
\begin{aligned}
P_{\text{int}} &= 1 - \frac{\alpha_{SR} e^{-K}}{\beta_{SR} - \delta_{SR}} \frac{\alpha_{SE}}{E_s / \lambda / \sigma^2} \sum_{k=0}^{m_{SR}-1} \sum_{l=0}^{k} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \sum_{r=0}^{m_{SE}-1} \frac{K^i}{i!\,j!\,l!\,r!} \\
&\times \binom{m_{SR}-1}{m_{SR}-1-k} \binom{m_{SE}-1}{m_{SE}-1-r} \left(\frac{\delta_{SR}}{\beta_{SR}-\delta_{SR}}\right)^k \left(\frac{\beta_{SR}-\delta_{SR}}{E_s/\sigma^2}\right)^l \\
&\times \left(\frac{\delta_{SE}}{E_s/\lambda/\sigma^2}\right)^r \left(\frac{K+1}{\overline{\gamma}_{RD}}\right)^j \int_0^\infty \gamma^{l+j+r} e^{-\left(\frac{\beta_{SR}-\delta_{SR}}{E_s/\sigma^2}+\frac{K+1}{\overline{\gamma}_{RD}}+\frac{\beta_{SE}-\delta_{SE}}{E_s/\lambda/\sigma^2}\right)\gamma} d\gamma.
\end{aligned}
\qquad (13)
$$

The antenna's dimension has the influence of the received energy thus the relay-eavesdropper antenna's surface area ratios is equivalent to relay-eavesdropper received power ratio.

Applying [14, Eq. (8.310.1)] in Eq. (13), the intercept probability can be determined in a closed-form expression as

$$
\begin{aligned}
P_{\text{int}} &= 1 - \frac{\alpha_{SR} e^{-K}}{\beta_{SR} - \delta_{SR}} \frac{\alpha_{SE}}{E_s / \lambda / \sigma^2} \sum_{k=0}^{m_{SR}-1} \sum_{l=0}^{k} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \sum_{r=0}^{m_{SE}-1} \frac{K^i (l+j+r)!}{i!\,j!\,l!\,r!} \\
&\times \binom{m_{SR}-1}{m_{SR}-1-k} \binom{m_{SE}-1}{m_{SE}-1-r} \left(\frac{\delta_{SR}}{\beta_{SR}-\delta_{SR}}\right)^k \left(\frac{\beta_{SR}-\delta_{SR}}{E_s/\sigma^2}\right)^l \\
&\times \left(\frac{\delta_{SE}}{E_s/\lambda/\sigma^2}\right)^r \left(\frac{K+1}{\overline{\gamma}_{RD}}\right)^j \left(\frac{\beta_{SR}-\delta_{SR}}{E_s/\sigma^2}+\frac{K+1}{\overline{\gamma}_{RD}}+\frac{\beta_{SE}-\delta_{SE}}{E_s/\lambda/\sigma^2}\right)^{-1-l-j-r}.
\end{aligned}
\qquad (14)
$$

Approximately 20 terms are needed to bound infinity summation in Eq. (14) in order to achieve sufficient accuracy in evaluating P_{int}.

IV. NUMERICAL RESULTS

In this section we present numerical and simulation results for observed system. Numerical results are obtained based on the derived analytical expression (14). The satellite links are modeled as shadowed-Rician fading channels and utilized parameters that define different shadowing severity levels are: frequent heavy shadowing (b_*=0.063, m_*=1, Ω_* =8.94×10^{-4}), average shadowing (b_*=0.126, m_*=10, Ω_*=0.835), and infrequent light shadowing (b_*= 0.158, m_*= 19, Ω_*=1.29) [10].
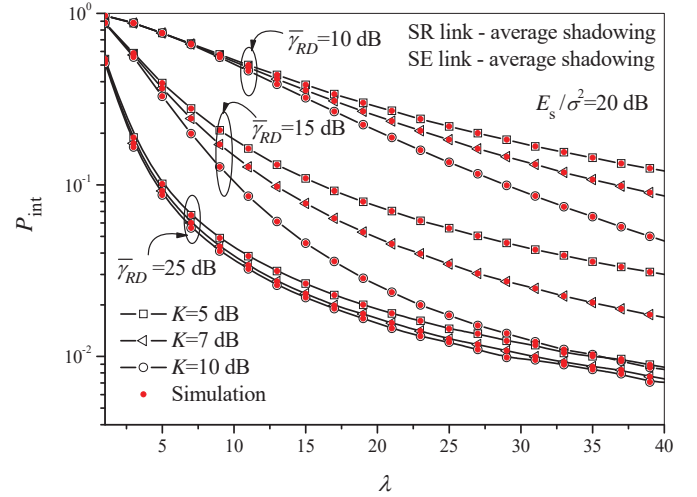
Fig. 2. Intercept probability versus the relay-eavesdropper antenna's surface area ratios

In Fig. 2, the intercept probability versus the ratio of relay and eavesdropper's antenna surface area is shown. We assume average shadowing conditions over shadowed-Rician links and constant satellite's emitting power. As expected, for higher values of parameter λ, the probability of intercept decreases and the communication between satellite and intended user via relay is secured. Also, when signal power at R-D link increases and/or Rician K factor increases, the probabilities of intercept are lower.
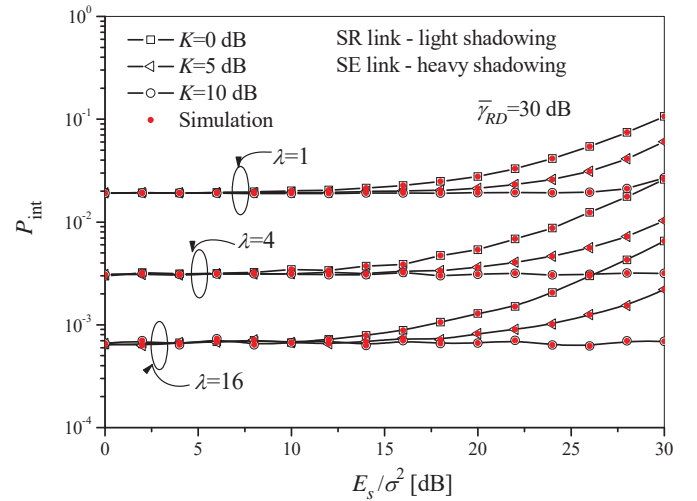
Fig. 3. Intercept probability versus the satellite emitting power for various K factors

The intercept probability dependence on the satellite's emitting power is presented in the Fig. 3. The S-R link is corrupted by light shadowing while S-E link is in heavy shadowing. We can notice that up to some specific value of satellite emitting power, the intercept probability is a constant defined by the ratio λ. For lower E_s / σ^2 values, the intercept probability depends exclusively on the S-R and S-E channel conditions and the influence of terrestrial channel conditions are marginal. Further amplification of the satellite's emitting power does not enhance the system PLS, even for favorable R-D channel conditions.

Intercept probability versus the average signal power on the R-D link for different satellite emitting power is shown in Fig. 4. It is obvious that scenario with S-E link in the heavy shadowing is beneficial for security issues. Also, we can notice that after some specific values of the average SNR, $\overline{\gamma}_{RD}$, the intercept probability tends to irreducible floor. Thus, a further variation of the system parameters don't affect security risk. In addition, the impact of satellite emitting power is more pronounced for lower-to-medium $\overline{\gamma}_{RD}$ values showing higher impact of satellite-terrestrial links.
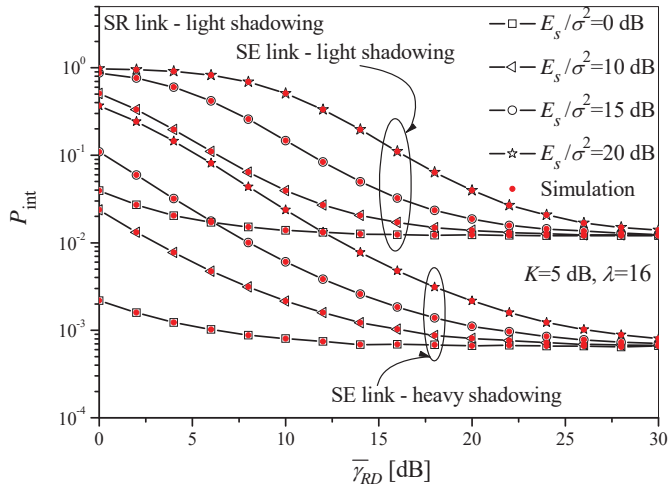
Fig. 4. Intercept probability versus the relay-destination average signal power under different shadowing conditions

In overall, the independent Monte Carlo simulations showed good agreement with analytical results presented in Figs. 2., 3. and 4.

V. CONCLUSION

In this paper, the probability of intercept of HSTRN with the DF relay in the presence of an eavesdropping attack is analyzed. Analytical expression for intercept probability is derived for integer fading parameter value of shadowed LoS component over satellite channels. The obtained results are in excellent agreement with the simulation results which were performed independently.

We showed that employment of dimensionally larger antenna at the relay brings benefits in the system PLS. Strengthen domination of LoS component at the R-D link also enhances the overall system performance. Lower intercept probability values are obtained when the wiretapped S-E link is in heavy shadowing, as expected.

ACKNOWLEDGEMENT

REFERENCES

[1] A. Vanelli-Coralli et al., "Satellite Communications: Research Trends and Open Issues", IWSSC 2007, Conference Proceedings, pp. 71-75, Salzburg, Austria, 2007.

[2] A. Roy-Chowdhury, J. S. Baras, M. Hadjitheodosiou, and S. Papademetriou, "Security Issues in Hybrid Networks with a Satellite Component", IEEE Wireless Commun., vol. 12, no. 6, pp. 50-61, 2005.

[3] F. Gandino, B. Montrucchio and M. Rebaudengo, "Key Management for Static Wireless Sensor Networks with Node Adding," IEEE Trans. Industrial Informatics, vol. 10, no. 2, pp. 1133-1143, 2014.

[4] G. Zheng, P.-D. Arapoglou and B. Ottersten, "Physical Layer Security in Multibeam Satellite Systems", IEEE Trans. Wireless Commun., vol. 11, no. 2, pp. 852-863, 2012.

[5] K. An, M. Lin, T. Liang, J. Ouyang, C. Yuan and W. Lu "Secrecy Performance Analysis of Land Mobile Satellite Communication Systems over Shadowed-Rician Fading Channels", WOCC 2016, Conference Proceedings, Chengdu, China, 2016.

[6] K. An, T. Liang, X. Yan and G. Zheng, "On the Secrecy Performance of Land Mobile Satellite Communication Systems", IEEE Access, vol. 6, pp. 39606-39620, 2018.

[7] Y. Li, K. An, T. Liang and X. Yan, "Secrecy Performance of Land Mobile Satellite Systems With Imperfect Channel Estimation and Multiple Eavesdroppers", IEEE Access, vol. 7, pp. 31751-31761, 2019.

[8] K. An, M. Lin, T. Liang, J. Ouyang, C. Yuan and Y. Li, "Secure Transmission in Multi-antenna Hybrid Satellite-Terrestrial Relay Networks in the Presence of Eavesdropper", WCSP 2015, Conference Proceedings, Nanjing, China, 2015.

[9] Q. Huang, M. Lin, K. An, J. Ouyang and W.-P. Zhu, "Secrecy Performance of Hybrid Satellite Terrestrial Relay Networks in the Presence of Multiple Eavesdroppers", IET Commun., vol. 12 , no. 1, pp. 26-34, 2018.

[10] Y. Ai, A. Mathur, M. Cheffena, M. R. Bhatnagar and H. Lei, "Physical Layer Security of Hybrid Satellite-FSO Cooperative Systems", IEEE Photonics J., vol.11, no. 1, 2019.

[11] V. Bankey, P. K. Upadhyay, "Physical Layer Security of Multiuser Multirelay Hybrid Satellite-Terrestrial Relay Networks", IEEE Trans. Veh. Technol., vol. 68, no. 3, pp. 2488-2501, 2019.

[12] W. Cao, Y. Zou, Z. Yang and J. Zhu, "Relay Selection for Improving Physical-Layer Security in Hybrid Satellite-Terrestrial Relay Networks", IEEE Access, vol. 6, pp. 65275-65285, 2018.

[13] A. Abdi, W. Lau, M.-S. Alouini, and M. Kaveh, "A New Simple Model for Land Mobile Satellite Channels: First and Second Order Statistics", IEEE Trans. Wireless Commun., vol. 2, no. 3, pp. 519–528, 2003.

[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA, USA: Academic, 2007.

[15] The Wolfarm Functions Site, 2008. [Online] Available: http://functions.wolfram.com/

[16] N. Milošević, J. Anastasov, A. Cvetković, D. Milović and D. Milić, "On the Intercept Probability of DF Relaying Wireless Communication", Wireless Pers. Commun., vol. 104, no. 4, pp 1523–1533, 2019.

Extreme Learning Machines for Wireless Channel Prediction in Microcell and Picocell Environments

Nikola Sekulović[1], Miloš Stojanović[2], Aleksandra Panajotović[3] and Miloš Banđur[4]

Abstract – **In this paper, examination of possibility and effectiveness of extreme learning machines (ELM) application to predict wireless channel conditions for single-input single-output (SISO) systems in microcellular and picocellular environments is carried out. Normalized mean squared error (NMSE) and time consumption are used as performance indicators. The experimental results on measured values for signal-to-noise ratio (SNR) show high accuracy of the ELM prediction model and short execution time.**

Keywords – **Channel prediction, Extreme learning machines, Microcellular environment, Picocellular environment.**

I. INTRODUCTION

Knowledge of information about state of wireless channel is increasingly important. The reason of that trend lies in demands for high-data services and limited wireless spectrum. Unfortunately, a state of wireless channel changes very quickly, so channel state obtained by channel estimation can become outdated due to delay caused by processing and feedback phases. The system performance enhancement can be achieved using channel prediction based on channel states in previous moments rather than using channel estimation [1].

In the open technical literature, there are several papers dealing with channel states prediction. Autoregressive (AR) model, support vector machine (SVM), discrete wavelet transform (DWT) method in combination with AR and linear regression (LR) algorithm (DWT-AR-LR) and echo state network (ESN) are widely explored in [2]-[6].

Extreme learning machine (ELM) is a learning algorithm for feedforward artificial neural networks with one hidden layer. Compared with traditional artificial neural networks, ELM may achieve better generalization performance for regression and classification cases. ELM tends to minimize training error with the smallest norm of weights. In addition, ELM has faster learning speed, i.e. significantly low computational time required for training (up to thousands of times) [7]-[9]. It increases training speed by randomly assigning weights and biases in the hidden layer, instead of iteratively adjusting its parameters by gradient based methods.

[1]Nikola Sekulović is with the College of Applied Technical Sciences, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: nikola.sekulovic@vtsnis.edu.rs.

[2]Miloš Stojanović is with the College of Applied Technical Sciences, Aleksandra Medvedeva 20, 18000 Niš, Serbia.

[3]Aleksandra Panajotović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[4]Miloš Banđur is with the Faculty of Technical Sciences, University of Priština, Knjaza Miloša 7, 38220 Kosovska Mitrovica, Serbia.

In this paper, the effectiveness of prediction scheme based on ELM is explored for microcellular and picocellular environments. Data sets used for training and testing contain measured signal-to-noise ratio (SNR) samples for scenarios described in details in [10]. Performance metrics used for analysis of the approach proposed is normalized mean squared error (NMSE) and time consumption.

The rest of the paper is organized as follows. Section II provides brief description of ELM. Section III describes communication scenario, data sets and ELM-based prediction algorithm. Experimental evaluation is presented in Section IV, while Section V concludes the paper.

II. EXTREME LEARNING MACHINES

Let's denote N training samples as $(\mathbf{x}_j, \mathbf{y}_j)$, $j=1,...,N$, where $\mathbf{x}_j = [x_{j1}, x_{j2}, ..., x_{jn}]^T \in \mathbf{R}^n$ represents the j-th n-dimensional training instance and $\mathbf{y}_j = [y_{j1}, y_{j2}, ..., y_{jm}]^T \in \mathbf{R}^m$ represents the j-th target value of the dimension m. ELM has the unified solutions for regression, binary and multiclass classification. In the case of regression, which is of interest for problem considered in this paper, it holds that $m=1$ [7]. Generally, the output of a standard single hidden layer feedforward network (SLFN) with L hidden neurons and activation function $h(x)$ is defined as

$$\sum_{i=1}^{L} \boldsymbol{\beta}_i h(\mathbf{w}_i \mathbf{x}_j + b_i) = \mathbf{f}_j, \quad j = 1,\ldots,N, \qquad (1)$$

where $\mathbf{w}_i = [w_{i1}, w_{i2}, ..., w_{in}]^T$, $i=1,...,L$, is the weight vector connecting the i-th hidden neuron and all input neurons, $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$ represents the weight vector connecting the i-th hidden neuron and all the output neurons, and b_i is the threshold of the i-th hidden neuron. According to ELM theory, $\mathbf{w}_i$ and b_i can be randomly and independently assigned a priori, i.e. without considering the input data [8].

The SLFN defined with (1) has approximation capabilities with zero error means $\sum_{i=1}^{L} \|\mathbf{f}_i - \mathbf{y}_i\| = 0$, i.e., there exist $\boldsymbol{\beta}_i$, $\mathbf{w}_i$ and b_i such that

$$\sum_{i=1}^{L} \boldsymbol{\beta}_i h(\mathbf{w}_i \mathbf{x}_j + b_i) = \mathbf{y}_j, \quad j = 1,\ldots,N. \qquad (2)$$

The previous equation can be expressed in matrix form resulting in

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}, \qquad (3)$$

with

$$\mathbf{H} = \begin{bmatrix} h(\mathbf{w}_1\mathbf{x}_1 + b_1) & \cdots & h(\mathbf{w}_L\mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ h(\mathbf{w}_1\mathbf{x}_N + b_1) & \cdots & h(\mathbf{w}_L\mathbf{x}_N + b_L) \end{bmatrix}_{N \times L}, \quad (4)$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L \times m}, \quad (5)$$

and

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix}_{N \times m}. \quad (6)$$

The matrix **H** represents the hidden layer output matrix of the neural network where the *i*-th column of **H** represents the *i*–th hidden neuron's output vector in regard to inputs $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_N$. The output weights can be analytically determined by finding the unique smallest norm least-squares solution of the linear system described by (3). In order to improve the performance, the constrained optimization problem can be formed for ELM, as shown in [7]:

$$Minimize: L_P = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\frac{1}{2}\sum_{j=1}^{N}\|\boldsymbol{\xi}_j\|^2$$

$$Subject\ to: h(\mathbf{x}_j)\boldsymbol{\beta} = \mathbf{y}_j^T - \boldsymbol{\xi}_j^T, j = 1,...,N \quad (7)$$

where $\boldsymbol{\xi}_j = \left[\xi_{j1},...,\xi_{jm}\right]^T$ is the training error vector of the *m* output nodes with respect to the training sample x_i, while C represents tradeoff parameter between model complexity and allowed errors $\boldsymbol{\xi}_j$ during training. Based on *Karush-Kuhn-Tucker* (KKT) theorem, the optimization problem previously defined is equivalent of solving the dual optimization problem

$$L_D = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\frac{1}{2}\sum_{j=1}^{N}\|\boldsymbol{\xi}_j\|^2 - \sum_{j=1}^{N}\sum_{i=1}^{m}\alpha_{ji}(h(\mathbf{x}_j)\boldsymbol{\beta}_i - \mathbf{y}_{ji} + \boldsymbol{\xi}_{ji}), \quad (8)$$

where $\alpha_j = \left[\alpha_{j1},...,\alpha_{jm}\right]^T$ are Lagrange multipliers.

After solving (8) based on *KKT* conditions, which can be found in detail in [7], the following solution is obtained:

$$\boldsymbol{\beta} = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{Y} \quad (9)$$

and the function of ELM is:

$$f(\mathbf{x}) = h(\mathbf{x})\boldsymbol{\beta} = h(\mathbf{x})\left(\frac{\mathbf{I}}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{Y}. \quad (10)$$

III. APPLICATION OF ELM FOR WIRELESS CHANNEL PREDICTION

A. Channel Description

The efficiency of the proposed machine learning technique for SNR prediction in wireless communication system employing a single transmit antenna and a single receive antenna is investigated. Namely, single-input single-output (SISO) channel in two different environments is considered:

1) B channel model represents a microcell environment where distance between mobile station (MS) and base station (BS) is in the order of 30 m. It assumes indoor-to-outdoor propagation with BS located outside and indoor environment usually consisted of several small offices.

2) E channel model refers to indoor-to-indoor scenario. It represents a picocell environment in modern open office with windows metallically shielded.

B. Data Sets

Data sets used for analysis in this work contain SNR channel values obtained based on measurement campaigns described in details in [10]. A series of SNR samples $x(k) = x(kT)$, $k = \overline{1, N}$, from [10], are used for network training and testing. Parameter *T* denotes sampling interval and parameter *N* is the total number of samples.

C. Prediction Algorithm

In general, for a given training set with *N* instances of *n* features, the sigmoid activation function $g(x)$ and *L* hidden neurons number, the ELM algorithm for regression can be summarized as follows:

Training procedure
(a) Assign random input weights $\mathbf{w}_i$, and biases b_i, $i = 1,...,L$;
(b) Compute the hidden layer output matrix **H** using (4);
(c) Compute the output weights $\boldsymbol{\beta}$ using (9);

Testing procedure
(a) Compute the hidden layer output vector $h(\mathbf{x})$ for current instance from the test set using (4);
(b) Compute the output $f(\mathbf{x})$ according to (10) using the $\boldsymbol{\beta}$ obtained in step (c) of the training.

IV. EXPERIMENTAL EVALUATION

In this section of the work, the accuracy of the ELM network for the time series prediction of SNR in the SISO system is evaluated using NMSE as a prediction error metrics. NMSE is defined as

$$NMSE = \frac{\sum_k \left(x(k) - f\left(x(k) \right) \right)^2}{\sum_k \left(x(k) \right)^2}. \qquad (11)$$

Data sets containing the measured instantaneous SNR values at the receiver side for the case when SNR at the transmitter side is 20 dB for both B and E channel model are used to test the proposed method. Analysis is carried out using N=4000 samples. The data sets are divided into two equal sets for training and testing (N_{tr}=N_{te}=2000). It is determined by simulation that there is no need to use more than 3 neurons in the input layer. Sigmoid function is used as activation. For the tests, the ELM is implemented in MATLAB. As an illustrative example, Fig. 1 shows target signal and prediction curve for the case of E channel.
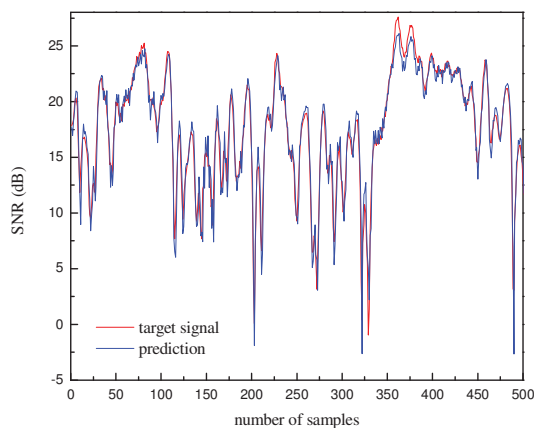
Fig. 1. Target signal and prediction curve for E channel
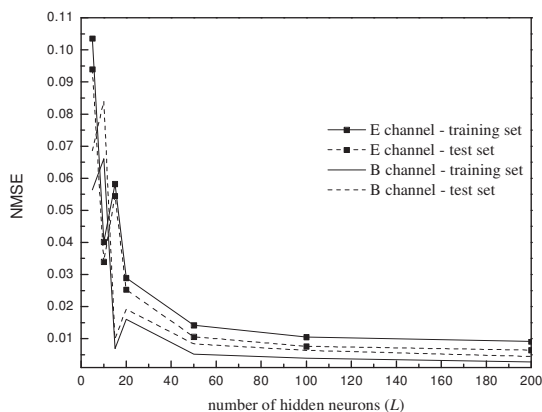
Fig. 2. NMSE versus number of hidden neurons

The NMSE is evaluated for both training and test set, with training and test times, measured in seconds on an *Intel Core i5* computer. Results for the NMSE as a function of the number of neurons in hidden layer are presented graphically in Fig. 2. It is notable that in the case of E channel, the NMSE on the test set remains in range from 0.1 to 0.006 for the number of neurons in hidden layer in range from 5 to 200. In the case of B channel, for the same prediction network, the NMSE values are slightly lower. They are in the range from 0.08 to 0.004. Furthermore,

the analysis shows that increasing number of neurons above a certain value does not improve prediction results significantly. We can also observe that the NMSE on the training set follows trend of the NMSE on training set regardless of number of hidden neurons, which implies that models are not overfitted. The obtained values of the NMSE for all ELMs, regardless the number of neurons in the hidden layer are within the expected range of precision and comparable to the results obtained in other studies [5, 6].

Table I contains training and test time in seconds for different number of neurons in hidden layer. We can see that training time for all 2000 instances in training set is only 0.001 seconds with 100 neurons in hidden layer, while prediction for all 2000 instances is done in 0.06 second. With increase of number of neurons up to 1000, these times slightly increase. For more than 2000 neurons in hidden layer training time increases significantly (greater than 1s), while test time increases slightly (but it is still less than 1s). These results demonstrate high performances in terms of training and test speed on this data set.

TABLE I
TIME CONSUMPTION OF THE ELM MODEL

Number of neurons	Training time (s)	Test time (s)
5	0.001	0.0312
10	0.001	0.0312
15	0.001	0.0468
20	0.001	0.0468
50	0.001	0.0468
100	0.001	0.0625
200	0.0468	0.0625
300	0.0781	0.0625
400	0.2031	0.0625
500	0.2187	0.0756
1000	0.8437	0.1406
2000	3.6875	0.3281
3000	8.9218	0.3593
4000	16.9063	0.4843
5000	29.1719	0. 5781

In order to compare the results of the ELM with other common classification techniques, we have measured accuracy of the Linear SVM and RBF SVM, on the same dataset. NMSE for Linear SMV was 0.0118, while NMSE for RBF SVM reached 0.0083. It can be noted that ELM outperforms Linear SVM in terms of NMSE, having the similar algorithm complexity. On the other side, ELM reaches results comparable to the RBF SVM, while operating significantly faster during the training and testing.

V. CONCLUSION

This paper has investigated the ELM-based prediction scheme for SISO systems in microcellular and picocellular environments. The effectiveness of the framework has been

confirmed using NMSE as a performance measure along with training and test time. Simulation results have shown that no more than several hundred neurons in hidden layer should be used. Further increasing the number of neurons will not result in significant prediction accuracy gain. The NMSE of the order of 10^{-3} and training and test time less than 0.22 and 0.08 seconds, respectively, can be expected.

ACKNOWLEDGEMENT

REFERENCES

[1] A. Panajotović, F. Riera-Palou and G. Femenias, "Limited feedback MU-MIMO-OFDM system", International ITG Workshop on Smart Antennas WSA 2016, pp. 348-355, Munich, Germany, 2016.

[2] P. Sharma and K. Chandra, "Prediction of state transitions in Rayleigh fading channels", IEEE Trans. Veh. Technol., vol. 56, pp. 416-425, 2007.

[3] X. Zhao, C. Hou and Q. Wang, "A new SVM-based modeling method of cabin path loss prediction", Int. J. Antennas and propagation, vol. 2013, pp. 1-7, April 2013.

[4] X. Long and B. Sikdar, "A wavelet based long range signal strength prediction in wireless networks", in Proc. IEEE Int. Conf. Commun., pp. 2043-2047, 2008.

[5] Y. Zhao, H. Gao, N. C. Beaulieu and Z. Chen, "Echo State Network for Fast Channel Prediction in Rician Fading Scenarios", IEEE Comm. Lett., vol. 21, pp. 672-675, 2017.

[6] N. Sekulović, A. Panajotović, M. Banđur, Đ. Banđur and D. Milović, "Channel Prediction in Wireless Microcell and Picocell Systems Using Echo State Network", AJER, vol. 7, no. 12, pp. 134-138, 2018.

[7] G.-B. Huang, H. Zhou, X. Ding and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification", IEEE T. Syst. Man CY-S., vol. 42, no. 2, pp. 513-529, 2012.

[8] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, "Extreme Learning Machine: Theory and Applications", Neurocomputing, vol. 70, no. 1-3, pp. 489-501, 2006.

[9] X. Sun, J. Xu, C. Jiang, J. Feng, S-S. Chen and F. He, "Extreme Learning Machine for Multi-Label Classification", Entropy, vol. 18, pp. 225-236, 2016.

[10] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen and F. Frederiksen, "A Stochastic MIMO radio channel model with experimental validation", IEEE J. Select. Areas Commun., vol. 20, pp. 1211-1226, 2002.

Comparative Analysis of MTP and DSDV Routing Protocols in VANET

Tsvetan Marinov[1], Maria Nenova[2] and Georgi Iliev[3]

*Abstract –***Nowadays we are witnessing the rapid development of telecommunication technologies. The transmission speed is constantly increasing and the routing protocols are improved. Intelligent Transport Systems are also part of this development. Vehicular ad-hoc network - VANET belongs to them. It is expected to solve serious problems such as road accidents, congestions and harmful air emissions. This report presents a comparative analysis between two protocols - Message Transmission Protocol - MTP and Destination-Sequenced Distance-Vector Routing - DSDV.**

*Keywords –***ITS, VANET, routing protocols, MTP, DSDV**

I. INTRODUCTION

The number of cars around the world is increasing every day. Road accidents are constantly occurring. Congestions are a common part of everyday life. The air is getting polluted. As a result, many people get sick. All these problems require the development of VANET. It is subclass of MANET (mobile ad hoc network) network. Originally MANET was developed for military purposes. The basic idea is that moving devices can communicate with each other. The main purpose of VANET is the same but here the communication will be between cars. If vehicles can exchange information among themselves, mentioned problems will be solved.

VANET differs from all known networks. This difference is determined by the movement of vehicles. As a result, the network has a dynamic topology. Architecture is varied and depends on the geographical area. Figure 1 shows the architecture of VANET. In general, the architecture may include all communication equipment.

The communication is divided into four types:

The first type is In-vehicle communication. Here each controller or computer in a vehicle can communicate with each other following the driver and vehicle behavior.

The second is Vehicle to Vehicle communications (V2V). The main idea is that cars can exchange information with each other. This is a new technology that is developed for VANET. In this area, a lot of research has been done to make the links between cars as reliable as possible.

The third communications is Vehicle to road infrastructure

[1]Tsvetan Marinov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, 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 1000, Bulgaria, E-mail: mariamnenova@gmail.com

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

(V2I). In these types of communications vehicles can communicate with road infrastructure as traffic lights, base stations and so on.

The last type of communications is Vehicle-to-broadband cloud (V2B). This allows wireless communication of automobiles over broadband connections such as 3G/4G. The broadband cloud has a great resource and may include more traffic information. It is also possible to use it for entertainment. This way the trip will be more pleasant.

Next section describes MTP and DSDV protocols. These protocols are selected because of their presence in the NS2-35 simulator (The MTP protocol is a modified version of Ad-hoc On-Demand Distance Vector (AODV). NS2-35 is open source code and allows modifying different types of protocols. This makes it suitable for the research.
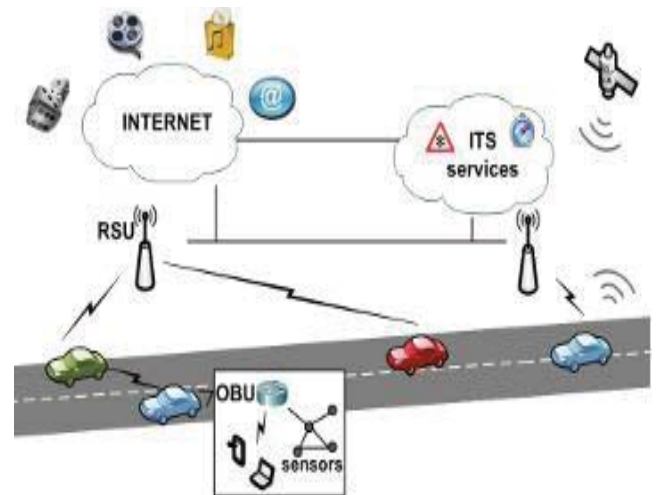
Fig. 1. Architecture of VANET

II. DESCRIPTION OF MTP AND DSDV ROUTING PROTOCOLS

A. Description of Message Transmission Protocol - MTP

This section describes shortly Message Transmission Protocol. Detailed description is provided in [1]. The proposed MTP restricts unreliable connections in VANET formed because of the high mobility of the network. During the route detection phase, cars send routing packets. When neighboring car accepts these packages, the speed of the vehicle is determined in order to form a reliable route of the package. If the vehicle moves too quickly, the neighbor refuses to give it the message. The new algorithm of the protocol helps to eliminate high speed vehicles, thus reducing unreliable links and saving bandwidth [1].

Decision making under this protocol depends on the duration of the link connection. In this research, the duration of the link is measured quantitatively from the time when two cars are connected without interruption. This amount is called LT (link time). The link time (LT) between two automobiles can be defined as a predicted time for connection between nodes [2]. In other words, this is the predicted time when two nodes have an active connection without interruption. LT is calculated using the GPS system of the nodes [3]. LT is defined by the following formula:

$$LT = \frac{-(ab+cd) + \sqrt{(a^2+c^2)r^2 - (ad-bc)^2}}{a^2+c^2} \quad (1)$$

Parameter "a" is the relative speed of the receiving vehicle with respect to the sending vehicle by axis X.

Parameter "b" is used to determine the distance between the receiving vehicle from the sender along the X axis.

Parameter "c" is the speed of the receiving vehicle with respect to the sending vehicle by axis Y.

Parameter "d" is the distance between the receiving vehicle and the sender.

The MTP protocol is a modified version of Ad-hoc On-Demand Distance Vector (AODV). The algorithm is added to the MAC layer in NS-2.35 (Network Simulator) and calculates LT. Pseudo-code of the algorithm is given in [1].

A. Description of Destination-Sequenced Distance Vector routing-DSDV

Destination-Sequenced Distance-Vector Routing - DSDV is a table-driven routing scheme for ad-hoc mobile networks based on the Bellman–Ford algorithm. It was developed by C. Perkins and P. Bhagwat in 1994 [3]. It is one of the basic protocols in mobile networks. The routing algorithm of the DSDV solves the routing loop problem. Routing information is recorded in a table. The table is arranged according to the sequence number of the data received. The number is generated by the destination node, and the transmitter needs to send out the next update with this number. Routing information is distributed between cars by sending full dumps infrequently and smaller incremental updates more frequently.

Upon receiving a new message, the cars use the most recent route number. A major advantage of the DSDV is the rapid creation of a route. The protocol is not suitable for dense networks. If the network has a large number of cars, the messages will be delayed due to the update of the routing table. Nowadays DSDV is not one of the most used protocols. In this study, the protocol was chosen for comparison because of its presence in the NS-2.35 simulator.

III. SIMULATIONS RESEARCH OF THE PROTOCOLS

A. The Scenario

Figure 2a shows link formation under the DSDV protocol. Figure 2b shows link formation under MTP. Car seven will send a data and car one should receive it. Dashed lines show the active link and the direction of motion of the vehicles. The selected speed is 20m/s. The speed is carefully selected to ensure the reliability of the connection. Car 7 should send a message to car 1. DSDV forms a road through cars 7-6-5-4-2-1. If cars are moving in the same direction, the message will get to car 1 without any problems. But car 2 will change its direction of movement. As a result, the connection will be interrupted. The DSDV will try to restore the link but without result. After a certain time, a link will be formed through 7-6-5-4-3-1. As a consequence of all this will get a delay in the delivery of the message. It is also possible to lose packets. High traffic will be generated and collisions may occur.

a.   b.
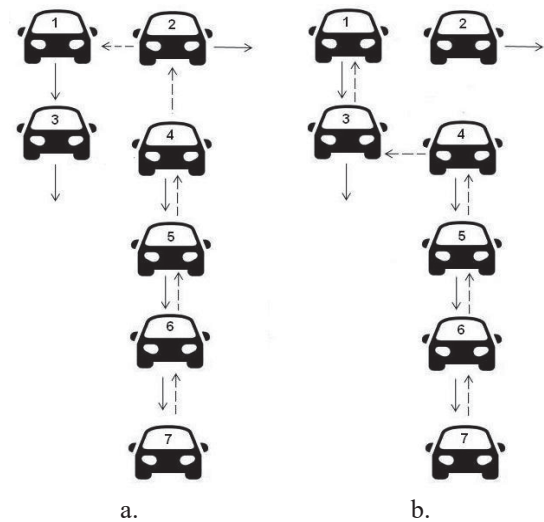
Fig. 2. a. Link formation under the DSDV protocol b. Link formation under the MTP protocol

MTP offers a solution to this problem. If the link time - LT is calculated and compared with the time it takes for the message to be delivered, it will be known whether it is possible to reach the final destination. MTP calculates the LT of the scenarios given in the study and establishes that a reliable connection through vehicle 2 will not exist. As a result, a route is formed through cars 7-6-5-4-3-1.

For the research done in this work NS-2.35 is used. A TCL file has been created for the execution of the given scenario. Initially, the file is compiled under the DSDV protocol and then recompiled under MTP. The simulator then creates two protocols analysis files. The channel parameters are given in Figure 3. The IEEE 802.11p standard is used. The scenario under consideration has a simple network topology because cars are few in number.

The network size is 700mX700m. The vehicles are selected to be seven in number. The Traffic Model is generated with CBR (constant bit rate). CBR is a feature set in the NS-2.35 simulator. In this research, the sending vehicle must transmit 512 bytes per second. The average data transmission rate is 256 kbps. The transport protocol used is TCP (Transmission Control Protocol). The cars have initial coordinates - car 0 (300, 700), car 1 (400, 700), car 2 (300, 600), car 3 (400, 600) car 4 (400, 500), car 5 (400, 400), car 6 (400, 300). They are set to move continuously.

Gt_	1
Gr_	1
L_	1.0
freq_	2.472e9
bandwidth_	11Mb
Pt_	0.031622777
CPThresh_	10.0
CSThresh_	5.011872e-12
RXThresh_	5.82587e-09
dataRate_	11Mb
basicRate_	1Mb

Fig. 3. The channel parameters

Where:
L - System Loss Factor
Freq – Channel frequency
Bandwidth – Channel bandwidth
Pt – Transmission power
CPThresh - Collision Threshold
CSThresh- Carrier Sense Power
RXThresh- Receive Power Threshold

B. Research parameters

The parameters tested for the two protocols are Packet Delivery Ratio, Normalized MAC Load and End-to-End Delay. After analyzing the results, we can see that the new protocol is doing well. This is achieved by the fact that there is no disconnection.

The Packet Delivery Ratio represents the ratio of the data packets delivered to the destination to those generated by the CBR sources. The PDR is calculates by following formula:

$$PDR = \frac{\text{Data packet delivered to all sources}}{\text{Data packet send by all sources}} \quad (2)$$

Figure 4 shows the Packet Delivery Ratio. The figure shows that MTP performs well with the DSDV for the given scenario.

The Normalized MAC Load is defined as the fraction of all control packets (routing control packets, Clear-To-Send (CTS), Request-To-Send (RTS), Address Resolution Protocol (ARP) requests and replies, and MAC ACKs) over the total number of successfully received data packets. This is the metric for evaluating the effective utilization of the wireless medium for data traffic. The NML is calculates by following formula:

$$NML = \frac{\text{Number of routing packets sent}}{\text{Number of data packets delivered}} \quad (3)$$

Figure 5 shows Normalized MAC load. MPT has a greater NML than DSDV. The main reason for the better performance of MTP is that the original route will be through 7-6-5-4-3-1

cars. There will be an interruption in the DSDV because the initial road is through 7-6-5-4-2-1.
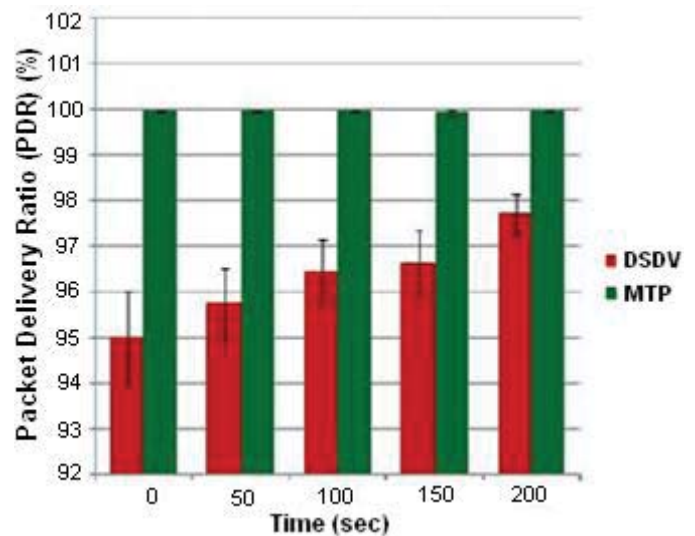
Fig.4 Packet delivery Ratio
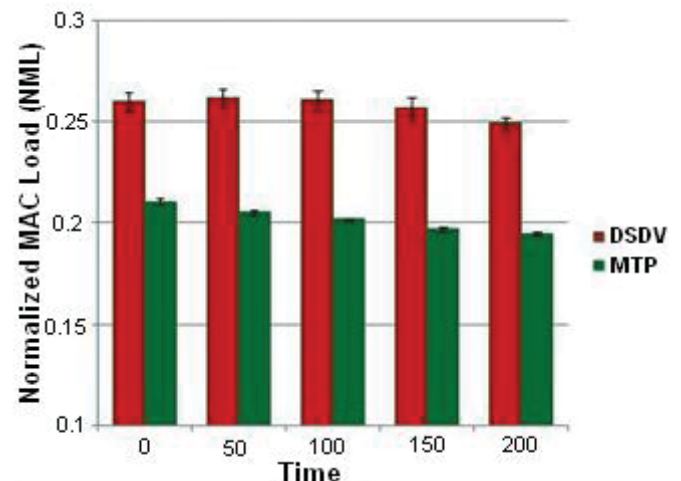
Fig. 5 Normalized MAC Load
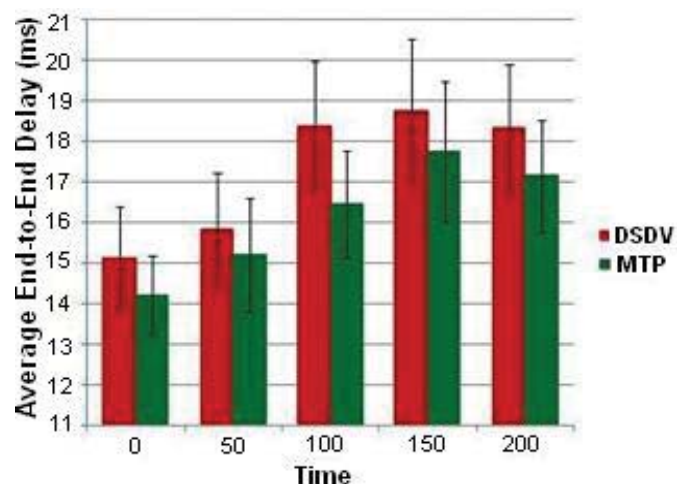
Fig. 6 End-to-End Delay

The average End-to-End Delay of data packets includes all possible delays caused by buffering during routing discovery, queuing at the interface queue, retransmission at MAC layer, propagation, and transfer time. The End-to-End Delay is calculates by following formula:

$$\text{Delay} = \frac{\Sigma(\text{Packet received time } - \text{ packet sent time})}{\Sigma(\text{packets received by destinations})} \quad (4)$$

Figure 6 shows End-to-End Delay. MTP performs better than DSDV. The figure shows that the delay is different in time. This is because every car is delaying the network.

IV. CONCLUSION

In this paper Message Transmission Protocol and Description of Destination-Sequenced Distance Vector routing protocol are described and compared. Most protocols have a problem with the formation of a proper route due to the high mobility of vehicles. As a result, the links are in a continuous process of connecting and disconnecting. MTP algorithm reduces unreliable links in VANET. The protocols are compared using the following three parameters - Packet Delivery Ratio, Normalized MAC Load and End-to-End Delay. The study shows that the MTP performs better than the DSDV.

In the scenario under consideration the cars are seven in number. These are a few vehicles. The topology of the network is simple. As a future work, the protocol should be tested with more vehicles. It is possible for a large number of vehicles and a densely populated network to get a long delay from the LT calculation. An attempt will also be made to improve some of the protocol parameters in selecting next hops during route discovery phase.

REFERENCES

[1] Tsvetan Marinov, Maria Nenova and Georgi Iliev "Message Transmission Protocol in VANET" Proc. X National Conference with International Participation Conference "Electronica 2019", May 16 - 17, 2019, Sofia, Bulgaria.

[2] R. Oliveira, M. Luis, L. Bernardo, R. Dinis, R. Pinto, P. Pinto. "Impact of Node"s mobility on Link-Detection Based on Routing Hello Messages." WirelessCommunications and Networking Conference (WCNC10). 2010.

[3] C. E. Perkins and P. Bhagwat, "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers," ACM SIGCOMM Computer Communication Review, vol. 24, no. 4, pp. 234–244, Oct. 1994.

[4] M. Zhang and R. S. Wol_. Routing Protocols for Vehicular Ad Hoc Networks in Rural Areas. IEEE Communication Magazine, 46(11):126{131, 2008.

[5] Tsvetan Marinov, Maria Nenova and Georgi Iliev "Comparative analysis of broadcasting protocols in VANET" Proc. IX National Conference with International Participation Conference "Electronica 2018", May 17 - 18, 2018, Sofia, Bulgaria.

[6] H. Moustafa and Y. Zhang. Vehicular networks: techniques, standards, and applications. CRC Press, 2009.

[7] S. Youse_, M. S. Mousavi, and M. Fathy. Vehicular Ad Hoc Networks (VANETs): challenges and Perspectives. In proceedings of 6th Interma-tional Coference on ITS Telecommunications, pp. 761-766, 2006.

[8] W. Chen, R. K. Guha, T. J. Kwon, J. Lee, and I. Y. Hsu. A survey and challenges in routing and data dissemination in vehicular ad hoc networks. In Proceedings of the 2008 IEEE International Conference on Vehicular Electronics and Safety, pp. 328-333, 2008.

[9] M. M. Al-Doori, F. Siewe, and A. H. Al-Bayatti. VSHDRP: Vehicle second heading directional routing protocol in VANET. In proceeding of 4th IEEE International Conference on Computer Science and Information Technology (ICCSIT2011), Chengdu, China, 2011.

[10] D. A. Maltz, J. Broch, J. Jetcheva, and D. B. Johnson. The effects of on-demand behavior in routing protocols for multi-hop wireless ad hoc networks. IEEE Journal on Selected Areas in Communications, 17:1439-1453, 1999.

[11] A. Festag, A. Hessler, R. Baldessari, L. Le, W. Zhang, and D. Westhoff. Vehicle-to-Vehicle and road-side sensor communications for enhanced road safety, 2009.

Providing tailor made managed solution by network operator to the business customers

Zlata Radak[1] and Goran Marković[2]

Abstract – **In this paper we will describe the process of implementing tailor made managed solution for business customers and how this process increases the availability of the solution for connection on the Internet, for the specific business customer. In this paper we will present one specific enterprise retail company and their requirements for the connecting on the Internet. At first, business customer has to choose between some of the predefined solutions that network operator already has in his service portfolio. But if the business customer has specific request, network operator need to adjust the solution in order to fulfill this request. In this specific case, business customer request is higher availability of the head office location. Passing through three different solutions, we will see how availability grows.**

Keywords – *Business customer, Internet connection, Network operator, Tailor made solution*

I. INTRODUCTION

Traditional network operators need to rethink their current strategies for delivering solutions for connecting business customers to the Internet [1]. They need a new strategy and instead of selling services from the service portfolio, they have to think of how to sell whole packet of telecommunication services to the specific customer and to satisfy their specific needs more carefully. Practice has shown that if operator satisfy customer needs more, the quality of experience (QoE) will be better. In these competitive environment the fulfillment of customer demands and QoE are becoming the main differentiators for the effectiveness of telecom operators [2]. Modern network operators made a decision to listen the voice of the customers and implement in the final product customer requirements and experience of using different services. In this way, operator will manage to provide to the customer additional performance and security improvements. This is very important in the term of customer engagement [5]. At first, operator will perceive the business customer network. According to this, operator will provide the best solution from the service portfolio. After that, operator will ask business customer about his impression of using this service. If the specific need arises from this questionnaire, operator will try to satisfy this need by changing existing solution to the new one that is more adjusted to the business customer needs. In this way, the operator develops and grows together with the business user to whom he provides his services.

Each company has its own specific requirements concerning networking and connecting with their central location and

pulling data from a central server, as well as in terms of Internet access, which must comply with certain rules and restrictions. To satisfy these requirements company must have highly specialized persons or whole team to deal with establishment of service, monitoring and solving every issue. Enterprise retail companies have large number of branches geographically dislocated throughout the country and abroad. Because of all these facts, managing of the telecommunication services within the company is becoming a challenging issue. More often business customers want to shift the responsibility for managing telecommunication services to the service provider – network operator [3].

The second major request concerning retail companies is higher availability of head office location for access by the branches. Branches have need to achieve sensitive and permanently necessary information and to be always updated so they could have business continuity and earn their income.

Critical issue for enterprise retail companies is also voice of the customer measured through Quality of Experience (QoE). In order to satisfy their customers, the company itself need to be satisfied as a customer, concerning the telecommunication services. Because of that, QoE is becoming crucial for delivering good service and stay competitive.

Many retail companies are evolving and spreading their business by opening their stores across the country even abroad, and every competitive telecommunication operator need to be ready to respond to all requests they have. According to our experience with one of the biggest network operator, we manage to signify some of the major requests that big retail company have. Further, in this paper will be describe one example of retail company and its needs and how the operator is adjusting to this. As a result, one tailor made managed solution will be created and modified for specific customer. Moreover, in the last chapter will be pointed the importance of measuring QoE and its application in creating new solutions for connection business customers to the Internet.

II. MANAGED SOLUTIONS FOR BUSINESS CUSTOMERS

A. The advantages of managed solutions

Sometimes it is better for both, network operator and for business customer, to have predefined business solutions for connecting to the Internet. In this way business customers who do not have specialized person or persons in IT field, could easier choose suitable packet of services for their business. This packet very often does not fully match customer requirements, but it is the closest one to ideal packet of telecommunication service for specific business customer.

[1]Zlata Radak is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, Belgrade, Serbia, E-mail: zlata.radak@sbb.co.rs.

[2]Goran Marković is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, Belgrade, Serbia, E-mail: g.markovic@sf.bg.ac.rs

On the other side, some specialized enterprises like retailers, have their own specialized team of competent persons who know exactly what are their needs and the operator have to adjust according to their needs. They expect from the operator to provide and manage the solution for connecting to the Internet and finally to change the solution if the specific need occurs.

The leading operator has a platform that could on the highest level provide innovative and high quality ICT services and solutions for business customers. The service portfolio is based on a regional multi-service network, supported by its own optical infrastructure, modern telecommunications equipment and modern data centers. The primary goal for the operator is to provide a comprehensive portfolio of services to all business customers regardless of the type of activity or size of company. Beside all of this, it should take into account the specific requirements of individual companies and provide flexible services and customized business solutions [3].

In order to achieve high-quality performance of the services, the operator provides fiber optic infrastructure at the all customer locations, head office and branches. These services are designed for organizations that need a reliable, highly available and fast Internet service. Due to the increasing needs of large companies for symmetric Internet access download and upload, speeds in this type of services are identical. An important feature is the SLA (Service Level Agreement), which guarantee the required level of service availability. The optical infrastructure provides the ability to add other services that the operator has in its services portfolio easily by virtual local area network (VLAN) separation. Equipment configuration and activation of the service need to be provided by a team of experts on the operator side, on the equipment collocated at the customer premises. This is so called managed service.

The main advantages of managed solutions must be defined and presented to the customer. Clearly it must be pointed out the cost savings as the most important fact as well as the improvements in the quality of service. Some of the main advantages of managed solutions could be identified as following:

1. Cutting costs: all the equipment is provided and managed by the operator,

2. Guaranteed technical parameters: providing service availability and performance quality by monitoring the Service Level Agreement (SLA) and Quality of Services (QoS) parameters,

3. Dedicated technical support: the engineers in the Network Operations Center (NOC) perform proactive monitoring of the managed equipment that is collocated at customer premises.

4. Compatible solution: the business customer gets from the operator managed solution that is easy to upgrade on the existing network design,

5. Technical support 24/7/365: They integrate, install and maintain all equipment and its configuration,

6. Business customer outsource solution for the connecting on the Internet: This way they have to focus on their core business [3].

Advantages and cost savings achieved through the implementation of a managed services are both short-term (reflected in the initial investment) and long term (lower monthly subscription fee in respect of costs before the introduction of managed services) [3]. The tendency of savings in terms of telecommunications and the growing demands in terms of quality leads to the development of innovative model for connecting the large number of locations through a Managed Layer 3 Virtual Private Network (L3VPN) with a central location. This involves setting up the Internet connection at the central location (head office) and create virtual connections - tunnels (L3VPN) through the operator's network to all other branches. In this way, it is achieved that IT administrator at the central location can easily and securely access data stored on remote locations and vice versa.

As a leading regional provider of information and communication technologies (ICT), the observed operator's network is able to offer to the business customers complete IT solutions and systems for unified communications between head office and all branches with no concern for the connections and equipment. Specific demands of the retail company will be explained in next chapter, but first the network configuration will be explained and the specific requirement will arise itself.

B. Enterprise retail network configuration

In this chapter we will present one specific enterprise retail company and their requirements for the connecting on the Internet. Retail company has 50 branches that are connected with the head office through firewalls. Head office communicates with the central servers through leased lines (Fig. 1). It is obvious that head office represents key point of the network, and it is necessary to protect this location as much as possible.
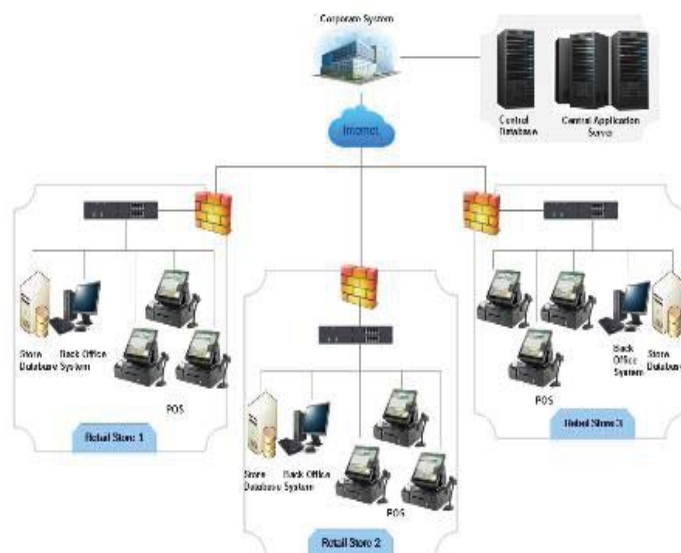
Fig. 1. Organization of the enterprise retail network

At first, operator creates managed L3VPN solution for connecting all location in star topology to the central location – head office. Operator provides optical cable to the all locations of the business customer in order to connect them on his network. After that, he provides manageable equipment – routers at all location of the business customer. Also, the operator configures manageable routers according to the

policies of the company and specific requests from the retail company. After putting the equipment in the operation, network operator documents in contract some Service Level Agreement (SLA) parameters that will be guaranteed. Some of these parameters related to the availability in time are illustrated in the Tab.1.

The availability is defined on a monthly basis, representing a percentage of the time in which the service was fully operational. For different types of SLA, the operator guarantees the following levels of availability [6]. Retail company chooses the Gold SLA that has the maximal availability offered by the operator (Tab.1).

Table I Availability levels according to the policy of the observed network operator

Type of the SLA	Availability (A)
Standard	98,5 %
Silver	99 %
Gold	99.5 %

To calculate the availability of a service, it is important to define the following terms:

Billing Period (OT): represents the period in which the SLA parameters are measured. It is expressed in hours (h). The accounting period is the same for all three types of SLA, which depends only on the number of days in the month, for the month that lasts 30 days OT (h) = 24x30 = 720.

Total loss of service (TTSL): Total loss of service is the total time the service was inoperative at the level of one month. This time is obtained when the service repair time (TTR) is summarized for all Trouble Tickets of critical severity (Severity 1) at the level of the month for which the billing is performed [6].

The availability of the service is calculated using the following formula:

$$TTSL = \Sigma TTR \,;$$
$$A\,(\%) = (OT-TTSL)/OT*100 \qquad (1)$$

Example: As an example we calculate the availability (A) during the month that has 30 days during which there were two critical incidents (TTID-1 and TTID-2), with the correction times for these incidents being TTR (TTID-1) = 2h and TTR (TTID-2) = 1h.

$$OT = 30x24h = 720$$
$$TTSL = TTR(TTID\text{-}1) + TTR(TTID\text{-}2) = 3h$$
$$A\,(\%) = (720\text{-}3)/720*100 = 99.58$$

This calculation shows that requested availability level is achieved because it is more than it is guaranteed by the contract i.e. 99,5% (Tab.1.). But still this is much time for the presented company, to stay without the connection to the central location and they requested from the operator to redefine the solution for them so they could get better performances, according to availability, for the head office. The operator conducted a questionnaire in a specialized team for telecommunication services of a retail company. As a result of customer requests, the operator suggests and implements tailor made solution with

redundant backup wireless link on the head office location in order to incise the level of availability.

C. Tailor made solution with redundant backup wireless link on the head office location

Although the presented managed solution is very reliable and well designed, the business customer still needs an additional level of security in the form of a separate and completely redundant link. Since the configuration of network equipment is not a focus of the operation of the retail company, it is also necessary to have equipment that will be preconfigured by transferring all services from the primary to the backup link if an interruption occurs. Bearing in mind that large amounts of data are transmitted in 24/7 working time and the backup link should have guaranteed quality and symmetric flow. In order to provide a backup link, it is suggested for an operator that an additional Internet link be provided to the user through another access technology, in this case of wireless technology, which would be connected to another optical hub of the operator [6]. Its realization is shown in Fig. 2.
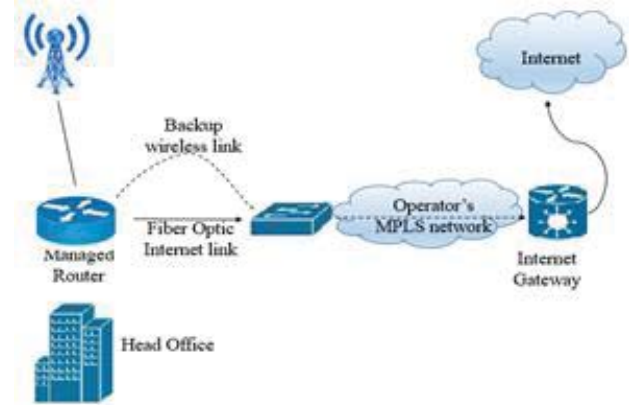
Fig. 2. Managed solution for connecting to the Internet with wireless backup link for head office [6]

D. QoE as a voice of the customer

The retail company that we observe in this paper every year confronts IT budgets that are increasingly limited. It is requested from IT department to meet the needs of companies that are more specific and to cut the budget. At other side demands for higher availability is constantly growing, demanding as much secure connection as possible. In this specific case, cloud services are imposed as a logical step forward.

Cloud services for the business customer mean the rental of IT and communications infrastructure and services from the network operator. This service excludes making capital investments in their own infrastructure, software and human resources. This is becoming more profitable and more common business model of companies and organizations around the world. They are best suited for companies that have a need for implementation of new IT services. Services are flexible and formed towards the specific needs [3].

The network operator conducted the survey and according to the answers find out about the future growth of the retail business customer, as well as budget cuts. According to this new findings, operator is on one side faced with need of the customer for more reliable head office location, and at the other side requests for virtualization and reduction of costs. According to this findings, network operator has to propose the improvement of the existing solution and more suitable tailor made managed solution for the customer.

E. Tailor made managed solution for connecting remote locations through L3VPN with virtualized central location in the operator's Data Center

Retail business company needs maximal reliability and availability of the head office location. By choosing predefined managed solution they get the availability of 99,5%. The second, tailor made managed solution with redundant backup link increases the availability, because when main link is not operational, they have a redundant wireless link and the same percentage of availability as the mail link. But, specific demand is placed to the operator to cut the cost and increase the availability at the same time.

The specific requirement is maximal availability of head office location. Also, it is very important to constantly run updates so branches have business continuity and earn their income. At the head office the server through a special application keeps track of the state of the warehouses, current prices and updated information as soon as changes occur. Current and accurate information are necessary at each branch office location and at any time. In order to achieve higher availability of head office location network operator takes into consideration to install one or more back-up links. The introduction of multiple optical links from different operators has proven to be an unjustified costly solution. Suggested solution by the expert presales team is to take a central server with all applications and data and virtualize it in operator's data center (Fig. 3) [3]. This solution achieves uptime reaches the number 99,999% and this is practically maximal uptime that could be guaranteed by the operator. Head office and branch offices have access to this virtualized server through managed L3VPN tunnel. All these results are exactly what the business customer was looking for.
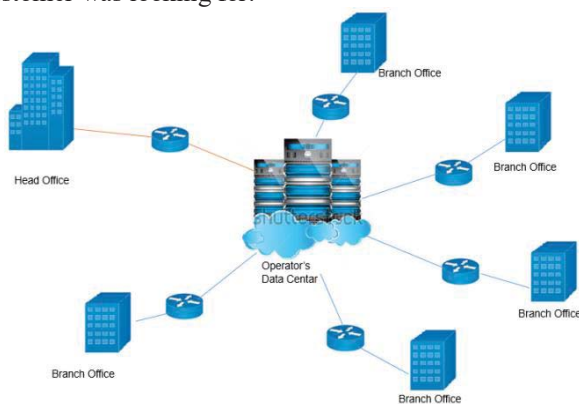
Fig. 3. Managed L3VPN Solution with central location in the operator's Data Centar [3]

In order to meet the customer needs, the network operator has made tailor made managed solutions (Fig. 3). In the data center is installed adequate virtual router with all the necessary applications and data. This virtual machine establishes a managed connection through L3VPN from a central location to all remote locations (had office and branch offices) which also have the managed equipment that was configured and monitored by the operator's expert team. Routers communicate over L3VPN tunnel that is established between the central site and all remote locations [3].

Thus configured solution ensures maximum performance for business customer and absolute availability of virtual servers in the data center. Also, it offers various advantages for administrator of the business customer side. Tailor made managed solution simplifies the work of administrator and allows him to devote additional time improving business applications and to introduce a new system in order to improve sales and facilitate the work for his employees [3].

III. CONCLUSION

Virtualization of the head office location in the network operator's data center provide the maximal possible availability of the central location and hence it is located in modernly equipped data center, offering maximal protection. This tailor made managed solution provides connection from the head office of the customer to the virtual central location in the network operator's data center. The virtual site is also connected to the branch offices providing them with maximum data availability. Costs and engagement of technical resources have not been taken into consideration in this paper. For the next survey, it is necessary to perceive the costs and justification of technical resources exploitation for service improvement. The proposal is to achieve this analysis using the Quality Function Deployment (QFD) method.

REFERENCES

[1] J. Walsh, S. Godfrey, "The Internet: A New Era in Customer Service", *European Management Journal* Vol. 18, No. 1, pp. 85–92, 2000
[2] K.U.R Laghari,. N. Crespi, K. Connelly: "Toward total quality of experience: a QoE model in a communication ecosystem", IEEE Commun. Mag. 50(4), 58–65, 2012
[3] Z. Radak, N. Gospić: "Transformation of the existing operator to the service provider for business customers", ICEST, Conference Proceedings, Macedonia, 2016
[4] A. Chakravarty, A. Kumar, R. Grewal: "Customer Orientation Structure for Internet-Based Business-to-Business Platform Firms", Journal of Marketing, 2014
[5] V. Kumar, A. Pansari: "Competitive advantage through engagement", Journal of Marketing Research In-Press, 2015
[6] Z. Radak, N. Gospić, G. Marković: "Predlog rešenja za pouzdan pristup Internetu", Tehnika, 2017.

SIGNAL PROCESSING
&
DIGITAL IMAGE PROCESSING

Design of Narrow Band-Stop Recursive Filter Based on Third Order Phase Corrector Application

Ivan Krstić[2], Miloš Živković[1], Ivana Kostić[1] and Goran Stančić[1]

Abstract —A method for design and software realization of infinite impulse response (IIR) digital filter with one notch frequency is presented in this paper. The proposed configuration has parallel nature with a pure delay in one path and third order all-pass phase corrector in another. The resulting filter fulfills all predefined specifications i.e. for arbitrary given maximal pass-band attenuation, band-stop boundary frequencies are symmetrical about notch frequency. The efficiency of the presented method is illustrated by few examples. All results are obtained using Matlab® package.

Keywords – Notch filter, phase corrector, all-pass filters, coupled all-pass, phase approximation.

I. Introduction

Digital narrow band-stop filters are widely used in applications where it is necessary to eliminate particular frequency component from input signal while the other frequency components need to remain untouched. Typical digital notch filters application areas include bio-medical engineering, seismology, speech processing, transmission of data through telephone channels, etc. The single-frequency interference can be removed by a notch filter tuned to that particular frequency. To remove unwanted signal components at several frequencies, it is possible to implement proposed filter in cascade configuration with adequately tuned their notch frequencies. In general, notch filters could be designed as recursive (Infinite Impulse Response-IIR) [1-2] or as non-recursive (Finite Impulse Response-FIR) structures [3]. In case of FIR filters it is easy to achieve the exact linear phase thanks to existing transfer function coefficient's symmetry. To achieve the given magnitude specifications FIR filter demands significantly higher order compared to corresponding IIR filter [4].

The transfer function of narrow band-stop recursive filter could be determined by applying bilinear transform to the starting analog prototype filter of second order [3], [5]. In that case, the location of central frequency of the band-stop (notch frequency) as well as band-stop boundary frequencies at which attenuation reaches 3 dB value are the main parameters for the filter design [6]. However, the given specifications could not be achieved by this approach, i.e. boundary frequencies of the band-stop do not exhibit symmetry about predefined notch frequency. This deviation from an ideal

[1]Goran Stančić, Miloš Živković and Ivana Kostić are with the Faculty of Electronic Engineering, University of Niš, ul. Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: goran.stancic@elfak.ni.ac.rs.
[2]Ivan Krstić is with the Faculty of Technical Sciences at University of Priština, 38220 Kosovska Mitrovica, Serbia.

solution is more evident in cases where the notch frequency is far from central frequency $\pi/2$, which corresponds to frequency $F_{sampling}/4$.

The proposed method is based on application of the third order phase corrector i.e. all-pass filter. The main goal is to achieve all given magnitude specifications. The realized notch frequency will be positioned exactly at predefined location, while at the same time cut-off edge frequencies are symmetrical regardless the value of the notch frequency. Practically, one wants full control of the magnitude characteristic at three frequencies. That is the main reason the third order polynomial is inevitable in all-pass transfer function. The coupled all-pass structure offers convenient way to solve the problem of filter design thanks to straightforward dependence of filter's magnitude and phase characteristic of corresponding all-pass sub-filter. Problem of design of resulting filter magnitude is easy to reformulate as the all-pass filter phase approximation problem.

The rest of the paper is structured as follows. In the Section II, the basic relations are derived according to which it is possible to obtain coefficients of the phase corrector transfer function. Comparison of the standard solution notch filter and notch filter obtained by the proposed procedure is done in Section III. Standard solution notch filter is realized by parallel connection of two all-pass filters with approximately constant phase in all pass-bands. The results of the simulation of designed filter performance are given in Section IV along with difference equations implemented in the Matlab® in order to obtain output of the notch filter.

II. Problem Definition

Specifications of notch filter magnitude characteristic precisely define the location of the notch frequency ω_n as well as band-stop width B_w. According to this available information it is possible to obtain band-stop lower ω_l and upper ω_r edge frequencies

$$\omega_l = \omega_n - B_w/2$$
$$\omega_r = \omega_n + B_w/2 \tag{1}$$

based on the mentioned symmetry. Given parameters uniquely define maximal attenuation in both pass-bands a. Usually, for attenuation at cut-off frequencies the value of 3 dB is adopted. In this paper, the proposed method allows arbitrary positive value for maximal attenuation in pass-bands to be chosen. The transfer function of the narrow band-stop recursive filter is given with

$$H(z) = \frac{1}{2}(z^{-1} + P(z)). \tag{2}$$

The transfer function of a stable phase corrector of the third order $P(z)$ is given with

$$P(z) = z^{-3} \frac{D(z^{-1})}{D(z)}, D(z) = 1 + p_1 z^{-1} + p_2 z^{-2} + p_3 z^{-3} \qquad (3)$$

in parallel with first order delay line, as it is shown in Fig. 1.

Fig. 1. Coupled all-pass notch filter

By substituting Eq. (3), alongside with $z = \exp(j\omega)$, into Eq. (2), after simple mathematical manipulations the magnitude characteristic of the notch filter could be obtained

$$|H(e^{j\omega})| = \left| \frac{p_1 + (1+p_2)\cos\omega + p_3\cos 2\omega}{D(e^{j\omega})} \right|. \qquad (4)$$

From the other hand, the following relation

$$p_1 + (1+p_2)\cos\omega + p_3\cos 2\omega = D_{Re}(\omega)\cos\omega - D_{Im}(\omega)\sin\omega \qquad (5)$$

holds, where with $D_{Re}(\omega)$ and $D_{Im}(\omega)$ are represented real and imaginary part of the frequency response of the stable, minimal phase filter $D(z)$, respectively. Based on trigonometric identities

$$\cos(\arg D(e^{j\omega})) = \frac{D_{Re}(\omega)}{|D(e^{j\omega})|},$$
$$\sin(\arg D(e^{j\omega})) = \frac{D_{Im}(\omega)}{|D(e^{j\omega})|} \qquad (6)$$

Eq. (4) could be rewritten in more compact form

$$|H(e^{j\omega})| = |\cos(\omega + \varphi(\omega))| \qquad (7)$$

where next notation is introduced

$$\varphi(\omega) = \arg D(e^{j\omega}) =$$
$$-\arctan \frac{p_1\sin\omega + p_2\sin 2\omega + p_3\sin 3\omega}{1 + p_1\cos\omega + p_2\cos 2\omega + p_3\cos 3\omega}. \qquad (8)$$

Taking into account the fact that function $\omega + \varphi(\omega)$ has value equal to zero for $\omega = 0$, and value π for $\omega = \pi$, coefficients p_1, p_2 and p_3 of the transfer function given with (2) need to be obtained such a way, to hold

$$|\omega + \varphi(\omega)| \le \varepsilon = \arccos 10^{-a/20}, \omega \in [0, \omega_l]$$
$$|\omega + \varphi(\omega) - \pi| \le \varepsilon, \omega \in [\omega_r, \pi] \qquad (9)$$

in both lower and upper pass-band.

According to the Eq. (4), one could conclude that amplification of the notch filter at frequency ω_n could be zero if the value of parameter p_1 is defined according to following equation

$$p_1 = -[p_3\cos 2\omega_n + (1+p_2)\cos\omega_n]. \qquad (10)$$

Incorporating Eq. (10) in Eq. (8) after some mathematical manipulations, expression

$$p_2[\sin(\varphi(\omega) + \omega)(\cos\omega - \cos\omega_n) +$$
$$\cos(\varphi(\omega) + \omega)\sin\omega] + p_3[\sin(\varphi(\omega) + \omega) \cdot$$
$$(\cos 2\omega - \cos 2\omega_n) + \cos(\varphi(\omega) + \omega)\sin 2\omega] \qquad (11)$$
$$= \cos\omega_n \sin(\varphi(\omega) + \omega) - \sin\varphi(\omega)$$

is obtained, which give mutual dependence among parameters p_2, p_3 and the phase characteristic $\varphi(\omega)$. In order to adequately obtain the values of unknown parameters p_2 and p_3 (what would uniquely define the value of parameter p_1 according to Eq. (10)), it is necessary to know exact values of phase characteristic $\varphi(\omega)$ at $M \ge 2$ frequencies. Based on that data, a system of M equations could be formed with two unknowns. For $M = 2$, the system of equations has the unique solution, while for $M > 2$ one is forced to find least-square (LSE) solution. To remind a reader that magnitude of resulting couple all-pass filter and phase of all-pass sub-filter are mutually related. The specifications adopted for cut-off could be written in the form

$$\omega_l + \varphi(\omega_l) = \varepsilon$$
$$\omega_r + \varphi(\omega_r) = \pi - \varepsilon. \qquad (12)$$

Eq. (12) define the precise location of two points at phase characteristic $\varphi(\omega)$ with coordinates $(\omega_l, \varepsilon - \omega_l)$ and $(\omega_r, \pi - \varepsilon - \omega_r)$. The system of equations generated by using Eq. (11) applied to that two particular points leads to the unique solution for parameters p_2 and p_3:

$$\begin{bmatrix} p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sin\varepsilon\cos\omega_n - \sin(\varepsilon - \omega_l) \\ \sin\varepsilon\cos\omega_n - \sin(\varepsilon + \omega_r) \end{bmatrix} \qquad (13)$$

where

$$a_{11} = \sin(\varepsilon + \omega_l) - \sin\varepsilon\cos\omega_n$$
$$a_{12} = \sin(\varepsilon + 2\omega_l) - \sin\varepsilon\cos 2\omega_n$$
$$a_{21} = \sin(\varepsilon - \omega_r) - \sin\varepsilon\cos\omega_n \qquad (14)$$
$$a_{22} = \sin(\varepsilon - 2\omega_r) - \sin\varepsilon\cos 2\omega_n.$$

The obtained parameters not necessarily fulfill the conditions given by Eq. (9). From the other hand, using the software package Wolfram Mathematica, it was numerically justified that filter which coefficients are obtained by Eqs. (13) and (10), would met specifications given by Eq. (9) in a wide range of maximal allowed attenuation i.e. $a \in [0.001, 3]$ dB, for arbitrary notch frequency location from a set $\omega_n \in [0.1\pi, 0.9\pi]$ rad and with band-stop width $B_w \in [0.001\pi, 0.1\pi]$ rad. In other words, it holds

$$\min \frac{20\log_{10}|\cos(\omega + \varphi(\omega, B_w, \omega_n, a))|}{a} > -1, \qquad (15)$$
$$\omega \in [0, \omega_l) \cup (\omega_r, \pi].$$

III. COMPARISON WITH STANDARD NOTCH FILTER

The transfer function of standard digital notch IIR filter [1], based on application of second order phase corrector is given with:

$$H_c(z) = \frac{1}{2}\left[1 + \frac{k_2 + k_1(1+k_2)z^{-1} + z^{-2}}{1 + k_1(1+k_2)z^{-1} + k_2 z^{-2}}\right] \quad (16)$$

where parameters k_1 and k_2 are given with

$$k_1 = -\cos\omega_n, \quad k_2 = \frac{1 - \tan B_w/2}{1 + \tan B_w/2} \quad (17)$$

The maximal band-pass attenuation has value of 3 dB.
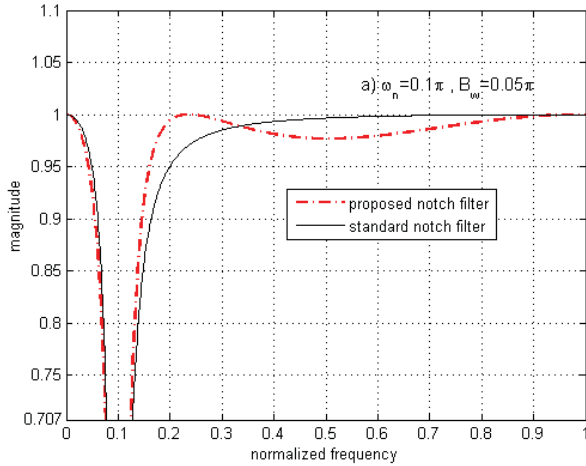
Fig. 2. Magnitude of proposed and standard notch filter with notch frequency $\omega_n = 0.1\pi$ and $B_w = 0.05\pi$
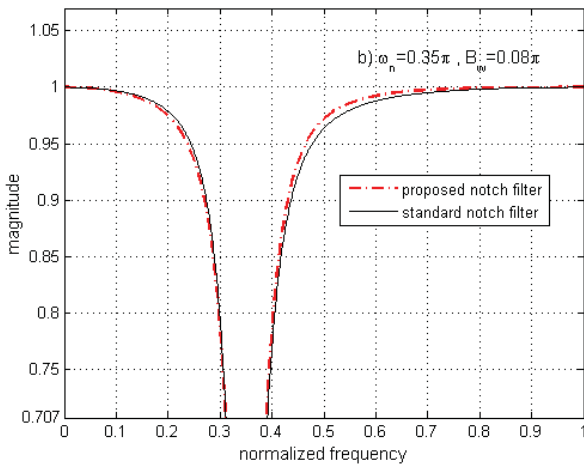
Fig. 3. Magnitude of proposed and standard notch filter with notch frequency $\omega_n = 0.35\pi$ and $B_w = 0.08\pi$

Magnitude characteristics of proposed notch filters alongside with standard solution, for different notch frequencies are presented in the following three figures. Fig. 2 corresponds to filters with notch frequency $\omega_n = 0.1\pi$ with band-stop width $B_w = 0.05\pi$. Figs. 3 and 4 show results for filters with specifications $\omega_n = 0.35\pi$, $B_w = 0.08\pi$ and $\omega_n = 0.8\pi$, $B_w = 0.1\pi$, respectively. In order to facilitate comparison, all presented filters have the same attenuation of 3 dB at boundary frequencies.

According to the obtained results, displayed in given three figures, one can conclude that improvement in magnitude characteristic is more significant if the notch frequency is close to π, or even better results for notch frequency in vicinity of zero. That fact provides opportunity to increase

sampling rate in order to improve filter efficiency in a simple way. The enhancement is negligible for filters with notch frequency which is located at central zone (about $\pi/2$). To remind the reader that this improvement is natural, behalf to slightly higher order of implemented filter. Standard notch filter solution involve the second order all-pass sub-filter in one of parallel branches, while the proposed filter apply third order all-pass sub-filter. One more unknown filter's coefficient offers the opportunity to expand the system of equations in order to fulfill the all given notch filter specifications.
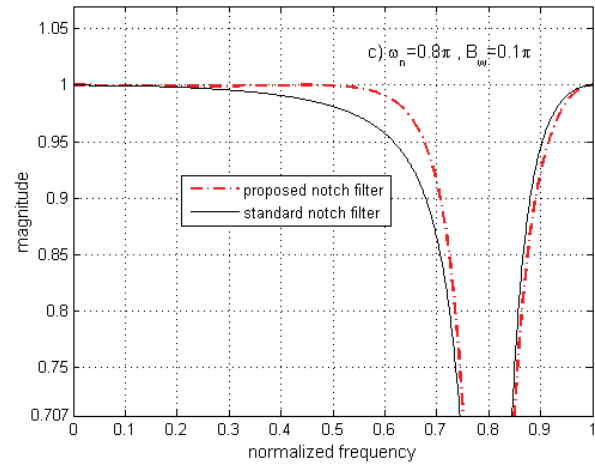
Fig. 4. Magnitude of proposed and standard notch filter with notch frequency $\omega_n = 0.8\pi$ and $B_w = 0.1\pi$
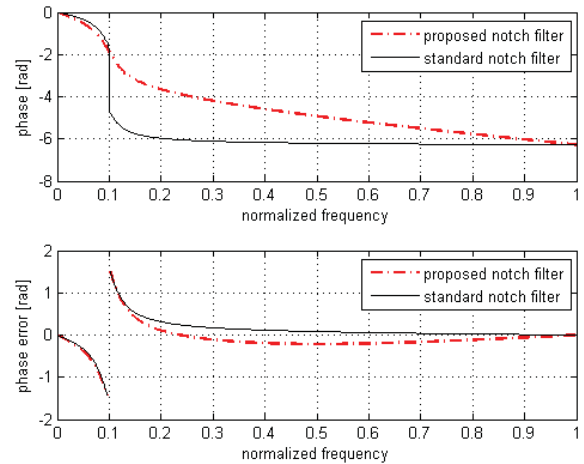
Fig. 5. Phase and phase error of proposed and standard notch filter with notch frequency $\omega_n = 0.1\pi$ and $B_w = 0.05\pi$

Phases of the proposed and standard notch filter and corresponding phase errors are displayed in Fig.5. Standard filter's phase approximates constant phase in both pass-bands, while proposed filter's phase approximate piecewise linear phase. It is not possible to obtain filter with only one notch frequency with standard solution if order of all-pass sub-filter from one of parallel branches is higher than 2. Taking into account the fact that constant equal to one is in one path (with zero phase) choosing all-pass filter of order $2k$ for another path will lead to resulting filter with k notch frequencies. In

other words, it is not possible to improve magnitude characteristic of standard filter with one notch frequency by choosing higher order all-pass sub-filter. On the other hand the proposed filter configuration does not have such constraint. Increasing of order of all-pass sub-filter has to be simply followed by increasing of delay line order. If difference in their orders remain to be equal to 2, resulting filter possess only one notch frequency. Choosing sub-filters of higher order give opportunity to reduce simultaneously the phase and magnitude error [8].

IV. SOFTWARE REALIZATION AND THE SIMULATION RESULTS

The Matlab® software package is used for design and realization of the proposed narrow band-stop filter. Sinusoidal noise with amplitude 0.2, at power-line frequency $F_n = 50$ Hz, is superimposed on the electrocardiogram (ECG) signal $s[n]$ downloaded from the database MIT-BIH [7], as it is given in Eq. (18). All available signals in MIT-BIH database are recorded after digitalization using sampling frequency $F_s = 360$ Hz

$$x[n] = s[n] + 0.2\sin(2\pi n F_n / F_s) \qquad (18)$$
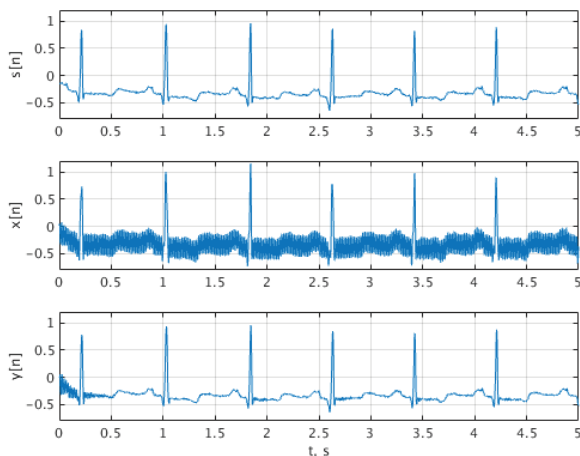
Fig. 6. ECG signal $s[n]$, ECG signal with added sinusoidal noise $x[n]$ and filtered signal at output of notch filter $y[n]$

Specifications of the digital filter in z domain are: $\omega_n = 2\pi F_n / F_s = 5\pi/18$, while for band-stop width and maximal attenuation in pass-bands are adopted $B_w = 0.02\pi$ and $a = 1$ dB, respectively. Software realization of the designed filter is based on linear difference equation derived according to Eq. (2)

$$y[n] = \frac{1}{2}(x[n-1] + w[n]),$$
$$w[n] = x[n-3] + \sum_{i=1}^{3} p_i(x[n-3+i] - w[n-i]) \qquad (19)$$

In Fig. 6, ECG signal $s[n]$ which corresponds to 100th sample in MIT-BIH database is shown, together with corrupted version of this signal $x[n]$ after addition of sinusoidal noise. Filtered version of the signal at output of proposed notch filter $y[n]$ is also presented in Fig. 6. It is evident that designed

notch filter successfully eliminated sinusoidal noise at power-line frequency from ECG signal. In the steady state, output signal $y[n]$ is noise free.

V. CONCLUSION

Design and software realization of IIR digital filter with one notch frequency is presented in this paper. The proposed filter is realized as parallel structure with pure delay in one path and all-pass sub-filter in another path. The existing methods for design of notch filter do not deliver solution with cut-off frequency symmetry about notch frequency. The proposed filter fulfills all predefined specifications i.e. for arbitrary given maximal pass-band attenuation, band-stop boundary frequencies are symmetrical about notch frequency. The efficiency of the presented method is illustrated by filtering ECG input signal corrupted with sinusoidal noise at power-line frequency. The proposed filter has approximately linear phase in pass-bands and minimal order. The phase and magnitude error of one notch filter could be further reduced by increasing simultaneously the order of delay element and all-pass sub-filter.

ACKNOWLEDGEMENT

REFERENCES

[1] Y. V. Joshi, S.C. Dutta Roy, "Design of IIR digital notch filters", Circuits, Systems and Signal Processing, vol. 16, no. 4, pp. 415-427, 1997.

[2] G. Stančić, S. Nikolić, "Design of digital recursive notch filter with linear phase characteristic", 11th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS), pp. 69-72, 2013.

[3] S. D. Roy, B. Kumar, B. Jain, "FIR notch filter design: A review", Facta Universitatis Series, vol. 14, no. 3, pp. 295-327, 2001.

[4] K. Aboutabikh, I. Haidar, N. Aboukerdah, "Design and implementation of a digital FIR notch filter for the ECG signals using FPGA", IJARCCE, vol. 5, no. 1, pp. 452-456, 2016.

[5] K. Hirano, S. Nishimura, S. Mitra, "Design of Digital Notch Filters", IEEE Transactions on Communications, vol. 22, no. 7, pp. 964-970, 1974.

[6] S. Nikolić , I. Krstić, G. Stančić, "Noniterative design of IIR multiple-notch filters with improved passband magnitude response", Int. Journal of Circuit Theory and Applications, vol. 46, no. 12, pp. 2561-2567, 2018.

[7] G. B. Moody, R. G. Mark "The impact of the MIT-BIH Arrhythmia Database", IEEE Eng in Med and Biol vol. 20, no. 3, pp. 45-50, 2001.

[8] G. Stančić, S. Nikolić, "Digital linear phase notch filter design based on IIR all-pass filter application", Digital Signal Processing, vol. 23, no.3, pp. 1065-1069, 2013.

Detection of Membrane Drying at Electrochemical Hydrogen Compressors

Gjorgji Nusev[1,2], Gregor Dolanc[1], Đani Juričić[1] and Pavle Boškoski[1]

Abstract – **Electrochemical hydrogen pump (EHP) is capable of extracting hydrogen from miscellaneous gases and compress it to very high pressures. Since during its operation it does not produce water the performance is heavily dependent on humidity of the membrane. Online humidity estimation is performed by means of performing electrochemical impedance spectroscopy (EIS).**

Keywords – **EHP, EIS, Wavelets, Impedance, DRBS.**

I. Introduction

Renewable energy sources such as wind turbine farms and solar power plants are ubiquitous in our energy systems. Despite their indisputable positive environmental impact, their main drawback is the intermittent power output. In many cases, the generated energy has a significant mismatch to the demand on the grid. For that reason the need for energy storage installations, that are capable of storing vast amounts of energy in times of surplus while being capable for instant on-demand delivery is of a great importance. Currently there are two competing technologies that are addressing these issues: batteries and fuel cells. Our focus in this paper will be on the latter.

Fuel cells operate on the chemical energy stored in form of hydrogen molecules. Hydrogen is one of the most abundant elements and has great properties of being an energy-rich gas and highly versatile. This makes hydrogen as one of the most suitable candidate to be used as an efficient energy storage media [1, 2]. Compressing the hydrogen to high pressures while maintaining high purity is a challenging process and consumes a lot of energy. There are two general categories of hydrogen compressors: mechanical, which uses adiabatic process and non-mechanical, which mainly uses isothermal process for compression.
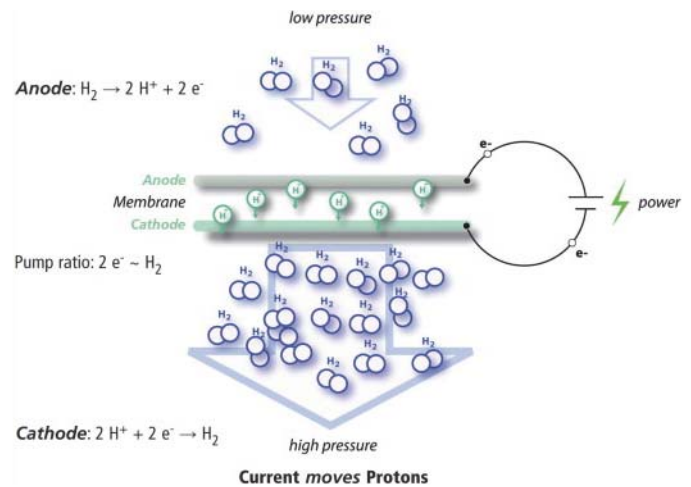
Fig. 1 Basic operational concept of how EHP works. I uses an external voltage source to efficiently pull only hydrogen protons through the proton conductive membrane [4].

Non-mechanical compressors such as electrochemical hydrogen pumps (EHP) uses electricity to extract hydrogen from miscellaneous gas mixtures and compress it in order to meet different application requirements. The lack of moving parts makes them more efficient when compared to the conventional mechanical compressors which are noisy, can induce impurities and less effective [3]. The structure of EHP is similar to that of a Proton Exchange Membrane Fuel Cells (PEMFC). The basic principles about how EHP works are shown in the Fig. 1.

Hydrogen molecules enter the anode side (the low pressure side) where are oxidised to protons and electrons. The applied voltage force the electrons through an external circuit while the protons are driven through the membrane to the cathode side where are recombined to form again a molecular hydrogen. As a result of the external voltage, the pressure at the cathode side increases with the increased number of transported hydrogen molecules.

Despite EHP advantages, they have mainly been used as a diagnostics test for measuring crossover in fuel cells. Ströbel et al. [5] was the first one to recognize the possibilities of the PEM-based structure for EHP. Hao et al. [6] performed electrochemical impedance spectroscopy (EIS) on EHP with included internal humidifier with dead-end anode. Nguyen et al. [7] has shown how the temperatures of the humidifier and the temperature of the stack are influencing the performance of the stack. From the previous reported analysis, they all address the problem of controlling the humidity of the membrane, which is essential for the transportation of protons through the membrane and is tightly connected to the overall performance of the EHP. On the other hand, shortage of humidity reduces the proton conductance. Due to the fact that EHP does not

[1]Gjorgji Nusev, Pavle Boškoski and Gregor Dolanc are with the Department of system and control - Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia, E-mail: gjorgji.nusev@ijs.si, pavle.boskoski@ijs.si, gregor.dolanc@ijs.si.

[2]Gjorgji Nusev is with Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia.

produce water, when compared to the PEM fuel cells, it has to be delivered via an external humidifier.

Online monitoring the level of humidity of the membrane by performing direct measurements is almost impossible due to the physical inaccessibility of the membrane. The only possible way to perform online estimation of the humidity is by performing an EIS. Classical way of performing EIS includes single sine excitation signals. As a result of this the characterisation process is often time consuming and lacks of accuracy especially in the low frequency region, which can be crucial and reduce the applicability for online applications. Using stochastic excitation, for instance discrete random binary sequence (DRBS), alleviates the aforementioned problems. This method has successfully been tested on PEM fuel cell [8]. For the purpose of this paper two experiments were performed, one in normal operating mode when the membrane was well humidified and an other when EHP was operating in drying mode. The experiments were performed in the frequency band between 0.1 Hz and 1000 Hz.

II. WAVELET TRANSFORM FOR IMPEDANCE CALCULATION

The conventional frequency domain signal analysis performed with the Fourier transform, provides a detailed picture of the frequency components present in the signal but without any information regarding their time occurrence and duration. Time-frequency analysis offers a solution to this problem thus providing the information about the temporal details as well. Typical examples are the Short-time Fourier transform, Wigner-Ville distribution, wavelet transform etc.

Regardless of the selected method there is a theoretical limitation on the joined time-frequency resolution. Unlike other methods, the wavelet transform enables flexible selection of the desired time-frequency resolution by introducing the concepts of scaling. Wavelet transform is based on a set of specifically designed functions called wavelets. The continuous wavelet transform (CWT) of a square integrable function $f(t) \in L^2(\mathbb{R})$ is defined as [9]:

$$Wf(s,u) = \int_{-\infty}^{\infty} f(t)\psi_{u,s}^*(t)dt$$

where wavelet function $\psi(t)$ is scaled and translated by introducing two additional parameters u and s:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$$

The key parameter in CWT is the selection of the wavelet function. EIS analysis requires information about the amplitude and phase of the excitation and response signals. Therefore, the CWT is performed with a complex wavelet function, in our case the Morlet wavelet [10]:

$$\psi(t) = \pi^{-\frac{1}{4}}\left(e^{-j\omega_0 t} - e^{-\frac{\omega_0}{2}}\right)e^{-\frac{t^2}{2}}$$

where ω_0 is the central frequency, whose value is linked to the time-frequency resolution of the wavelet. The link between the Morlet wavelet's scale parameter s and the actual frequency is through the following relation:

$$\frac{1}{f} = \frac{4\pi s}{\omega_0 + \sqrt{2 + \omega_0^2}}$$

The time and frequency localisation is determined through the parameters u and s respectively. The scale parameter s acts as a dilatation factor of the mother wavelet, which determines the frequency region of the wavelet analysis. The translation parameter u defines a time location where the instantaneous phase and amplitude are estimated. More details regarding the properties of the Morlet wavelet and the application of CWT for EIS analysis can be found in [11, 12].

III. EXPERIMENT SETUP

An experimental environment was established in order to perform diagnostic experiments on 5 cell EHP stack. All connections, pipings and instrumentations are shown in Fig. 2.
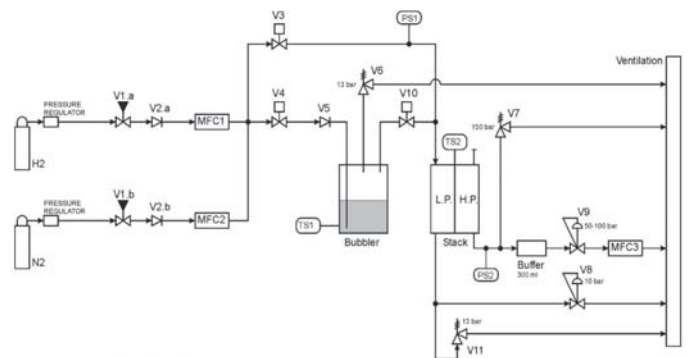
Fig 2. Gas connections of the experimental setup for performing diagnostic experiments on EHP
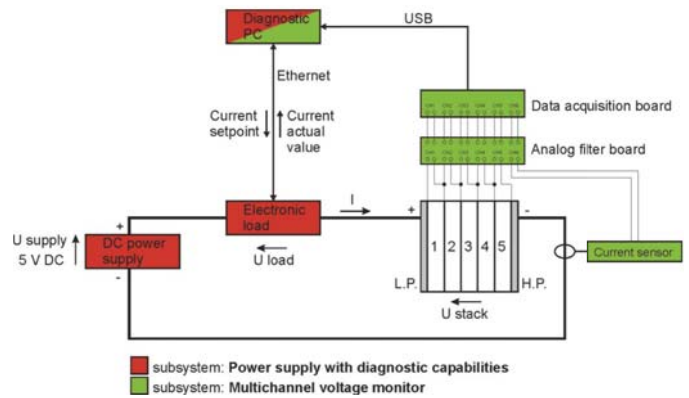
Fig. 3. Electrical connection of the EHP for performing electrochemical impedance spectroscopy

From the Fig. 2 mass flow controller MFC1 is used to supply hydrogen to the stack, while MFC2 is used to supply the "impurity", for example nitrogen. Solenoid valve V3 is used to deliver the hydrogen mixture to the stack directly, without humidification, while valves V4 and V10 deliver the hydrogen mixture to the stack through the humidifier (bubbler). Backpressure regulator V8 controls the pressure at the low pressure side of the stack. On the high pressure side there is a buffer to accumulate the pumped hydrogen and back pressure regulator V9 to control the pressure in the tank and at the high pressure side of the stack at constant value. Mass flow

controller MFC3 is used as a mass flow meter to measure the output flow rate. In steady state, MFC3 reading equals the hydrogen flow rate from the high pressure side of the stack.

Electrical connections on the EHP, which are essential to perform diagnostics, are shown in the Fig. 3. Characterization of the cells is performed at DC currents I_{DC}=20 A with peak-to-peak amplitude I_{AC}=1A by using classical single sine wave and stochastic DRBS excitations. The perturbed current is induced using programmable digital electric load (Rigol DL3031A) with excitation frequencies starting from 0.1 Hz up to 1000 Hz. The voltage response of the cells together with the current are sampled with sampling frequency f_s=50 kHz using 16-bit data acquisition card (NI USB 6215). Before DAQ signals are low pass filtered at f_c=10.8 kHz.

All controllable variables are kept constant during the process of characterization. Due to the fact that we were not able to directly measure and to give a precise quantification of the humidity level of the membrane, the first experiment was performed at nominal conditions with the humidifier (bubbler) and the second experiment was performed when the humidifier was turned off.

IV. RESULTS

In order to perform EIS, operating condition of the EHP must fulfil the linearity conditions, which are tightly connected to the amplitude of the excitation signal. Therefore, the excitation amplitude must be small enough in order to operate within the linear region but big enough to measure its response.

Before performing an EIS characterisation, the IV curve of the cell was first measured in order to determine the AC amplitude at desired DC current. From the IV curve (Fig. 4) it was decided to perform the EIS characterisation at I_{DC}=20 A with peak-to-peak amplitude I_{AC}=1 A using *sine wave* and *DRBS* excitation signals.
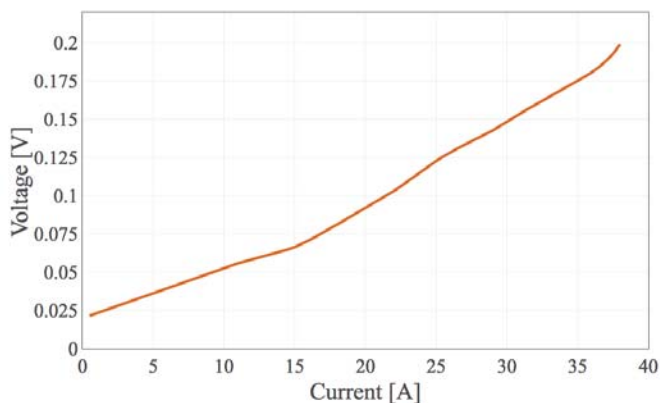
Fig. 4. IV curve of the 2nd cell from the HyET electrochemical hydrogen pump

The impedance of the cell under a test was first calculated using single sine wave excitation signals. It was calculated at 20 different frequency pints starting from 0.1 Hz up to 900 Hz (5 points per decade). In order to have good calculation of the impedance at every frequency point, at least 10 cycles are needed in order to obtain good quality of the acquired signal. For that reason, the excitation time at 0.1 Hz is 100 seconds.

Following this, the time needed to calculate the impedance at 20 different frequency points is almost 200 seconds.

By using DRBS as excitation signal (Fig. 5) the time needed for characterization is only 110 seconds, which represents 45 % decrease in time when compared to the classical sine wave excitation method. However, the main advantage of this method would be the ability to calculate the impedance in desired number of frequency points. In Fig. 6 is shown the comparison between the impedances calculated using sine wave and DRBS excitation signals. From the figure it can be concluded that the impedance calculated using DRBS is almost the same as the impedance calculated using classical single sine wave. The only difference would be in the low frequency region (at 0.1 Hz). The reason for this behaviour is due to the small changes in the operating point caused by stack temperature fluctuation.
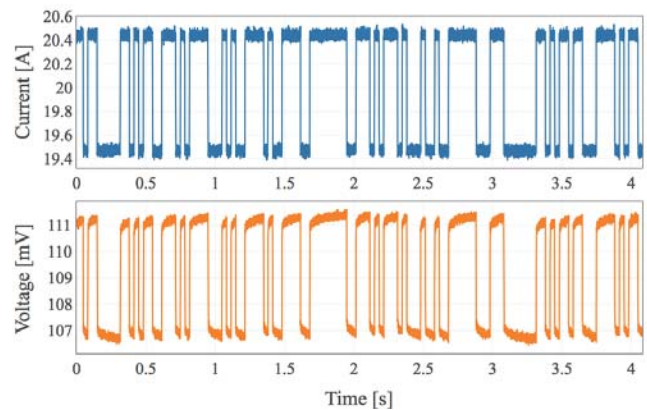
Fig. 5. DRBS current excitation signal and the voltage response of the cell
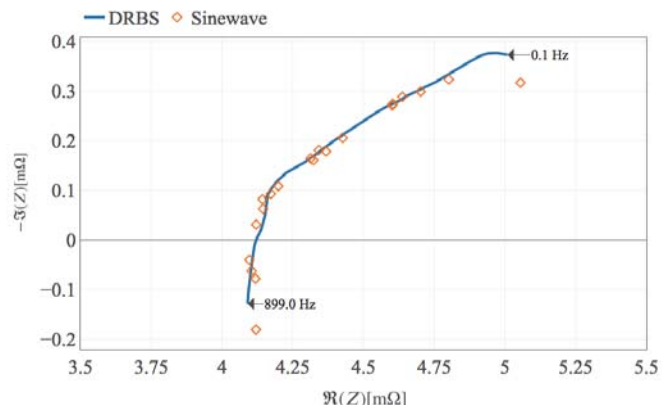
Fig. 6. Comparison between impedances calculated using classical single sine and DRBS excitation signal

In Fig. 7 is shown how the impedance changes when the cell operates in normal operating mode and in drying mode. From the shape of the impedances it can be concluded that drying operation mode causes a right shift and swelling. This behaviour is caused due to increase of the resistivity, which is connected to the size of the water clusters within the polymer microstructure. Dehydration of the EHP membranes leads to narrowing of the interconnecting channels which decreases the

mobility of the protons and increases the electrical resistance [13]. Since calculation of the impedance using DRBS signals takes less time, it is more suitable for performing online monitoring of the humidity of the membrane inside the EHP.
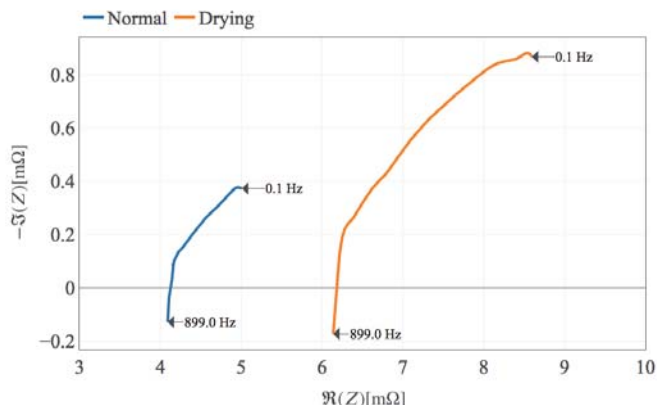
Fig. 7. Impedance of the cell at normal and drying operating conditions

V. CONCLUSION

Monitoring the level of humidity inside the membrane is crucial for maintaining its performances. Performing direct measurements of the humidity of the membrane is almost impossible due to the physical limitations. The only possible way is by using electrochemical impedance spectroscopy.

Impedance of the EHP was measured in the frequency interval between 0.1 Hz and 900 Hz using classical single sine and DRBS excitation signals at two different operating modes *normal* and *drying*. Classical way of performing EIS, which included calculation of the impedance in 20 different frequency points. Despite the accuracy of the calculation, classical way is usually time consuming and lack of accuracy especially in the low frequency region. By using stochastic excitation (DRBS) combined with wavelet transform, the impedance can be calculated in in desired number of frequency points. In addition to this, the time needed for characterization, when compared to the classical way, was decreased by almost 45 %. Therefore, performing EIS by using DRBS excitation signals is more adequate method for monitoring the level of humidity of the membrane.

From the obtained impedances can be seen that drying shifts the impedance to the right, which is directly connected to the increase of the membrane impedance. From this we can assume that by monitoring how the impedance is shifting it is possible to measure the humidity of the membrane while in operating mode.

ACKNOWLEDGEMENT

REFERENCES

[1] D. S. Satyapal, "Hydrogen and fuel cells overview," U.S. Department of Energy Fuel Cell Technologies Office.

[2] D. Fraile, J. Lanoix, P. Maio, A. Rangel, and A. Torres, "Overview of the market segmentation for hydrogen across potential customer groups, based on key application areas," CertifHy, 2015.

[3] K. Onda, K. Ichihara, M. Nagahama, Y. Minamoto, and T. Araki, "Separation and compression characteristics of hydrogen by use of proton exchange membrane," Journal of power sources, vol. 164, no. 1, pp. 1–8, 2007.

[4] P. Bouwman, "Electrochemical hydrogen compression (ehc) solutions for hydrogen infrastructure," Fuel Cells Bulletin, vol. 2014, no. 5, pp. 12–16, 2014.

[5] R. Ströbel, M. Oszcipok, M. Fasil, B. Rohland, L. Jörissen, and J. Garche, "The compression of hydrogen in an electrochemical cell based on a pem fuel cell design," Journal of power sources, vol. 105, no. 2, pp. 208–215, 2002.

[6] Y. Hao, H. Nakajima, H. Yoshizumi, A. Inada, K. Sasaki, and K. Ito, "Characterization of an electrochemical hydrogen pump with internal humidifier and dead-end anode channel," International Journal of Hydrogen Energy, vol. 41, no. 32, pp. 13 879–13 887, 2016.

[7] M.-T. Nguyen, S. Grigoriev, A. Kalinnikov, A. Filippov, P. Millet, and V. Fateev, "Characterisation of an electrochemical hydrogen pump using electrochemical impedance spectroscopy," Journal of applied electrochemistry, vol. 41, no. 9, p. 1033, 2011.

[8] A. Debenjak, P. Boškoski, B. Musizza, J. Petrovčič, and D. Juričić, "Fast measurement of proton exchange membrane fuel cell impedance based on pseudorandom binary sequence perturbation signals and continuous wavelet transform," Journal of Power Sources, vol. 254, pp. 112–118, 2014.

[9] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Elsevier Academic Press, 3 edn., ISBN 9780080922027, 2008.

[10] D. Iatsenko, Nonlinear Mode Decomposition, Springer Theses, Springer International Publishing, 2015.

[11] P. Boškoski, A. Debenjak, B. M. Boshkoska, Fast Electrochemical Impedance Spectroscopy, Springer, 2017.

[12] D. Iatsenko, P. V. McClintock, A. Stefanovska, Linear and synchrosqueezed time–frequency representations revisited: Overview, standards of use, resolution, reconstruction, concentration, and algorithms, Digital Signal Processing 42 (2015) 1 – 26, ISSN 1051-2004.

[13] P. D. Beattie, F. P. Orfino, V. I. Basura, K. Zychowska, J. Ding, C. Chuy, J. Schmeisser, and S. Holdcroft, "Ionic conductivity of proton exchange membranes," Journal of Electroanalytical Chemistry, vol. 503, no. 1- 2, pp. 45–56, 2001.

Algorithm Selection for Automated Audio Classification based on Content

Ivo Draganov[1] and Krasimir Minchev[2]

Abstract – **In this paper an approach for algorithm selection for automated audio classification is proposed based on content. Three popular algorithms for speech vs. music discrimination are incorporated, namely the zero-crossing rate, average frame power and 4-Hz energy modulation which prove effective at different level depending on the content of the audio being processed. Extensive testing is done over various musical pieces from different countries and in different styles to evaluate the accuracy and time consumption for each one. Then based on registered levels and with statistics from the initial record for particular audio recording the most proper could be switched on for processing in a complete audio classification system. Obtained results re considered promising for future use on a wider scale.**

Keywords – **Audio Classification, Music Speech Discrimination, 4 Hz Energy Modulation, Zero-Crossing Rate, Average Frame Power.**

I. INTRODUCTION

Discrimination between speech and music signals is applied in various areas of speech signal processing, such as Voice Activity Detection (VAD). On this occasion, many solutions, both in the time domain and in the frequency, have been proposed. The most common are: 4 Hz energy modulation, entropy modulation, spectral center, spectral flow, and zero-crossing rate (ZCR). Less frequent are spectrum overturning, spectral centroid, spectral flux variation and others [1, 2]. We will look at some of them by striving to discriminate given signals with accuracy we will seek maximum knowledge of the content but at the same time we will discuss the complexity of the calculations. We will focus on the discrimination of speech and sound signals based on their energy. The energy distribution of speech and musical signals will be evaluated by looking at the frequency of zero-crossings, short-term energy, and 4 Hz modulation also considering the Minimum Energy Density (MED), Low Energy Frames (LEF), and the Modified Low Energy Ratio (MLER) [3].

In order to discriminate speech and musical signals, features that are different for both classes are used. A simple look at the waveform of a one-minute excerpt of a voice signal, pop music, classical or opera signal shows great differences between classes. The shape of the signal of the speech signal shows a great difference in energy and amplitude that none of the musical signals show. The highly compressed waveform of a song seems to lack dynamics, while the classic instrumental and the opera performance have a low amplitude peak and show great dynamic deviations.

Even if the classes are easy to identify in wave form, the exact position of the transitions is difficult to detect. An excerpt from pop music shows possibilities of the song dynamics stopping abruptly when vocals appear. The transition is difficult to be spotted [4].

The rock music samples are not the same in comparison to the four examples described above but the most important and interesting thing here are the changes. In all the examples containing music the square root of RMS never goes down to zero and does not deviate significantly from the average value. There are many snippets in the speech signals containing null or near null values of the frames variation and the differences are sharp. Investigating the spectra of the four examples reveals that all musical examples have a higher peak at the low frequencies, although the peak occurs at different frequencies. This peak is mostly responsible for the fundamental frequency of vocal components. The classical music that vocals are missing is not as sharp as the other three examples. The speech signal has more energy in the frequency range around 1 to 3 kHz, unlike the musical examples.

Sandars notes, "It is well known that the energy contour is capable of separating speech from music". Discrimination of speech from music is most likely to be based on the differences of the continuous change in the envelope of the energy curve. In speech signals the vocals and the consonants are clearly distinguishable and on the other hand, the shape of the musical signal that is more stable is also easily detectable. Furthermore, we are aware that the speech signal has a 4 Hz energy modulation characteristic, which coincides with the frequency of the syllabi. Sandars uses a simple method of discriminating speech and musical signals. He found that using statistics calculated on the basis of a zero crossing factor, he could reach a classification of about 90 percent. By adding more information on the energy contour, he upgraded the accuracy to 98% [5].

In this paper we are investigating wide range of sound recordings differing in content which may utilize a proper algorithm for discrimination of music vs. speech with higher reliability and in the same time in some cases with reduced computational complexity based on three audio metrics – the ZCR, average frame power (FPOW) and 4 Hz energy modulation (4Hz). In Section II the selected metrics are described. Then, experimental results follow in Section III with discussion on the overall performance of the tested implementations. A conclusion is made in Section IV.

[1]Ivo Draganov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: idraganov@gmail.com.

[2]Krasimir Minchev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: krasimirminchew@gmail.com

II. Measures used

The first method we select for discriminating speech from music signals is the zero-crossing rate (ZCR). The frequency of zero crossings is a simple method of describing the content based on its most energetically pronounced frequency. We observe the definition of zero crossing frequency in the next expression [1]:

$$Zn = \sum_{m=-\infty}^{\infty} |sgn[x(m)]-sgn[x(m-1)]| \, w(n-m), \quad (1)$$

where:

$$sgn[x(n)]= 1, \quad x(n)\geq0, \quad (2)$$
$$sgn[x(n)]=-1, \quad x(n)<0. \quad (3)$$

And $w(n)$ is the window comprising of N number of frames:

$$W= 1/2N, \ 0\leq n\leq N-1, \quad (4)$$
$$W = 0, \text{ otherwise.} \quad (5)$$

The short-term energy method is a little more complicated than the above-mentioned, but it's simpler to apply than finding the 4Hz measure. Given that the amplitude of the non-intuitive segments is noticeably lower than those of the speech segments, the short-term energy of the speech signals reflects the amplitude dispersion. By observing a speech signal, we can notice that the peak of the signal amplitude is noticeable as well as the fundamental frequency in the speech parts of the signal. This suggests that simple time processing techniques could derive useful information about signal characteristics. Most short-term processing techniques that derive features from the time domain Q[n] can be represented mathematically as [6]:

$$Qn = \sum_{m=-\infty}^{\infty} T[x(m)]w(n-m). \quad (6)$$

T is the transposed matrix, which may be linear or non-linear, $x(m)$ represents the information sequence, and $w(n-m)$ represents a time-limited window sequence. The energy of a discreet signal is defined by the expression:

$$E_t = \sum_{m=-\infty}^{\infty} s^2(m), \quad (7)$$

where E_t is the total energy and $s(m)$ is the discreet signal. For the calculation of Short-Term Energy, the signal is considered in short frames, whose size is usually between 10 and 30ms. It is necessary to take all samples in a frame from the signal from $m = 0$ to $m = N-1$, where N is the length of the frame. Then:

$$E_t = \sum_{m=-\infty}^{-1} s^2(m)+ \sum_{m=-\infty}^{N-1} s^2(m)+ \sum_{m=-\infty}^{\infty} s^2(m). \quad (8)$$

The samples value's are zero outside the frame. Therefore:

$$E_t = \sum_{m=0}^{N-1} s^2(m). \quad (9)$$

From (9) it could be estimated that the total energy of a frame for the signal from 0 to N-1 samples. The short-term energy is defined as the sum of the squares of the samples in a frame according to:

$$e(n) = \sum_{m=-\infty}^{\infty} [s_n(m)]^2. \quad (10)$$

After splitting in frames and windows, the N-th frame of the signal becomes $s(m).w(n-m)$ and therefore Eq. (10) becomes:

$$e(n) = \sum_{m=-\infty}^{\infty} [s(m).w(n-m)]^2, \quad (11)$$

where $w(n)$ is represented with a window function of limited duration, and n is the offset of the frame in number of samples. This shift may be as small as one sample or as large as one full frame.

The most complex for realization of the three methods is 4Hz modulation [7]. This method implies better discrimination and a higher rate of success than the previous two methods, but its implementation goes through more stages. First it is needed to derive MEL coefficients. Pre-emphasis is introduces, then the signal spectrum is re-emphasized and the constant component is removed. A low-order digital filter (most commonly a first order FIR filter) is attached to the input $x(n)$ so it is aligned in its spectrum:

$$H(z)=1-az^{-1}, \quad 0.9<a<1. \quad (12)$$

Then follows splitting in frames and the spectral analysis is performed on them. This is because human speech does not change much over time and can be treated as a quasi-static process. Very popular frame length is 20-30ms. Hamming window weighting of each frame takes place according to:

$$y(n)=x(n)w(n), \quad (13)$$

$$W(n)=0.54-0.46cos(\frac{2\pi}{N-1}). \quad (14)$$

Every frame undergoes Discrete Fourier transform:

$$X(k)= \sum_{n=0}^{N-1} y(n)e^{-\frac{j2\pi}{N}} \quad 0\leq n,k\geq N-1, \quad (15)$$

A set of triangular band-pass filters that simulate the characteristics of the human ear are applied to the signal spectrum. This process is called Mel filtering. Human hearing perception analyzes audible spectrum of groups based on the number of overlapping critical bands. These bands are allocated in such a way that the frequency resolution is high in the low frequency area and low in the high frequency area. Mel frequency is found from linear frequency based on:

$$f_m=2525 \, x \, log(1+\frac{1}{7}). \quad (16)$$

The energy of the filter bank is found by:

$$E=\sum_{k-1}^{N} |X(k)|^2. \, \psi_i \, (k) \quad (17)$$

and after DCT over all MEL coefficients, finding their second derivatives and filtering the result in 40 channels with another FIR filter, all the energies contained in them are summed

together and the total energy is normalized with the average energy of the frame. The modulation is characterized by the variation of the filtered energy in dB per a second from the whole signal.

III. EXPERIMENTAL RESULTS

Experimental testing relies on a custom built database comprising of 5 speech and 10 music recordings in non-compressed format. Sampling frequency is 44100 Hz with resolution of the samples of 16 bits and all being stored in single channel wav files. The duration for all sounds captured is 1 minute.

Table I contains the ZCR, FPOW and 4Hz values obtained from the speech signals where *speechf* labels identify recording with a female voice and *speechm* – a male one.

TABLE I
SPEECH SIGNALS MEASURES

Test set	ZCR	FPOW	4Hz
speechf1	0.0947	0.0018	45.5400
speechm2	0.0671	0.0034	45.3575
speechf3	0.0743	0.0036	36.4281
speechm4	0.0853	0.0035	41.4678
speechm5	0.0793	0.003	43.5407
Average	0.0801	0.0031	42.4668

In Table II the resulting values for the three parameters are given when found over the 10 music recordings some of which are typical folklore works from different geographical locations.

TABLE II
MUSIC SIGNALS MEASURES

Test set	ZCR	FPOW	4Hz
Rock	0.1971	0.0024	0.6245
Jazz	0.0971	0.0016	13.8121
Latino	0.2086	0.0035	0.5937
Folk	0.0899	0.0067	5.4985
Hindi	0.0916	0.0056	5.6968
Nordic	0.1083	0.0023	2.5773
African	0.0888	0.000837	64.8428
Chinese	0.0421	0.0036	25.4098
Russian	0.0612	0.0038	14.1327
Classic	0.0688	0.0029	11.7992
Average	0.1053	0.0033	14.4987

In order to discriminate speech and music signals based on these methods using full validation, we must use the average values for all files of a given type - musical or speech signals, and by them to calculate a threshold to use. To compute the threshold of a method, we collect the two average values of the musical and speech signals and divide them into two, and thus we get the threshold that will discriminate against the signals. For the zero crossing frequency, the values below the threshold, i.e. the signals with a lower frequency of zero crossings, will be defined as speech signals and those with

higher values - as musical. Related values found for the thresholds are t_{ZCR} = 0.0928, t_{FPOW} = 0.0032, and t_{4Hz} = 28.4830.

The results from full validation of the speech database are shown in Table III.

TABLE III
CLASSIFICATION ACCURACY OF SPEECH SIGNALS

Measure	ZCR		FPOW		4Hz	
Threshold	0.0928		0.0032		28.4830	
Classify	Right	Wrong	Right	Wrong	Right	Wrong
speechf1	0	1	1	0	1	0
speechm2	1	0	0	1	1	0
speechf3	1	0	0	1	1	0
speechm4	1	0	0	1	1	0
speechm5	1	0	1	0	1	0
Accuracy,%	80	20	40	60	100	0

Let's take a look at the data in the second and third columns referring to the ZCR method. We determine the accuracy of the validation against the correct classification of the speech signal due to the threshold of the different methods that are used according to the method described above. After the classification of all speech files for the zero crossing frequency method, we obtain accuracy of 80% at full validation. Taking into account the simplicity of this method, its accuracy is very satisfactory. We continue with the examination of the fourth and fifth columns where the short-term energy method is described. When it is classified after all the speech files, we obtain the accuracy of the 40% full validation, which is an unsatisfactory result. The following method is observed in columns six and seven, where the dispersion method of 4 Hz modulation energy is described. For it, we get 100% validation accuracy, which is the highest possible result we are looking for. By comparing the three methods, we can categorically define the 4Hz modulation energy method as the best method for determining speech signals, given its complexity compared to the other two methods, the result is expected. But the simplest method - the frequency of zero crossings is more effective for determining speech signals than the more sophisticated method - the short-term energy method.

The first wrong classification we notice for the ZCR method in the first female voice record. Given the simplicity of the method, it is not the most reliable classifier since the values of the source data do not differ dramatically from the wrong classification and may be due to the low energy in this record and the presence of more consonant letters or silence mixed with some noise. Let's look at the short-term energy method, and we see that all but one values are very close. We see that wrong classification is in both female and male speech records. Their values are so close that we can not identify features that are clear and suggest an increase or decrease in energy to a subsequent misclassification. These close values may be due to the peripheral devices used, the non-isolated environment, etc. So the accuracy of the short-term energy method is unsatisfactory for end-user applications.

The most effective and accurate method that end-user can use is the dispersion of 4 Hz modulation energy. Depending

on the priority, if it is the accuracy and not time consumption, the best and most appropriate is this method.

The classification accuracy for the music signals is presented in Table IV.

TABLE IV
PAGE LAYOUT DESCRIPTION

Measures	ZCR		FPOW		4Hz	
Threshold	0.0928		0.0032		28.4830	
Classify	Right	Wrong	Right	Wrong	Right	Wrong
Rock	1	0	0	1	1	0
Jazz	1	0	0	1	1	0
Latino	1	0	1	0	1	0
Folk	0	1	1	0	1	0
Hindi	0	1	1	0	1	0
Nordic	1	0	0	1	1	0
African	0	1	0	1	0	1
Chinese	0	1	1	0	1	0
Russian	0	1	1	0	1	0
Classic	0	1	0	1	1	0
Accuracy,%	40	60	50	50	90	10

After the necessary calculations for all the music signals for the zero crossing method, we get the accuracy of the full validation of 40%. The accuracy of this method of recognizing musical signals is not satisfactory. Some of the errors that are introduced in classifying using this method may be due to the type of music used. We observe a proper classification for rock, jazz, etc., while in classical, Russian, Chinese music, etc., we observe a wrong classification, which may be due to the instruments and pauses contained in a certain type of music. For example, classical and Chinese music is experiencing energy accumulation at low and high frequencies, as are the moments of silence caused by instruments used in this type of music such as string, wind, keyboard instruments - flute, violin, piano.

The next method that we are looking at is the short-term energy method for it after the classification of all the music files we get the accuracy of the full validation 50%. The accuracy of this method for recognizing a musical signal is better than the accuracy of the zero crossing frequency method and its accuracy of speech recognition but is still unsatisfactory and the use of this method is not very reliable.

After the classification of all music files with 4Hz method, modulation energy has a 90% accuracy of full validation. The only error that has been made is with African music, yet this method is the closest to the maximum accuracy we are looking for. By comparing the three methods, we can conclude that the method of 4 Hz modulation energy dispersion is the most reliable method for recognizing musical signals. But this time for the classification of musical signals, the short-term energy method is more applicable than the zero crossing frequency method. Observing Table IV we can see that the wrong classification, which is present in the 4Hz method, is not correctly classified in the other two methods.

In Fig. 1 the overall performance is given for the three algorithms of classifying the test audio content into speech and music with the complete variability in different styles.
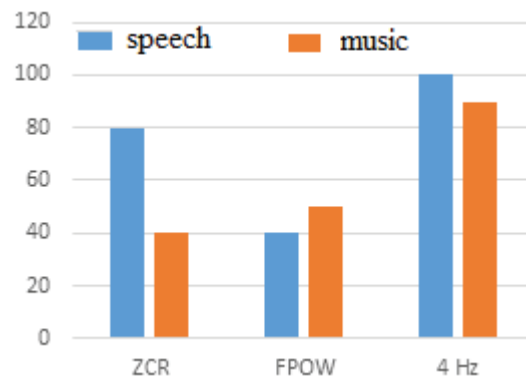
Fig. 1. Overall accuracy of classification for the three tested algorithms

IV. CONCLUSION

The best and most accurate method that most satisfies the discriminatory condition and can be used by an end user in an objective environment with publicly available and relatively cheap peripherals for speech and music discrimination is the 4Hz method. The following method, which certifies the discrimination condition to some extent and can be used is the method of the zero crossing rate, although it is not very reliable, but the calculation is much simpler, which makes it an ideal choice if the main goal is speeding-up the process rather than achieving accuracy close to 100%. For future improvement of the zero crossing method and the short-term energy method, professional peripheral devices and soundproofing environment can be used to increase their classification accuracy. End user use can use those along with other methods or features to improve their performance.

REFERENCES

[1] Carey, M., E. Parris, S. Eluned, S., H. Lloyd-Thomas. A comparison of features for speech, music discrimination. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 149-152, 1999.

[2] Lu, L., H. Jiang, H. Zhang, A robust audio classification and segmentation method. In Proceedings of the 9th ACM International Conference on Multimedia, pp. 203-211, October 2001.

[3] Velayatipour, M., B. Mosleh, A review on speech-music discrimination methods, International Journal of Computer Science & Network Solutions, February 2.2, 2014.

[4] Ericsson, L., Automatic speech/music discrimination in audio files. Skolan för datavetenskap och kommunikation, Kungliga Tekniska högskolan, 2010.

[5] Spina.S. Analysis and transcription of general audio data, PhD Thesis. Massachusetts Institute of Technology, 2000.

[6] Shete, D., S. Patil, S. B. Patil, Zero crossing rate and Energy of the Speech Signal of Devanagari Script. IOSR-JVSP, 4.1: 1-5, 2014.

[7] Logan, B. Mel Frequency Cepstral Coefficients for Music Modeling. In: ISMIR, p. 1-11, 2000.

Study of Parasitic Effects in Two-Integrator Loop G_m-C Filters If Realized with Single Stage OTAs

Boncho Nikov[1], Ivan Uzunov[2] and Marin Hristov[3]

Abstract – **A generalized study concerning the effects of OTA imperfections in the second order Gm-C filters based on two-integrator loop configuration is done in the paper. Single stage CMOS OTAs are assumed in the investigations and the most important OTA imperfections in this case are their input resistance and input capacitances. Most of the considered circuits have similar properties concerning these imperfections and it is shown that the gyrator biquad has the best behaviour.**

Keywords – **Gm-C filters, operational transconductance amplifiers (OTA), imperfections, biquads.**

I. INTRODUCTION

The active filters based on operational transconductance amplifiers (OTA) and capacitors, known as G_m-C filters, are still widely used in analog signal processing [1,2]. The frequency domain of their implementation is very wide – it ranges from Hertz area [3] up to several hundreds of MHz [4,5] and even to GHz [6]. They attract with several benefits: easy for integration; versatile configurations satisfying different requirements; wide frequency tuning range, covering more than one decade when necessary [3,4,6]. These filters are known for long time, however they are still object of investigations, related to their extending and modifying applications and to the use of the new technologies for their realization. The permanent trends for reducing of the supply voltages, reduction of the sizes, requirements for low consumed dc power, etc., change the OTA parameters and as consequence the effects of these parameters in filter circuit.

Most often as second order G_m-C sections (biquads) are used the circuits belonging to the wide class of two-integrator loop filters [1,2]. This is due basically to their versatility – for the most of them one circuit is able to realize every transfer function, which is achieved by applying the input signal or by taking the output signal from different points of the circuit. Other advantages are their abilities for realizing of high Q-factors and for operation at high frequencies.

Usually single stage CMOS OTAs are used for design of G_m-C biquads. Multistage OTAs are appropriate for realizing of high G_ms, which is achieved at the price of reduced frequency bandwidth due to appearance of high-impedance node between the stages [7]. OTA with high G_ms are necessary for high frequency filters, which is in contradiction with their limited bandwidth. Thus, it is desirable in such cases to try to design OTAs with high G_mS.

The influence of the single-stage OTA parameters on the behavior of G_m-C sections differs in some extent compared with multistage OTA. The major parasitic parameters of a multistage OTA are its input and output impedances and the frequency dependence of G_m, represented usually by single pole approximation [1]. The basic parasitic parameters of single-stage CMOS OTA are their input capacitances and their output impedances. The frequency dependence of G_m in this case (its increasing with frequency) can be represented by one zero [1]. However it is more appropriate to consider G_m as frequency independent and add a parasitic transition capacitance to be accounted for the frequency dependence – in fact this is C_{gd} of the transistors in input differential pair of the OTA. Since this capacitance is much less than the input capacitance it has less effect.

This paper tries to compare the effects of the OTA parasitic parameters in the most popular Gm-C biquads based on two integrator loop configuration. The goal is to estimate how these influences change the possibilities of the circuits for realization of high Q-factor filters and their ability for operation in wide frequency range. The second chapter, describes shortly the considered circuit if ideal OTAs are used. Third chapter considers the effect of OTA output resistances on the behavior of the circuits, and in the fourth chapter is discussed the changes due to the input capacitances of the amplifier – so called excess phase effect.

II. SHORT DESCRIPTION OF THE CONSIDERED CIRCUITS

The filter circuits, which will be investigated here, are given in Fig. 1. They represent most of the two-integrator loop biquads [1,2] and their single-ended versions are shown for simplicity. All circuits are able to realize different biquads (low-pass, high-pass, band-pass, with complex zeros) depending on the way of applying the input signal and taking the output signal. The transfer functions of the circuits can be written in the general way:

$$H(s) = \frac{N(s)}{s^2 + d_1 s + d_0} = \frac{N(s)}{s^2 + \frac{\omega_{p0}}{Q_{p0}} s + \omega_{p0}^2}, \tag{1}$$

where $N(s)$ is either first or second order polynomial or a constant, depending on the section type. The parasitic effects, considered here, concern basically the denominator of the transfer function and are significant at high pole Q-factor. For this reason their influence on the numerator $N(s)$ will be not considered and inputs and outputs are shown in Fig. 1 for the

[1]Boncho Nikov is with Melexis Bulgaria Ltd; 2 Samokovsko Shosse blvd.; 1138 Sofia, Bulgaria; E-mail: bni@melexis.com.

[2]Ivan Uzunov is with Smartcom Bulgaria AD, 133 Tzarigradsko shosse Blvd., 1784 Sofia, Bulgaria, e-mail: ivan_uzunov@smartcom.bg.

[3]Marin Hristov is with the Faculty of Electronics and Electronic Technlgies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

case of band-pass sections – circuits, which most often require high Q-factors.

The coefficients of the transfer function denominators and the pole parameters are summarized in Table I. In all formulas a is ratio of identically marked g_ms, while b differs in Fig. 1(a) and Fig. 1(d):

$$a = \frac{g_{m3}}{g_{m4}}; \quad b = \frac{g_{m5}}{g_{m6}} \text{ in (a)}; \quad b = \frac{g_{m5}}{g_{m4}} \text{ in (d)}. \quad (2)$$
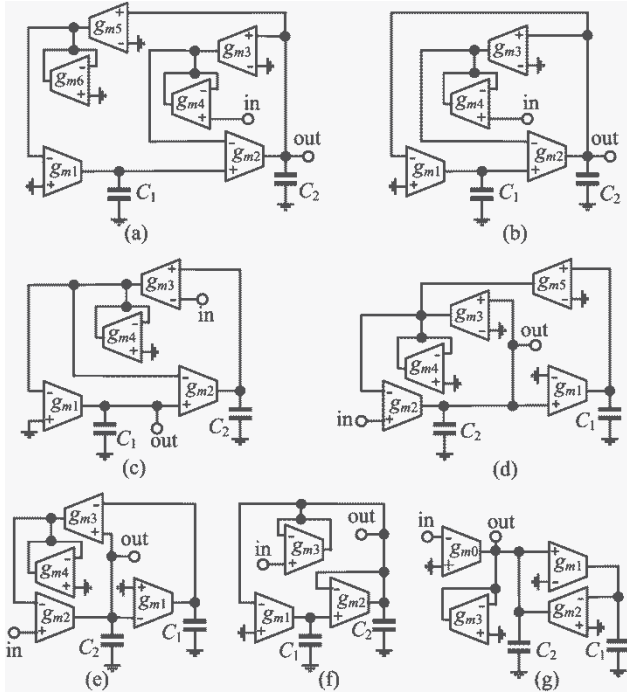
Fig. 1. Two-integrator loop biquads [1]: (a), (b), (c) distributed-feedback configurations; (d), (e) summed-feedback configurations; (f) Tow-Thomas G_m-C circuit; (g) gyrator band-pass biquad.

TABLE I.
TRANSFER FUNCTION PARAMETERS OF THE BIQUADS IN FIG. 1.

	d_1	$d_0 = \omega_{p0}^2$	Q_{p0}
Fig. 1(a) Fig. 1(d)	$a\dfrac{g_{m2}}{C_2}$	$b\dfrac{g_{m1}g_{m2}}{C_1C_2}$	$\dfrac{1}{a}\sqrt{b\dfrac{g_{m1}}{g_{m2}}\dfrac{C_2}{C_1}}$
Fig. 1(b)	$a\dfrac{g_{m2}}{C_2}$	$\dfrac{g_{m1}g_{m2}}{C_1C_2}$	$\dfrac{1}{a}\sqrt{\dfrac{g_{m1}}{g_{m2}}\dfrac{C_2}{C_1}}$
Fig. 1(c) Fig. 1(e)	$a\dfrac{g_{m2}}{C_2}$	$a\dfrac{g_{m1}g_{m2}}{C_1C_2}$	$\sqrt{\dfrac{1}{a}\dfrac{g_{m1}}{g_{m2}}\dfrac{C_2}{C_1}}$
Fig. 1(f)	$\dfrac{g_{m2}+g_{m3}}{C_2}$	$\dfrac{g_{m1}g_{m2}}{C_1C_2}$	$\sqrt{\dfrac{C_2}{C_1}}\dfrac{\sqrt{g_{m1}g_{m2}}}{g_{m2}+g_{m3}}$
Fig. 1(g)	$\dfrac{g_{m3}}{C_2}$	$\dfrac{g_{m1}g_{m2}}{C_1C_2}$	$\dfrac{\sqrt{g_{m1}g_{m2}}}{g_{m3}}\sqrt{\dfrac{C_2}{C_1}}$

III. INFLUENCE OF OTA OUTPUT CONDUCTANCES

OTA output impedances consist typically of parallel connected conductance and capacitance. The output capacitances in the considered circuits are connected in parallel either to the integrator capacitors C_1 and C_2 or to the input capacitance of the same or another OTA. In the first case they can be considered as parts of C_1 or C_2; in the second case their influence should be considered together with the influence of the OTA input capacitances. For this reason the effect of output impedances will be considered as effect of output conductances only. The inspection of the circuits in Fig. 1 shows that only output conductances g_{o1} and g_{o2} of OTAs g_{m1} and g_{m2} affects the circuit parameters, since they are in parallel to C_1 and C_2. The other output conductances are in parallel to OTAs, connected as resistors, and can be absorbed in these resistors.

After these observations the analysis can be done easily – it is enough to replace sC_1 by $sC_1 + g_{o1}$ and sC_2 by $sC_2 + g_{o2}$ in the expressions for the transfer functions. Proceeding in this way, the following expression for the transfer function of the circuits in Fig. 1(a) and 1(d) is derived:

$$H(s) = \frac{N(s)(1+\omega_1/s)(1+\omega_2/s)}{s^2 + s\left(\frac{\omega_{p0}}{Q_{p0}}+\omega_1+\omega_2\right)+\omega_{p0}^2+\frac{\omega_{p0}}{Q_{p0}}\omega_1+\omega_1\omega_2}, \quad (3)$$

where $\omega_1 = g_{o1}/C_1$ and $\omega_2 = g_{o2}/C_2$. The changes due to output conductances appear in three places. It increases pole frequency however this increase is small. Both terms which are added to ω_{p0}^2 are much smaller since the ratios ω_1/ω_{p0} and ω_2/ω_{p0} are

$$\frac{\omega_1}{\omega_{p0}} = \frac{g_{o1}}{\sqrt{bg_{m1}g_{m2}}}\sqrt{\frac{C_2}{C_1}}, \quad \frac{\omega_2}{\omega_{p0}} = \frac{g_{o2}}{\sqrt{bg_{m1}g_{m2}}}\sqrt{\frac{C_1}{C_2}}. \quad (4)$$

They are significantly less than 1 if $g_{o1,2} \ll g_{m1,2}$ (well-designed OTA) and also high-Q circuits are considered.

It seems that the multipliers $1 + \omega_1/s$ and $1 + \omega_2/s$ increase the gain to infinity at very low frequencies. However they present in the formula only because the change of the numerator is not considered. The numerator $N(s)$ is also function of C_1 and C_2. The capacitors in the numerator will produce the same terms in a way, which will cause canceling with the terms coming from the denominator. Therefore in the final expression for $H(s)$ these two multipliers will not exist.

The most important effect of the output conductances is in the first order term in the denominator. The quantities ω_1 and ω_2, which are added to ω_{p0}/Q_{p0}, can be of the same range as ω_{p0}/Q_{p0} and limit the maximum achievable Q-factor. If assume $(\omega_{p0}/Q_{p0}) = 0$, i.e. $Q_{p0} \to \infty$, then the pole Q-factor is determined by the sum $\omega_1 + \omega_2$ and is equal to

$$Q_{p\,max} \approx \frac{1}{\frac{g_{o1}}{g_{m1}}\frac{a}{b}Q_{p0}+\frac{g_{o2}}{g_{m2}}\frac{1}{a}\frac{1}{Q_{p0}}}. \quad (5)$$

Formulas (3), (4) and (5) are derived for Fig. 1(a) and 1(d), however they are valid for all circuits except the gyrator one (Fig. 1(g)), if set $b = 1$ for Fig. 1(b); $b = a$ for Fig. 1(c) and 1(e); and $b = a = 1$ for Fig. 1(f). Both quantities Q_{p0} and $Q_{p\,max}$ determine the real pole quality factor Q_p according the formula

$$\frac{1}{Q_p} = \frac{1}{Q_{p0}} + \frac{1}{Q_{p\,max}}. \quad (6)$$

These results can be interpreted in the following way: If Q_{p0} is specified then exists a parameter $Q_{p\,max}$, dependent on Q_{p0} and on the ratios $g_{o1,2}/g_{m1,2}$, which limits Q_p. Since Q_{p0} is ratio of g_ms and of capacitors and can't be done infinitely large, the real Q_p of the circuit is always combination of Q_{p0} and $Q_{p\,max}$. The

value of $Q_{p\,max}$ is necessary to be high if high Q_p is needed – it should be of the range of Q_{p0} or higher. Since Q_{p0} enters in the denominator of $Q_{p\,max}$, this requirement leads to hard demand on the ratio g_{o1}/g_{m1} – it should be less than $1/Q_{p0}^2$.

The expressions for ω_p and Q_p for the gyrator circuit are

$$\omega_p^2 = \left(1 + \frac{g_{o1}g_{23}}{g_{m1}g_{m2}}\right)\omega_{p0}^2 \approx \omega_{p0}^2; \quad Q_p \approx \sqrt{\frac{g_{m1}g_{m2}}{g_{o1}g_{23}}}\frac{1}{\sqrt{\frac{\tau_{o1}}{\tau_{o2}}}+\sqrt{\frac{\tau_{o2}}{\tau_{o1}}}}, \quad (7)$$

where $g_{23} = g_{o2} + g_{m3}$, $\tau_1 = g_{o1}/C_1$, $\tau_2 = g_{o2}/C_2$, and ω_{p0} is given in the last row of Table I. The maximum pole Q, defined by OTA output conductances only, is achieved when $g_{m3} = 0$, i.e. when the corresponding OTA is missing (in fact this OTA is used only for fixing the desired Q_p); and when $\tau_1 = \tau_2$. Thus the gyrator circuit has more potential for realization of high-Q biquads.

IV. Limitations from OTA Input Capacitances

Other parasitics, which may affect significantly the behavior of the circuits, are OTA's input capacitances. Those of them, which are connected in parallel to integrator capacitors C_1 and C_2 are not so dangerous. Usually they are smaller than C_1 and C_2 and can be considered as parts of them. Their effect is some deviation of the pole frequency, which can be compensated by proper adjustment. More critical is the influence of the input capacitances, which appear in parallel to OTAs, connected as resistors: g_{m4} and g_{m6} in Fig. 1(a) and g_{m4} in Fig. 1(b), (c), (d) and (e). They introduce parasitic capacitances C_{p4} in parallel to g_{m4} and C_{p6} in parallel to g_{m6}, which are equal correspondingly:
- in Fig. 1(a): $C_{p4} = C_{in,2} + C_{in,4}$, $C_{p6} = C_{in,1} + C_{in,6}$;
- in Fig 1(b), (d) and (e): $C_{p4} = C_{in,2} + C_{in,4}$;
- in Fig. 1(c): $C_{p4} = C_{in,1} + C_{in,2} + C_{in,4}$

where $C_{in,k}$ is the input capacitance of OTA g_{mk}.

The capacitances C_{p4} and C_{p6} together with g_{m4} and g_{m6} form parallel RC circuits. They introduce additional phase shift (so called excess phase) in the gains of the voltage multipliers, created with the help of g_{m4} and g_{m6} (for example of the multipliers g_{m3}/g_{m4} and g_{m5}/g_{m6} in Fig. 1(a)). In fact the coefficients a and b in the formulas for d_1 and d_0 in Table I are not frequency independent and the following formulas are valid

$$a = \frac{a_0}{1+s/\omega_4}; \quad b = \frac{b_0}{1+s/\omega_6}\text{ in (a)}; \quad b = \frac{b_0}{1+s/\omega_4}\text{ in (d)}, \quad (8)$$

where a_0 and b_0 are the values of these parameters according (2) and frequencies ω_4 and ω_6 are given by

$$\omega_4 = g_{m4}/C_{p4}; \quad \omega_6 = g_{m6}/C_{p6}. \quad (9)$$

The replacement of the expressions for a and b changes the formulas for filter transfer functions in the following way:

for Fig. 1(a):

$$H(s) = \frac{N(s)(1+s/\omega_4)(1+s/\omega_6)}{s^2\left(1+\frac{s}{\omega_4}\right)\left(1+\frac{s}{\omega_6}\right)+s\frac{\omega_{p0}}{Q_{p0}}\left(1+\frac{s}{\omega_6}\right)+\omega_{p0}^2\left(1+\frac{s}{\omega_4}\right)}; \quad (10a)$$

for Fig. 1(b): $\quad H(s) = \dfrac{N(s)(1+s/\omega_4)}{s^2\left(1+\frac{s}{\omega_4}\right)+s\frac{\omega_{p0}}{Q_{p0}}+\omega_{p0}^2\left(1+\frac{s}{\omega_4}\right)}; \quad (10b)$

for Fig. 1(c), (d) and (e): $\quad H(s) = \dfrac{N(s)(1+s/\omega_4)}{s^2\left(1+\frac{s}{\omega_4}\right)+s\frac{\omega_{p0}}{Q_{p0}}+\omega_{p0}^2}. \quad (10c)$

The terms $(1 + s/\omega_4)$ and $(1 + s/\omega_6)$ appear in the denominator and also multiply the numerator. Their effect in the denominator of (10c) will be considered firstly. Now it is of 3rd degree, i.e. the complex poles of the filter are changed and appears a third pole, which is real. The influence of the term $(1 + s/\omega_4)$ depends basically on the ratio ω_4/ω_{p0}. The numerical investigation shows existence of minimal allowed value of this ratio, at which the pole Q-factor is equal to infinity and below this value $Q_p < 0$, i.e. the circuit is unstable. This is illustrated in Fig. 2(a). The minimal value of ω_4/ω_{p0} is equal to Q_{p0}, which can be proved mathematically. At $(\omega_4/\omega_{p0}) = Q_{p0}$ the denominator of (10c) can be written as

$$D(s) = \left(s^2 + \omega_{p0}^2\right)\left(1 + s/(\omega_{p0}Q_{p0})\right), \quad (11)$$

i.e. the complex poles are purely imaginary and the circuit is at the boundary of stability. Of course the value of ω_4 in the real circuits should be higher of the limit ($= \omega_{p0}Q_{p0}$) in order to have reserve of stability. For example, if assume 20% allowed increasing of Q_p then ω_4 should be not less than $60\omega_{p0}$, $178\omega_{p0}$ and $589\omega_{p0}$ for $Q_{p0} = 10$, 30 and 100 correspondingly – significantly stronger requirement than the limit $\omega_{p0}Q_{p0}$.
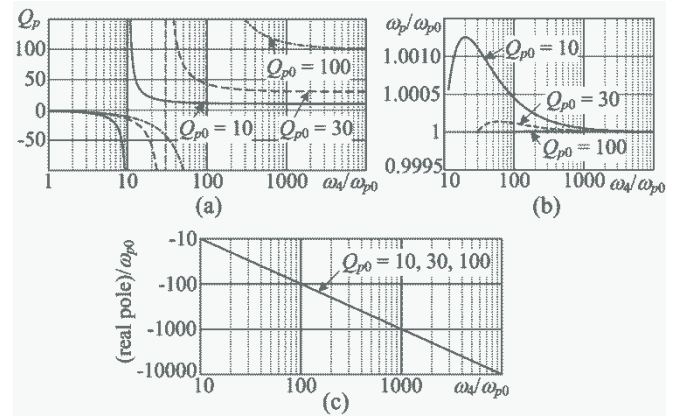
Fig. 2. Influence of the ratio ω_4/ω_{p0} for the circuits in Fig. 1(c), (d) and (e): (a) change of the pole Q-factor; (b) change of the pole frequency; (c) extra real pole.

The other influences of ω_4 are negligible. Fig. 2(b) shows that the relative variation of the frequency of the complex poles is less than 0.1% for values of ω_4/ω_{p0}, for which the circuit is stable. The extra real pole is equal to ω_4 as can be concluded from Fig. 2(c) and it will cancel with the zero introduced by the multiplier $(1 + s/\omega_4)$ in the numerator.

The considerations above are valid also for the circuit in Fig. 2(a) when $\omega_4 = \omega_6$. Then identical multipliers $(1 + s/\omega_4)$ appear in the numerator and denominator of (10a), they cancel themselves and the denominator of (10a) becomes the same as those of (10c). When ω_4 and ω_6 differ, the studying of the transfer function can be done numerically, calculating the poles at different values of ω_4 and ω_6. Elaborating in this way it is found out for moderate ratios ω_4/ω_6 between 0.5 and 2 that the instability is defined from ω_6 only and the limit is $\omega_6/\omega_{p0} \approx Q_{p0}$. Which is interesting, this limit doesn't depend on ω_4. The other influences are insufficient as in the previous case: the frequency of the complex pole pair is stable and two real pole appears, which are far from the complex poles.

The variations of Q and of the pole frequency of the circuit in Fig. 1(b) from the parasitic frequency are illustrated in Fig. 3. The Q-factor also increases at low ω_4, however it is not so much and the circuit stays stable even for values of ω_4 unrealistically close to ω_{p0}. The change of the pole frequency also is small. The extra real pole, which also appears in this circuit, is far from the pole frequency for $\omega_4 > 10\omega_{p0}$ (normal values of ω_4) and its effect can be neglected.
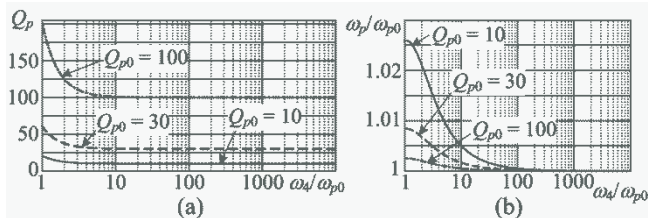
Fig. 3. Influence of parasitic frequency $\omega_4 = g_{m4}/C_{p4}$ in Fig. 1(b): (a) variation of pole Q-factor; (b) variation of pole frequency.

Generally, the excess phase due to parasitic capacitances in parallel to OTAs, connected as resistors in Figs. 1 (a) to (e), causes the effect, known in the two-integrator loop biquads with other types of amplifiers [1]: increasing of the pole Q-factor, which may bring even to instability of the circuit. In the considered case the reason is the input capacitances of some OTAs. This effect limits the high frequency application of the circuits. High pole frequencies are achieved by large G_ms of OTAs g_{m1} and g_{m2} and small capacitances C_1 and C_2. Large G_ms require transistors with larger sizes in single stage OTAs and as consequence – higher input capacitances. The consideration here shows existence of lower limit for the ratio ω_4/ω_{p0}. Thus, raise of ω_{p0} leads to higher g_{m1} and g_{m2}, higher input capacitances, lower ω_4 and as result – approaching the limit.

A way to compensate the effect of excess phase is to add a proper resistor in series to one of the capacitors C_1 and C_2 – the one which is in the loop immediately before g_{m4}. This resistor together with the capacitor shifts the phase of the voltage over them in opposite direction to the phase shift, caused by g_{m4} and C_{p4}. This approach is necessary to be considered for each circuit separately. It is applied for some circuits [8, 9] and can help partly. Its disadvantage is different dependences of g_{m4} and compensating resistor from temperature, process, etc.

The last two circuits in Fig. 1 – the one in (f) and the gyrator biquads in (g) do not suffer from excess phase. This is because all OTA input capacitances in these circuits are in parallel to integrator capacitors and there is no place, where an extra phase shift can arise. This superior property of both circuits is only when they are realized with single stage CMOS OTAs. For example one of the first papers concerning the excess phase and its compensation is about gyrator circuit [8], of course when gyrator amplifiers are realized in different way.

V. CONCLUSION

A generalized consideration of G_m-C biquads based on two integrator loop configuration is done concerning the influence of OTA output resistances and input capacitance. These OTA imperfections have the most negative impact on the circuit behavior when OTAs are single-stage CMOS. The effects of both types of imperfections are considered separately – an approximate approach, however it allows more clear characterizations of the influences.

The OTA output resistances reduce pole Q in all circuits. In the most of the circuits except the gyrator biquad Q is defined by ratio between G_ms of some OTAs and the output resistances additionally reduce it. The gyrator circuit is an exception of this rule and its Q can be defined only from the output resistances of the OTAs in the gyrator, which makes the gyrator able to realize higher Q.

OTA input capacitances, when they are in parallel to amplifiers connected as resistors, change the phases of some voltages in the circuit, which causes intolerable increasing of Q and even instability. The effect is more visible when the circuit is intended for operation at higher frequencies. Again the gyrator is an exclusion – it doesn't suffer from this disadvantage.

REFERENCES

[1] T. Deliyannis, Y. Sun, J. K. Fidler, *Continuous-Time Active Filters*, CRC Press, 1999.

[2] P. V. Ananda Mohan, *VLSI Analog Filters. Active-RC, OTA-C, and SC*, Birkhäuser, 2013.

[3] S.-Y. Peng et al., A Power-Efficient Reconfigurable OTA-C Filters for Low Frequency Biomedical Applications, *IEEE Trans. on Circuits and Systems – I: Regular Papers*, vol. 65, No. 2, pp. 543-555, Feb. 2018.

[4] K. Kwon, A 50- to 300-MHz Gm-C Tracking Filter Based on Parallel Operation of Saturation and Triode Transconductors for Digital TV Tuner ICs, *IEEE Trans. on Circuits and Systems – II: Express Briefs*, vol. 62, No. 6, pp. 522-526, June 2015.

[5] J. Lechevalier et al, A Forward-Body-Biased Tuned 450MHz Gm-C 3rd Order Low-Pass Filter in 28nm UTBB FD-SoI with > 1dBVp IP3 over a 0.7-to-1V Supply, *Proc. 2015 IEEE Int. Solid-State Circuit Conference (ISSCC 2015)*, pp. 96-97.

[6] F. Houfaf, M. Egot, A. Kaiser, A. Cathelin, B. Nauta, A 65nm CMOS 1-to-10GHz Tunable Continuous-Time Low-pass Filter for High Data Rate Communications, *Proc. 2012 IEEE Int. Solid-State Circuit Conference (ISSCC 2012)*, pp. 362-363.

[7] B. Nauta, *Analog CMOS Filters for Very High Frequencies*, Kluwer Academic Publishers, 1993.

[8] V. Th. van Looij, K. M. Adams, Phase compensation in electronic gyrator circuits, *Electronics letters*, vol. 4, No. 20, pp. 430-431, 4th Oct. 1968.

[9] I. Uzunov, B. Nikov, S. Ouzounov, M. Hristov, Frequency Compensation in a Two-Integrator Loop Gm-C Biquad When Realized With Single Stage OTAs, *XXVI Internat. Scientific Conference Electronics – ET-2017*, Sozopol, Bulgaria, Sept. 13-15, 2017.

Using object recognition benefits in a system for fire detection in the nature

Maria Pavlova[1]

Abstract – **The subject of fire recognition is quite popular and many studies discus it. This paper presents a research on a fire recognition field. A camera mounted on the unmanned aerial vehicles (UAV) records videos of the surrounding environment. The method of recognising fire captured on the videos, part of this research, aims to reduce time needed to detect fire. The method assisted by OpenCV library capabilities recognises fire event. The results are described in this paper.**

Keywords – **Object recognition, OpenCV, Fire recognition, Machine learning.**

I. INTRODUCTION

Fire in a building or in the nature is a problem of very high concern. To find a way to prevent a fire is a task of significant importance. For the matters of the indoor fires there are already different fire preventions systems. These systems proof to be highly efficient in buildings and another indoor facilities.

The problem of detecting fire in the outdoors is quite complex. An efficient fire detection requires more advanced technologies such as neuro networks, object recognition and, in addition, unmanned aerial vehicles (UAV).

This paper presents the results of a research in the area of fire recognition in inaccessible locations. For the purposes of the research is used UAV with a small lightweight digital camera mounted on it. The camera records video clips of the surrounding environment and then transfers them to a remote workstation, running fire recognition software.

Nowadays, different object recognition algorithms are available for any purpose. One of them is Machine learning with feature object recognition used in the current research. For the purpose of the research, Python wrapper is used to encapsulate the C/C++ code, thus increasing overall performance. Additional benefit is that the wrapper can be used in any Python module and OpenCV library as such, and OpenCV opens a large spectrum of possibilities in the Machine Learning.

Video cameras and computer vision methods are used for fire detection in chemical factories and other high fire risk industrial. The Deep machine learning will increase the accuracy and reduces the costs for high level of security [8].

This paper has four contributions with following contents: the first contribution is introduction; second contribution of this paper is the software implementation; the third is about results

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

and the fourth contribution is the conclusion and the References. In each part of this paper the subject will be explained accordingly.

II. SOFTWARE DESCRIPTION

Fire detection using infrared camera [1], [2] is a popular decision among the building's fire detection systems. There are also researches, using this same camera for outdoor uses.

The solution using infrared cameras is unquestionably reasonable, but it gives its best results in short distances. When talking about open spaces, that's almost never the case – often the areas of fire are inaccessible and thus the results are usually far from the necessary precision.

Therefor this research is aiming to offer a fire recognition method in the outdoors with a camera, capturing the surface. The camera is shown on fig. 1
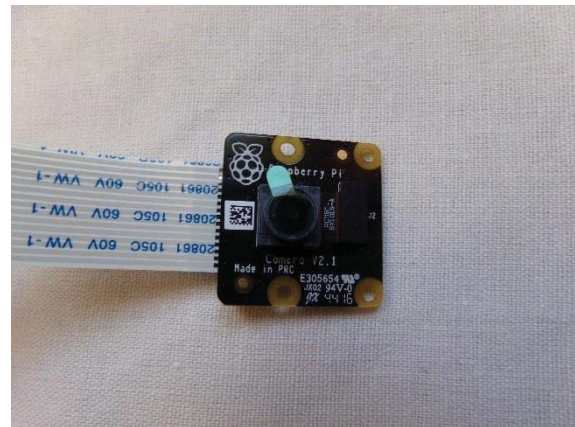
Fig. 1. Camera mounted on UAV

The application has been written and tested on Linux operation system. Captured video is processed by deep learning algorithm for detecting the fire in captured video.

There are many systems based on pixel object recognition like Pixel-to-pixel machine [4] and they have their application in their own fields. This approach is based on Paul Viola and Michael Jones algorithm for rapid and reliable object detection [3]. The algorithm which is proposed by Paul Viola is improved by Rainer Lienhart [5]. In this research the object detection classifier is based on features technique, which is faster than pixel system. Furthermore the complicated fire image (color, texture) will be recognized better with feature classifier. The algorithm allows for additional own classifiers to be created.

Basically the cascade of boosted classifiers works with haar-like features, from now on called classifiers. It is named cascade, because several simple classifiers (stages) are parts of

the whole classifier. The word boosted means that every stage of classifier is a result of another. It is possible to use one of four supported techniques – Adaboost, Real Adaboost, Gentle Adaboost and Logitboos. Adaboost is the technique used in this research. As mentioned before, the classifier has more than one classifier inside. The haar-like features are inputted data of the main classifier and it is deeply explained in [6].

Each feature used in a separate classifier is determined by shape, position, scale. The feature are computed by subtracting the sums of pixels by "integral image" in respected areas under white and black rectangles.[WK1] Black and white rectangles areas, which the image is separated. In their article "Rapid Object Detection using a Boosted Cascade of Simple Features" authors described deeply their method [3]. They apply the results for face detection classifier.

Base on above mentioned and the possibilities to prepare own cascade classifier, this research present the own cascade classifier fire detection. The algorithm use for classifier training is shown in table 1.

TABLE I
ALGORITHM FOR PREPARING THE CASCADE CLASSIFIER

1	Collect the negative images
2	Collect or Create positive images
3	Create a positive vector file by stitching together all positives
4	Train cascade classifier

The process of creating the classifier and all related works are presented in author's article "Object Recognition System Operating from Different Type Vehicles Using Raspberry and OpenCV". As the aim of this paper is to present the results of this research, the algorithm steps will be mark shortly.

To create the data base of negatives images is important task and first should be done. The data base should be at least 1000 pictures. Important requirement for the pictures: do not contain the object of recognition. In this case the fire.

Fig. 2. Result from creation of positive vector and stitching together all positives

Second step of this algorithm is creation of the positive data base. It may create from one image or collection of more. OpenCV has a powerful module opencv_createsampels [7]. With this function in terminal under Linux the script is one row:

"opencv_createsamples -img fire5038.jpg -bg bg.txt -info info/info.lst -jpgoutput info -maxxangle 0.5 -maxyangle 0.5 -maxzangle 0.5 -num 1300"

The result is very interesting and is shown in fig.2. We can see that the positive image fire is covered over the background image. The positive object is marked with red circle on the shown figure.

After preparation is possible to do the last step to train the cascade classifier. About cascade classifier is explained from Viola and Jones in "Rapid Object Detection using a Boosted Cascade of Sample Features" [10]. In OpenCV library this is one row of code, but behind this row has a lot of work for devising. Create simples first and after this is possible to train the stages. The number of stages is important. It depends of number of created samples. One towsend and two hundred positive samples will give opportunity around 10 stages. One stage contain the number of stage, the number of features in each stage and thresholds of each stage. On fig. 3 is shown stage 2 and stage 7 after cascade training.
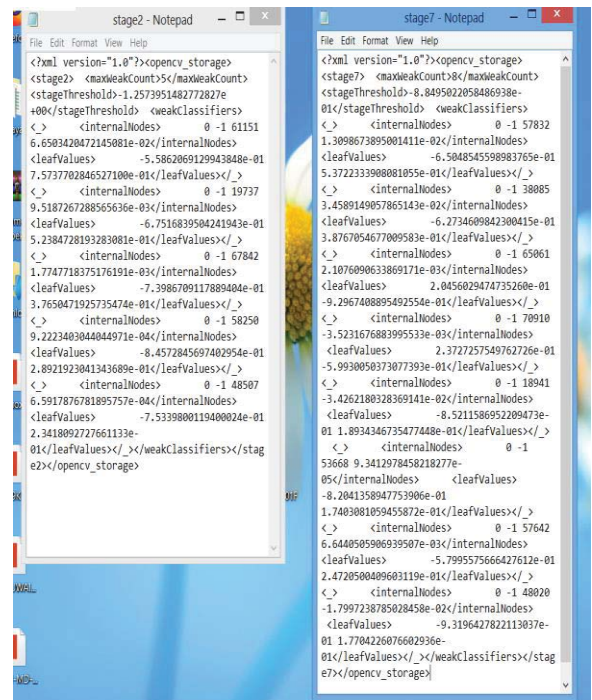
Fig. 3. Stage 2 and stage 7 after Cascade Training

III. RESULTS OF SOFTWARE IMPLEMENTATION

Finally the result of all this steps is the XML file from Classifier Training. This file is argument of function cv2.CascadeClassifier. Implementation of this function give result fire recognition. For this application this XML file is named cascade_training_fire.xml.

In additional one blue rectangle marks the recognized object. For this purpose is used the function cv2.rectangle(). Its

arguments are video or image, size and colour of the outline.

On fig. 3 is shown the screenshot of one of the videos, which are used for test. Blue rectangle define very well the fire and the smoke.

Fig. 3. Result from fire recognition

From another video we have also good results shown in fig. 4 and fig. 5.

Fig. 4. Positive result from fire recognition

On fig. 5 is shown very good result - the recognition of smoke. In this moment of video the classifier works really precise. That means that the training of machine for smoke gives very good results. From all tested pictures the smoke is recognized completely, with only one mistake: if there is the clouds in the picture the result will be wrong. It is shown in fig. 7. If we take into account the application area of the system, namely the observation of the ground surface, we may draw the conclusion that the program recognizes smoke in 98% of the cases.
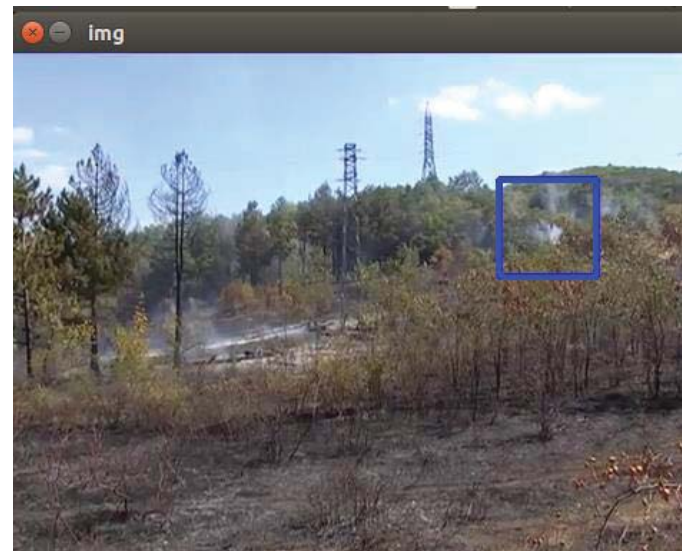
Fig. 5. Positive result from fire recognition

To recognize the fire precisely is task with high importance, because of people security and prevents of forest fire. It means the accuracy should be very high. But unfortunately in the results we have object recognition different than fire.

On fig. 6 is shown one of this negative results.
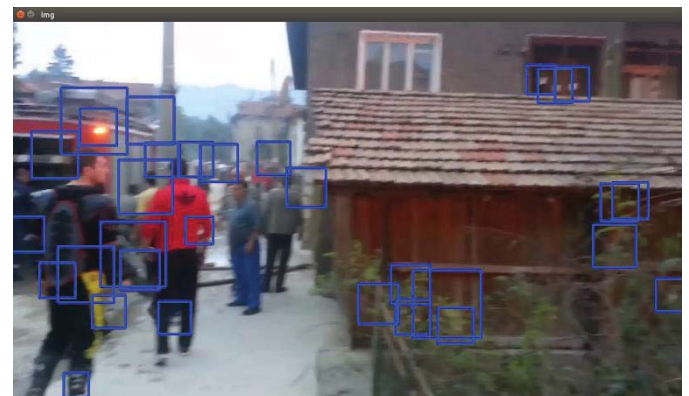
Fig. 6. Recognition of face instead of fire

The problem here is that the classifier recognize the face and the moving of all this people give the negative result. The problem with detection is serious and one of possible way to resolve is present in [9]. The authors have proposed a method, which extract the regions with movement. This regions are the potentials places with fire.

Second problem is that the machine recognise the clouds like smoke. This effect is shown in fig. 7.

Fig. 7. Recognizes the clouds instead of smoke

There is one more problem on this picture – the blue rectangle on the right side surrounds leaves instead of fire.

The last problem that was registered is that the classifier recognized the car lights instead of fire and this bad result is shown on fig. 8.

Fig. 8. Recognizes the car lights instead of fire

IV. CONCLUSIONS

The results of this research show that 40 % of all tested videos give bad results in fire recognition. In the remaining 60% of test videos the classifier recognizes the fire very well. It's clear that the described approach is necessary to be improved. This will be the next aim in this research.

REFERENCES

[1] https://doi.org/10.3182/20050703-6-CZ-1902.01380
[2] https://www.tandfonline.com/doi/abs/10.1080/01431169921229 0
[3] P. Viola, M. Jones "Rapid object detection using a Boosted Cascade of Simple Features", ACCVPR, 2001 https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf
[4] Keysers D., Deselaers T., Ney H. (2004) Pixel-to-Pixel Matching for Image Recognition Using Hungarian Graph Matching. In: Rasmussen C.E., Bülthoff H.H., Schölkopf B., Giese M.A. (eds) Pattern Recognition. DAGM 2004. Lecture Notes in Computer Science, vol 3175. Springer, Berlin, Heidelberg
[5] Rainer Lienhart and Jochen Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002. This paper, as well as the extended technical report, can be retrieved at http://www.multimedia-computing.de/mediawiki//images/5/52/MRL-TR-May02-revised-Dec02.pdf
[6] https://docs.opencv.org/2.4/modules/objdetect/doc/cascade_clas sification.html?highlight=cascadeclassifier#lienhart02
[7] https://pythonprogramming.net/haar-cascade-object-detection-python-opencv-tutorial/, 2017
[8] Hao Wu, Deyang Wu, Jinsong Zhao, An intelligent fire detection approach through cameras based on computer vision methods, Process Safety and Environmental Protection, Volume 127, 2019, Pages 245-256,ISSN 0957-5820, https://doi.org/10.1016/j.psep.2019.05.016
[9] Mahdi Hashemzadeh, Alireza Zademehdi, "Fire detection for video surveillance applications using ICA K-medoids-based color model and efficient spatio-temporal visual features, Expert Systems with Applications", Volume 130, 2019, Pages 60-78, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2019.04.019.
[10] Paul Viola, Michael Jones "Rapid Object Detection using a Boosted Cascade of Simple Features", ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001

.

Modeling of 3D Human Body for Photorealistic Avatar Generation: A Review

Nicole Christoff [1]

Abstract – **One of the challenges in computer graphics is the human body modeling. The adequate design of the human shape comes from the complexity of the anthropology itself. Numerous methods for the 3D human figure processing tasks have been provided for a range of identity-dependent body shapes and its various poses. In this paper, a survey of existing methods for 3D human body modeling is given.**

Keywords – **3D body shape, Human modelling, Geometric segmentation.**

I. INTRODUCTION

The way people perceive the world around them and interact with each other change; the current communication technology involves the transmission of audio and visual data between objects. This way, users can interact with others at a distance. Undoubtedly, video communication tools such as Skype and Viber are useful for many applications and tasks. However, text, 2D image and voice are not sufficient enough to satisfy as an alternative to human interaction, because there is no personal contact between the interlocutors and there is a limited sense of presence (sharing the same space). For the user these facts create a feeling of incompleteness and dissatisfaction with the communication process.

Currently, although in an early stage of development, augmented reality (AR), mixed reality (MR) and virtual reality (VR) offer great potential to include the five human senses in the communication process to become more meaningful and natural to all participants.

To simulate the particular physical function in the virtual environments, on each element is applied numerous algorithms. More particularly, those algorithms are developed for the human body models. Building a model involves many issues, such as 3D body training dataset preparation, designing a proper body model, and training the model to fit the prepared data.

The rest of the paper is organized as follows: in Section II, some of existing human body representations are briefly described. Some of free available databases are presented in Section III. In Section IV, the human body shape modeling techniques can be found. The conclusion is given in Section V.

II. HUMAN BODY REPRESENTATION

To compare a population groups and to create an anthropometric dataset, the anthropometric measurements can

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

be used as a basis. ISO 7250-1 gives such list of body measurement descriptions. A list of primary and secondary dimension indicators is given by ISO-8559 standard Part 2 [1]. Fig. 1 represents some of the measurement landmarks that are used by the both standards.
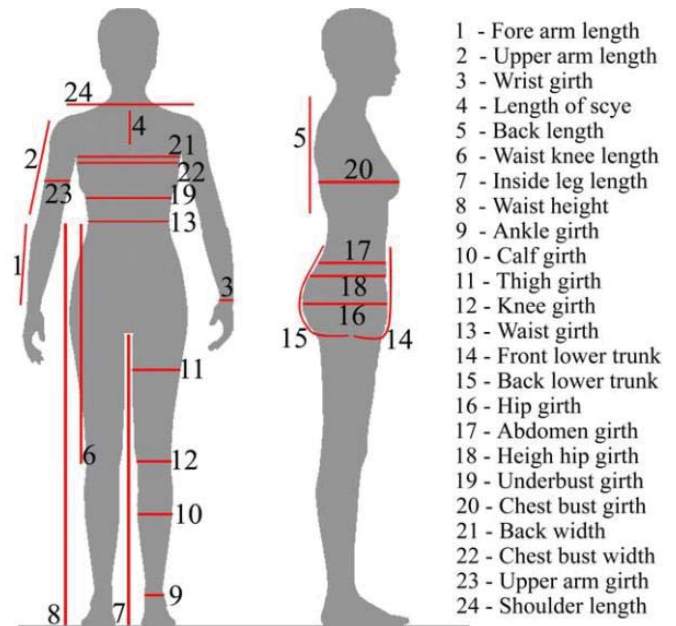
Fig. 1. Some of the major body measurement landmarks extracted from ISO-7250 and ISO-8559 standards (issue from [2]).

The anthropometric measurements are used to describe the human shape. They represent an estimation of the distances (linear and curvilinear) between anatomical landmarks or circumferences at particular regions of interests of the human body. For example, in medicine, in case of difficulty in determining the height of a patient, upper arm circumference is used for estimation of his body mass index [3]. Body mass index (BMI) is the ratio of the stature and the weight. The anthropometric measurements are widely used in many simulation systems [4].

Common anthropometric measurements include height (stature), weight (body mass), erect sitting height, triceps skinfold (upper arm girth), arm circumference (upper arm girth), abdominal circumference (waist circumference), calf circumference, knee height and elbow breadth [5].

Male (see Fig. 2) and female (see Fig. 3) bodies can be scaled by modifying standard anthropometric dimensions. Such examples are shown in Fig. 1 and Fig. 2, where height, weight and erect sitting height are used to represent a range of body sizes. Unfortunately, not all dimensions of the figures are

adjusted, which sometimes leads to unrealistic shapes of these scaled figures [4].
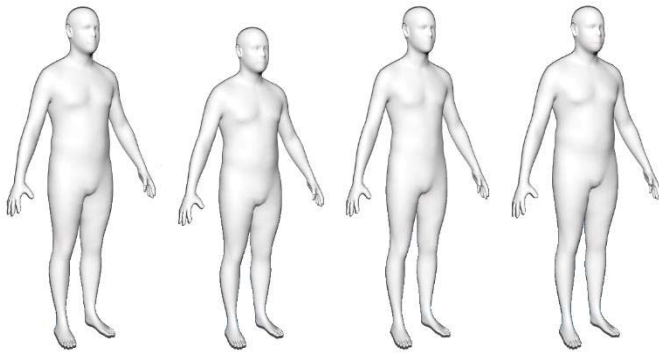
Fig. 2. Editing tree anthropometric parameters of a men's body. Left to right: average shape, with height decreased by 10 cm, weight decreased by 20 kg and decreased the ratio erect sitting height (decreased by 5 cm) to the stature (unchanged from the initial).
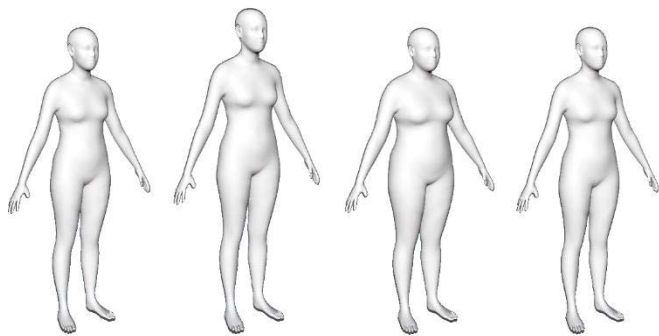
Fig. 3. Editing tree anthropometric parameters of a women's body. Left to right: average shape, with height increased by 10 cm, weight increased by 20 kg and increased the ratio erect sitting height (increased by 5 cm) to the stature (unchanged from the initial).

III. DATABASES

Regarding 3D body shape datasets, we can mention several of them, freely available online [6], [7]:

- Civilian American and European Surface Anthropometry Resource (CAESAR) project is the earliest research on anthropometric surveys conducted by 3D scanners. This is the largest commercially dataset, which contains raw 3D colored meshes with missing regions. It contains approximately 4000 human bodies of different shapes (male and female bodies) [6]. An example is shown in Fig. 3 a).
- Shape Completion and Animation for PEople (SCAPE) dataset is built by only one subject in different poses. As a result, 71 meshes using only geometric information, reconstructed from real data, are obtained [8]. In Fig. 3 b), one of those registration is visualized.
- FAUST - This data set containing 300 real, high-resolution human scans of 10 subjects in 30 poses, is created with a common template using a texture-based registration technique [9]. An example can be seen in Fig. 3 c).

- Dyna - Using over 40,000 scans of ten subjects, a dynamic body shape dataset is created by Pons-Moll et al. [10]. The dataset is a physics simulation of soft tissue motions in dynamic mesh sequences. A frame of a female subject is illustrated in Fig. 3 d).
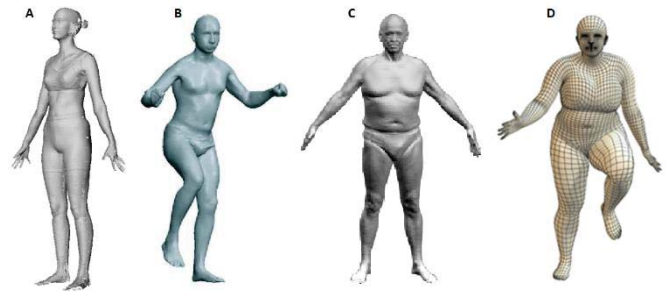
Fig. 4. Examples, from left to right, from CAESAR database [6], SCAPE database [8], FAUST [9] and Dyna database [10].

IV. HUMAN BODY SHAPE MODELING

The methods for the human body shape modeling can be categorized into four categories: direct model acquisition, image based reconstruction, template model-based scaling and statistics-based model synthesis [11].

A. Direct model creation

This method represents a construction of a human form and anatomical models via 3D scanning. 3D scanners or RGB R scanners with Kinect are used for the object/human model acquisition. The huge problem with this method is the holes and gaps in the scanned data, due to the self-contact (for example the armpits). This causes the topological structure of the scanned model. In the case of phase changes, the human will obtain unrealistic shape. Apart from the acquisition error of the capturing device (e.g. holes and noise in a scan), the clothes are additional difficulties for the body representation.

A good practice is meshing hands, feet and head separately of the body [4]. They contain more details, like the hair on the head.

Fig. 5. Example of scan of a human torso, clothed with a t shirt.

For example, in Fig. 5 a 3D scan of human (male) torso, clotted with a T shirt is presented. Referring the measurement landmarks (see Fig. 1), there are 11 landmarks which we want to obtain, but the information is missing (due to the T shirt, which gives wrong information about the real sizes): 2 - upper arm length, 4 - length of scye, 5 - back length, 13 - waist girth, 17 - abdomen girth, 18 - height hip girth, 19 - underbust girth, 20 - chest bust length, 21 - back width, 22 - chest bust width, 24 - shoulder length.

B. Image-based reconstruction

The image-based reconstruction, or photogrammetry is a process for recovering the exact positions of object's surface points applying measurements made in multiple separated 2D images from a set of angles. The creation of a digital human model consists of the following steps: 3D point cloud generation; structuring and modelling (segmentation the relevant anthropometric information and mesh generation); texturing and visualization [12]. The produced model often still contains gaps (the noise and depending on the background of the subject), but is cheap and accessible (requires a set of 2D images).

To create a 3D representation of a human figure, Ramakrishna et al. used 2D coordinates of anthropometric landmarks in a 2D image. The method can work with a simple body poses and estimates the camera using a matching algorithm, working on the image projections. Ramakrishna et al. propose a condition for the sum of squared limb-lengths, computed in closed form, necessary for the creation of realistic 3D human configurations. The method is robust to missing data [13].

C. Template model-based scaling

The template model-based scaling method is a process of deformation of a single template human body model to obtain an additional model. From this template model, its landmarks (limbs) and proportions are scaled and adjusted to create new body shapes. The first step of the template model is marking the specific regions and identifying them with unique numbers. After that, a correspondence of specific regions of the main body mesh (set of vertices) is done. This process is based on the measurement landmarks extracted from ISO-7250 and ISO-8559 standards, illustrated in Figure 1 [2]. It is a simple method, which generate new datasets with small resources (only one body scan).

Đonlić et al. [14] propose anthropometry-based approach can successfully segment various body-types. They utilize a generated 3D body model to segment six specific body parts of the human: head, torso, left and right arm, and left and right leg. Their approach is sensitive to outliers and the quality of the 3D reconstruction, due to the low-cost scanner that they had used.

D. Statistics-based model synthesis

A statistical shape model can be built using a set of examples of a shape. For the training set, each shape is represented by a set of n labelled landmark points. They are consistent from one shape to the next one. Given a set of such labelled training examples, they need to be aligned into a common coordinate frame. The idea is to minimize the sum of squared distances to the mean of the set. Then, each shape can be represented by a 2n element vector. The aligned training set forms a cloud in the 2n dimensional space. It can be considered as a sample from a probability density function, the cloud is approximated with a gaussian. The Principle Component Analysis (PCA) is used to pick out the main axes of the cloud. It models only the first few, which accounts for the majority of the variation.

First parametric modeling method of 3D body shape is that of Anguelov et al. [8], introducing the fundamental Shape Completion and Animation for PEople (SCAPE) method. SCAPE is a statistical model that captures variation in both individuals' shape and pose. The method is based on a decomposition of the body shape into both articulated and non-rigid deformations. The method is presented in Fig 6. The 3D body shape is represented as a triangular mesh. For a triangle face f_k, pose deformation Q_k and body identity-dependent shape deformations S_k, with identity-dependent shape parameters β, are applied. The deformation Q_k is predicted from the nearest body part joint angles α. To take into account all changes between different humans, the deformation matrix S_k, calculated from the parameters β in the PCA subspace, is computed. After that, each body part (a single bone) to which the triangle f_k is assigned, is rotated by an articulated rigid rotation R_k [15].
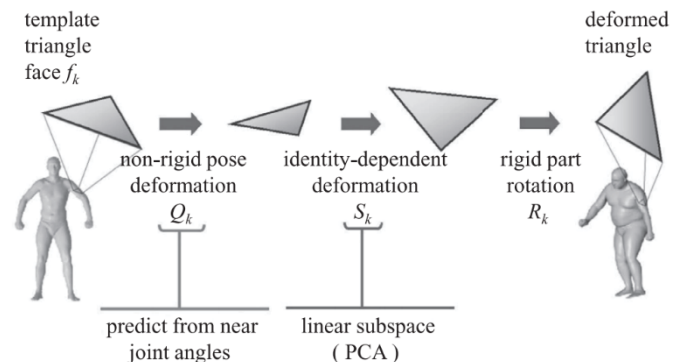
Fig. 6. Working mechanism of SCAPE (issue from [15]).

Many follow works are based on this method. For example, a variant of a SCAPE model is Loper et al. [16] and Pishchulin et al. [7].

Loper et al. [16] propose a realistic learned linear model of human body shape and pose, using PCA. The resulting principal components become body shape blend shapes. As a preprocessing step, authors register a template mesh for each scan and pose normalize the data. Loper et al. trained and tested the algorithm with the CAESAR dataset. On the trained meshes, they fit the algorithm and then compare the vertex errors. The similarity with SCAPE is that the Loper's model decomposes the body shape on the same way. The difference is that Loper's model is based on a vertex approach and corrective blend shapes are used [16].

Pishchulin et al. [7] uses a statistical model of 3D human shape and pose to rebuild the statistical body representation and learned from a database, based on CAESAR dataset [6]. For the human body reconstruction, they use sparse input data. They propose a method for scan alignment that quantitatively lead to the best learned models.

V. CONCLUSION

Study of methods for generating human body models with different measurement ranges is one of the main problems of computer graphics. It comes from the complexity of the anthropology itself. We have summarized the common technics used in such 3D human models, and mention some advantages and disadvantages. We have listed some of available 3D body datasets.

ACKNOWLEDGEMENT

REFERENCES

[1] https://www.iso.org/obp/ui/#iso:std:iso:8559:-1:ed-1:v1:en

[2] M. Kasap and N. Magnenat-Thalmann, "Parameterized Human Body Model for Real-Time Applications", International Conference on Cyberworlds (CW'07), pp. 160-167, 2007.

[3] H. A. Nygaard, "Measuring body mass index (BMI) in nursing home residents: The usefulness of measurement of arm span", Scand J Prim Health Care, vol. 26, no. 1, pp. 46-49, 2008.

[4] M. P. Reed, U. Raschke, R. Tirumali and M.B. Parkinson, "Developing and Implementing Parametric Human Body Shape Models in Ergonomics Software", Proc. 3rd International Digital Human Modeling Conference. Tokyo, Japan, 2014.

[5] D. J. Goldstein and B. McQuiston, "CHAPTER 40 - Nutrition and Renal Disease", Nutrition in the Prevention and Treatment of Disease, pp. 617-636, 2001.

[6] K. Robinette, S. Blackwell, H. Daanen, M. Boehmer, S. Fleming, T. Brill, D. Hoeferlin, and D. Burnsides, "Civilian American and European Surface Anthropometry Resource (CAESAR) final report", Tech. Rep. AFRL-HEWP-TR-2002-0169, US Air Force Research Laboratory, 2002.

[7] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt and B. Schiele, "Building Statistical Shape Spaces for 3D Human Modeling", Pattern Recognition, vol. 67, pp. 276-286, 2017.

[8] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers and J. Davis, "SCAPE: Shape Completion and Animation of People", ACM Trans. Graph, vol. 24, pp. 408-416, 2005.

[9] F. Bogo, J. Romero, M. Loper, and Michael J. Black, "FAUST: Dataset and Evaluation for 3D Mesh Registration", 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 3794-3801, 2014.

[10] G. Pons-Moll, J. Romero, N. Mahmood, and M. Black, "DYNA: a model of dynamic human shape in motion", ACM Transactions on Graphics, (Proc. SIGGRAPH), vol. 34, no 3, pp. 120:1-- 120:14, 2015.

[11] S.-Y. Baek and K. Lee, "Parametric human body shape modeling framework for human-centered product design", Computer-Aided Design, vol. 44, pp. 56-67, 2012.

[12] F. Remondino and S. El-Hakim, "Image-based 3D Modelling: A Review", The Photogrammetric Record, vol. 21, no. 115, pp. 269-291, 2006.

[13] V. Ramakrishna, T. Kanade, and Y. Sheikh "Reconstructing 3D Human Pose from 2D Image Landmarks", Computer Vision – ECCV, pp 573-586, 2012.

[14] M. Đonlić, T. Petković, S. Peharec, F. Berryman, and T. Pribanić, "On the segmentation of scanned 3D human body models", 8th International Scientific Conference on Kinesiology, Croatia, pp. 694-697, 2017.

[15] Z.-Q. Cheng, Y. Chen, R.R. Martin, T. Wu, and Z. Song, "Parametric modeling of 3D human body shape—A survey", Computers and Graphics, vol. 71, pp. 88-100, 2018.

[16] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black "SMPL: A Skinned Multi-Person Linear Model", ACM Transactions on Graphics (TOG), vol. 34, no 6, Article No. 248, 2015.

[17] D. J. Goldstein and B. McQuiston, "CHAPTER 40 - Nutrition and Renal Disease", Nutrition in the Prevention and Treatment of Disease, pp. 617-636, 2001.

On image preprocessing methods for preparation liver CT image series database for intelligent segmentation and classification algorithms

Melinda Kovács[1] and Szilvia Nagy[2]

Abstract – **In this paper entropy-, and other statistical parameters- as well as convolutional filter-based preprocessing methods are studied in order to prepare series of images from various phases of computer tomography takes about human liver. are given. The purpose is to build a database for intelligent image segmentation and lesion classifying methods with both the liver and the different types of roundish objects within the liver having their own masks as well as their classification. The masks resulted from our methods are to be cross-checked, labelled and corrected by medical experts manually, they only give a basis for helping to build the database.**

Keywords – **Image processing, liver, computer tomography, image series database.**

I. INTRODUCTION

Computed tomography was introduced by Godfrey Hounsfield in 1969. It is based on X-ray imaging, a method was used from the beginning of the 20th century, but unlike plain X-ray method, which produces 2D summational images from a body part, computed tomography takes a data map of X-ray absorbing measurements, from which a computer can create 2D or 3D objects. The first use in a real patient was a brain scan in 1971. Hounsfield (and Allan M. Cormack) received the Nobel Prize in Physiology or Medicine in 1979 for the development of computer assisted tomography.

As an essence of the CT machine, an X-ray tube and a detector row are opposed to each other. The detector row is measuring the output beam intensity integrated along a line between X-ray source and detector. It depends from the generation of the given CT machine, whether the X-ray tube is rotating or the detector row, or both, and how many tubes and detector rows are present. Since either the detector or the tube is rotating, data is produced from a given point's X-ray attenuation from many angles, which means that the attenuation of the point can be calculated.

"Tomography" means "imaging by sectioning". From the raw data of thin sections acquired by the machine, a series of algorithms create a thin slice and then a reconstructed image. The mathematical background of CT imaging reconstruction is the Radon transformation invented in 1917 by Johann Radon, who showed that a function could be reconstructed from an infinite set of its projections [1]. Stefan Kaczmarz developed a method to find an approximate solution to a large system of linear algebraic equations in 1937 [2].

Image reconstruction is a mathematical process that generates visual images from the X-ray attenuation data acquired by the CT machine from many different angles. Image reconstruction has an impact on image quality as well as on the radiation dose. Two major reconstruction method categories are exist. Analytical reconstruction methods are usually based on filtered back projection (FBP), which uses a 1D filter on the projection data before back projecting the data onto the image space. [3]. Iterative Image Reconstruction methods can be found in [4].

Medical imaging uses a special file standard named Digital Imaging and Communications in Medicine (DICOM). It is used for storing and transmitting medical images without data loss. It can be exchanged between to receiving-capable entities. DICOM Conformance Statements state the DICOM classes they support, and the standard includes the definition of the file format, and the protocol for network communication. Every radiological modality has a standard in DICOM [5]. In the case of computer tomography, images are stored usually as matrices of unsigned integers (even though the intensity corresponding to air is –1000 in hounsfield units (HU): everything is shifted by 1000) with dynamic range about 4-5000. The dynamic range of the liver in a CT image is usually not much more than 256. This means, that for image processing purposes, in the case of the liver, standard image formats are also sufficient, and probably more favorable, as the DICOM files have too high dynamic range, they may contain information about the patient and the environment, and they need special readers.

Mostly liver segmentation is done by intelligent systems, that need training sets and tests sets consisting of rather large number of images and masks. Preparation of such a training set is rather time and resource consuming. Also, medical experts are needed for checking and correcting the masks, even if the preparation is done by less qualified staff. Our goal is to provide preliminary masks for both the liver contour and the lesions in order to decrease the human labour need of building databases.

Also, usually independent images are used for image processing purposes, as for larger number of neighbouring images to be studied together much higher computational capacity is necessary. From the point of view of medical experts, however, individual images are usually hardly interpretable, thus it is more favourable to study the series of images in order to be able to determine the nature of the finding.

In this article we are presenting the first steps on building a database. After thresholding the image to the intensity domain of the liver, we apply different kinds of preprocessing methods in order to provide preliminary masks for the liver. As a next

[1]Melinda Kovács is with the Multidisciplinary PhD School of Engineering Sciences, Széchenyi István University, Egyetem tér 1. Győr, H-9026 Hungary.

[2]Szilvia Nagy is with the Department of Telecommunications at Széchenyi István University, Egyetem tér 1. Győr, H-9026 Hungary, E-mail: nagysz@sze.hu.

step we apply an active contour method for generating a more fitting mask and compare the resulting masks to one another. The applied preprocessing techniques are mean and median filtering, gradient filtering, calculating Rényi entropies for the environment of the pixel and structural entropies.

In Section II, we summarize the properties of the liver on a CT image of various phases, in Section III we give the applied mathematical tools. In Section IV. the results are shown and in Section V the results are summarized and a conclusion is drawn.

II. ON THE CT IMAGES OF HUMAN LIVERS

In diagnosing liver diseases especially tumours, usually contrast material is used, and CT image series are taken before injecting the contrast material (native phase), at the time when the contrast material is mainly in the arteries (arterial phase), at the time, when it is mainly at the veins (portal phase), and later, after ten-fifteen minutes (late phases). The most interesting and informative phases are the arterial and portal ones, though for some differential diagnosis purposes the other phases can be necessary, too.

The liver itself on CT images is a rather diffuse object, with high intensity variation within the organ from pixel to pixel, but generally a rather uniform average intensity throughout the whole organ, except for the blood vessels and the different lesions, like cysts, benign or malignant tumours, as it can be seen in Fig. 1.
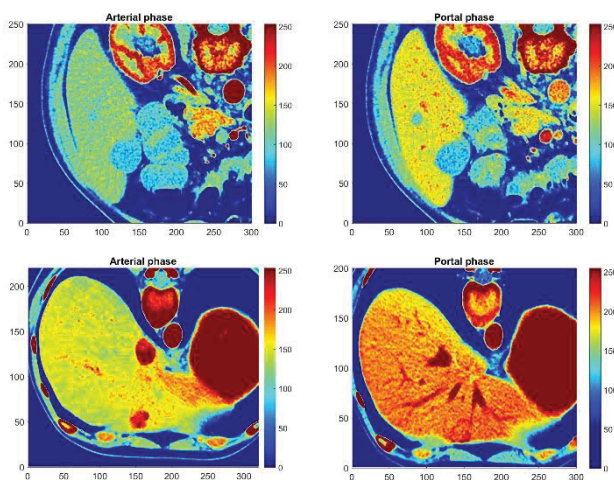
Fig. 1. Arterial and portal phases of two livers with different lesions. The upper row liver has a small cyst around position (90,130), which is more visible at the portal phase. The liver in the lower row has a larger FNH (focal nodular hyperplasia) around position (150,50), which is more visible at the arterial phase. The scale of the colors is shifted by −50 HU in order to show the most significant 256 HU wide part of the dynamic range, all the units below −50 HU are substituted by 0, and all the levels above 206 HU are substituted by 255.

It seems to be a rather simple thresholding problem to select the liver, however, the dynamic range of the muscle tissue usually overlaps with that of the liver, so it is rather complicated to distinguish between the stomach or blood vessel walls as well as the muscles between the bones and the liver, if they are touching, like in the bottom part of Fig. 1.

III. ON THE APPLIED FILTERS

We applied local filters with rather small radius (2, 3 or 5) in order to help the segmentation of the liver from the environment.

A. Mean, median, standard deviation, gradient

As a first step, we studied the mean and median filtered versions of the images. As awaited, they make the liver rather homogeneous, with small variation. We used the difference of the two filtered images, as it can be seen in Fig. 2., as the first possible tool for finding the borderlines of the liver and the lesions within.
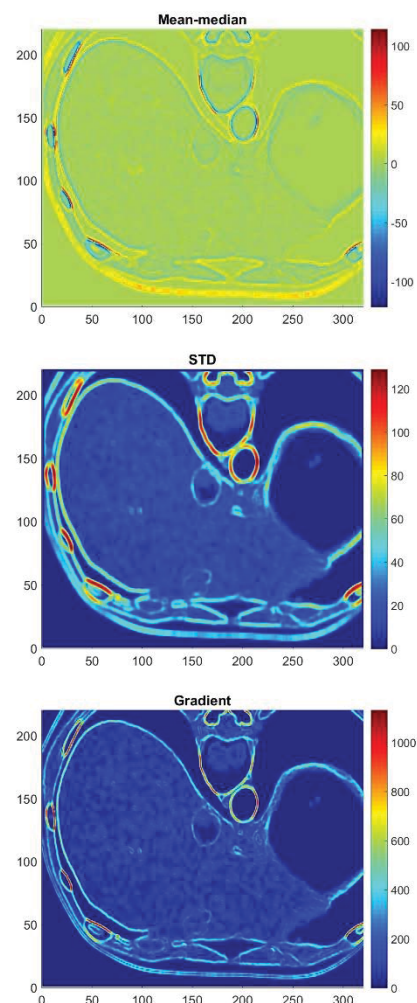
Fig. 2. Filtered versions of the third subfigure of Fig. 1 with the difference of mean and median filtered image, the standard deviation and gradient filters.

Also, the standard deviation was used similarly to the mean and median operations in a 5×5 sliding window: this was our second candidate for aiding the segmentation.

Gradients are usually the basis of edge detections, so we also applied gradient filters to the images. Their results can be seen in Fig 2. for the first subplot in Fig. 1. Both the gradient and the deviation are large at the borderlines of the liver and the tumor, while they remain rather small, but significantly larger than zero at the healthy liver parts.

B. Entropies

Also, the Rényi entropy based structural entropy and the filling factor was used as in sliding windows of size 5 by 5, similarly to the previous cases.

The Rényi entropy

$$S_n = \frac{1}{1-n} \ln\left(\sum_{i=1}^{N} I_i^{\,n}\right), \tag{1}$$

was introduced by Alfréd Rényi [7] and proved to be an extension of Shannon's entropy definition. Pipek and Varga used the differences of Rényi entropies, namely the

$$S_{str} = S_1 - S_2, \tag{2}$$

structural entropy and the logarithm of the spatial filling factor

$$-\ln q = S_0 - S_2, \tag{3}$$

to characterize the localization type, i.e., the average shape of electron distributions [8-10]. These quantities were generalized for any type of probability distributions and used for characterizing microscopy [11-13] and surface scanner images [14]. They were also applied antecedents in colonoscopy image analysis [15-17], based on the fuzzy classification method of [18-20]. Generally it can be said that for distributions with rapid decrease the structural entropy is larger, or the logarithm of the filling factor is smaller compared to distributions with less steep slopes.

In Fig. 3, as an example, the 1st and 2nd Rényi entropy filtered versions of the third image in Fig. 1 are given together with the structural entropy (2). (The zeroth Rényi entropy is constant, this is why the image from a filter based on (3) is not shown)

C. Thresholding and active contour

Usually the patients are diagnosed based on the ALARA principles, i.e., the radiation should be "As Low As Reasonably Achievable", which means that for some cases only one phase is available, in some cases more, depending on the purpose of the diagnosing process. As it can be seen from Fig. 1, the liver itself is much more separable in the portal phase image. Therefore we suggest starting the segmentation of the liver from the portal phase, if available.

The lesions however might not be distinguishable and very probably they are not diagnosable from a single phase only, this is why we suggest searching for smaller, roundish objects from the arterial and also the later phases as well.
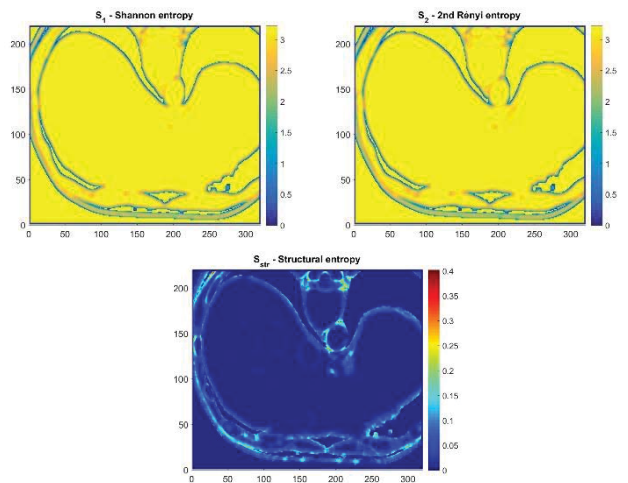
Fig. 3. Filtered versions of the third subfigure of Fig. 1 with Rényi entropies and structural entropy.

The first step for creating masks for the lesions is to provide a mask for a liver in order to limit the search to its area. Using a rather simple thresholding can mark the liver area rather well. If two masks are used, one for giving the area above the muscle tissue level, i.e., a lower mask, and one for filtering out the too high intensities (the bones, blood vessels, stomach or bowel content, etc.), their difference can be used as a basic mask for the liver and the blood vessels, stomach and other objects that are separated from the liver by a non-masked area. These later objects can be eliminated from the mask by an area filling algorithm, while the others usually have thin wall, which can be the basis of the separation, though this tends to leave not nice artifacts.

Using the area filled mask as a basis of an active contour algorithm [21,22] can provide more precise masks.

IV. Results

For the basis of the lower and upper masks mentioned in the previous subsection, the median filtered images are of much more use than the original or mean filtered versions, as it can be seen in the example in Fig. 4.

The area filling segmentation algorithm needs a point that is within the boundaries of the liver. Although people differ in size, points in the liver can be found, moreover, a basic form corresponding to the liver could also be found, similarly to [21] in order to have a larger starting are for the filling algorithms.
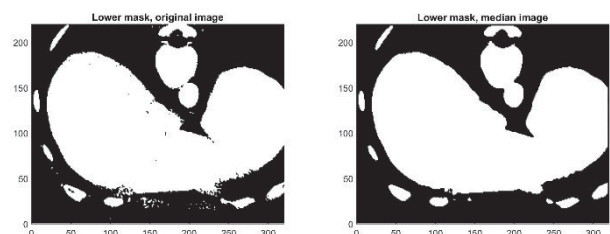
Fig. 4. Lower mask from the original and from the median filtered images. The threshold level is at 100 HU.

The conditions for quitting the filling for a line can be based on all the filtered images given in the previous section, however, the most effective one seems to be the gradient combined with the median and the structural entropy forming a condition:

- median larger than a threshold (100 HU), or some of the few neighbors from the center of the area are larger than the threshold
- gradient between the limits (10 and 200 HU)
- structural entropy decreasing to a small number (almost 0, e.g. 0.01) from its first local maximum from the center of the area

The combination of these three conditions can give rather
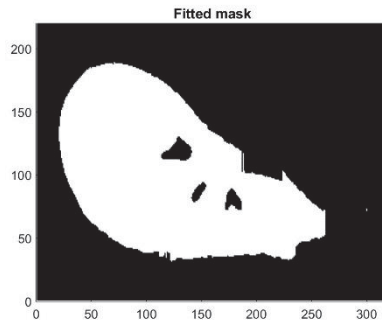
Fig. 5. Segmentation mask for the liver after active contour.

nice segmentations, as it is visible in the example in Fig 5.

V. SUMMARY AND CONCLUSION

Basics of a method for providing automatic masks for liver CT images is given. The masks are to be supervised manually in order to provide a database for machine learning algorithms and diagnostics based on computational intelligence methods.

The masks for the liver are generated from a threshold mask based on median filtered portal phase images, using area filling from a starting liver point. The conditions for filling can be based on several processed versions of the image, from gradient and standard deviation filtered to Rényi entropy filtered images. A combination of filling conditions is given using the median filtered image as well as the gradient and structural entropy images for determining the borderlines of the liver.

For segmentation of the lesions within the area of the liver, similar methods can be used: higher and lower median density roundish spots are to be searched.

REFERENCES

[1] H. Hornich, Translated by PC. Parks "A Tribute to Johann Radon". IEEE Trans. Med. Imaging.;Vol. 5, No. 4, pp. 169-169, 1986. DOI: 10.1109/TMI.1986.4307774

[2] S. Kaczmarz, "Approximate solution of system of linear equations", Int. J. Control, Vol. 57, No. 6, pp. 1269-1271, 1993. DOI: 10.1080/00207179308934446.

[3] A.C. Kak, M. Slaney, *Principles of Computed Tomographic Imaging*. SIAM, Philadelphia, PA, 1987.

[4] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, 2nd ed: SPIE Press, Bellingham, Washington, 2009

[5] C.E. Kahn Jr, J.A. Carrino, M.J. Flynn, D.J. Peck, S.C. Horii, "DICOM and radiology: past, present, and future". J. American College Radiol. Vol. 4, No. 2, pp. 652-657, 2007. DOI 10.1016/j.jacr.2007.06.004

[6] A.E. Flanders, J.A. Carrino, "Understanding DICOM and IHE". Seminars in Roentgenology Vol. 38, pp. 270–281. 2003

[7] A. Rényi, On measures of information and entropy. In Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, Berkeley, CA, USA, 20 June–30 July; pp. 547–561,1960.

[8] 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.

[9] 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.

[10] Sz. Nagy, B. Sziová, J. Pipek, "On Structural Entropy and Spatial Filling Factor Analysis of Colonoscopy Pictures", Entropy, Volume 21(3), ID: 256, 32 pages 2019.

[11] 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.

[12] 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.

[13] A. Bonyár, AFM characterization of the shape of surface structures with localization factor. Micron, 87, 1–9. 2016.

[14] 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.

[15] Sz. Nagy, F. Lilik, L.T. Kóczy, Entropy based fuzzy classification and detection aid for colorectal polyps. In: Proceedings of the IEEE Africon, Cape Town, South Africa, 18–20 September 2017; pp. 78–82.

[16] 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. In Proceedings of the FuzzIEEE, Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7.

[17] Sz. Nagy, B. Sziová, L. Solecki, The effect of background and outlier subtraction on the structural entropy of two-dimensional measured data, under publication at IJRIS 2019.

[18] F.; Lilik, J. Botzheim, Fuzzy based Prequalification Methods for EoSHDSL Technology. Acta. Tech. Jaurinensis, Volume 4, pp. 135–144. 2011.

[19] F. Lilik, L.T. Kóczy, Performance Evaluation ofWire Pairs in Telecommunications Networks by Fuzzy and Evolutionary Models. In Proceedings of the IEEE Africon, Pointe-Aux-Piments, Mauritius, 9–12 September 2013; pp. 712–716.

[20] F. Lilik, Sz. Nagy, L.T. Kóczy, Improved Method for Predicting the Performance of the Physical Links in Telecommunications Access Networks. Complexity **2018**, Article ID 3685927 2018.

[21] A. Mihaylova, V. Georgieva, Spleen segmentation in MRI sequence images using template matching and active contours Proc. Computer Sci. Vol. 131, pp. 15-22, 2018.

[22] V. Georgieva, S. Ermakov, GUI for CT image segmentation via active contours, 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 2016.

Estimating the Width of an Oblong Shape Using a Single Camera

Vladimir Sibinovic[1]

Abstract – **In this paper a hypothesis is tested that a single, cheap web camera can be used to estimate the width of an oblong shape, e.g. wire or thread. The designed system is calibrated and tested under controlled parameters. The goal is to achieve the error in estimation less than ±0.1 mm.**

Keywords – **Camera, Image processing, Estimation.**

I. Introduction

Manufacturing industries rely on real-time data input for continuous and unattended process optimization [1]. Because this is a crucial part of the production process, expensive and precise laser sensors are usually used for on-line and non-invasive width measurement. With the availability of cheap electronics nowadays, many simple processes are being automated with the focus being on price. With that in mind, we try to estimate the width using cheap CMOS camera and image processing.

For the purpose of this paper a testing rig has been created to provide a fixed distance between the camera and target plane. The rig was printed on a 3D printer and is shown on Figure 1. Camera is position at 30mm from the base which is used as the target area. This distance was chosen arbitrary as a lowest distance at which the camera produces clear images.
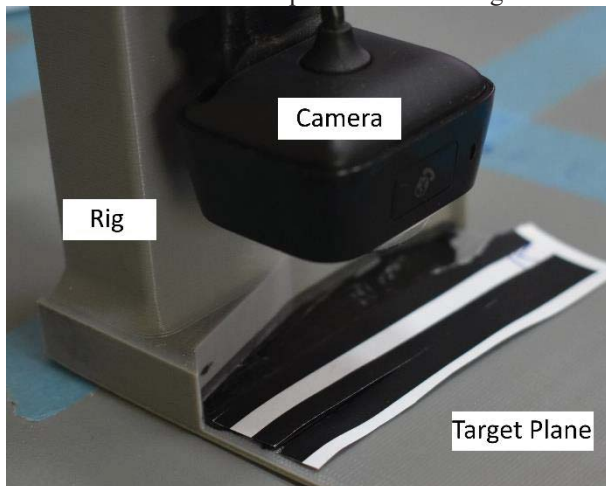
Figure 1. Testing rig for the system

[1] Vladimir Sibinovic is with the Faculty of Electronic Engineering at University of Nis, Aleksandra Medvedeva 14, Nis 18000, Serbia, E-mail: vladimir.sibinovic@elfak.ni.ac.rs

Camera is positioned so that the oblong shape target, e.g. white line on a black background is on the vertical centre of the image and spans across the whole image on the horizontal axis.

II. Setup

The system is based around Microsoft LifeCam-5001 HD web camera, and its parameters are shown in Table 1.

Table I
Camera Parameters

Resolution	1280 x 720
Field of View	66°
Communication	USB 2.0
Sensor	OV9712

The camera uses OmniVision CMOS sensor that has image area of 3888 μm x 2430 μm, with a pixel size of 3 x 3 μm.

The area that the camera can capture at the distance of 30 mm, from camera front, was experimentally determined to be 43 x 24 mm. From this and the vertical resolution of the camera we can calculate that 1 pixel represents 0.033 mm on the target plane.

By using simple geometry shown on Figure 2, and trigonometry we can calculate the exact distance of the sensor from the target plane.
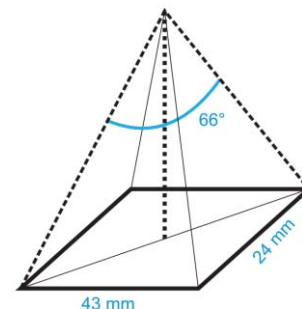
Figure 2. Field of view geometry

The determined distance is 37.9 mm. Which correlates to the camera geometry, as we defined the distance of 30 mm from the front of the camera and the camera sensor is about 8mm inside of the camera.

III. Image processing

Image processing was done using OpenCV library implemented on a PC in C++ programming language. First a picture is loaded in grayscale. Since the effect of color is not the scope of this paper, we have chosen a white object on black background to provide maximum contrast. Next, a Gaussian blur was applied. This blurs the sharper edges, and has proven to increase the accuracy of the system.

After a detailed analysis of different thresholding options [2], the best results, for the given problem, were achieved using binary threshold with Otsu's method [3]. After the preparation of the image the user moves the region of interest and positions it over the area where the estimation is needed. User can define the length and width of the ROI window using the provided GUI. The key objective here is to select a ROI that has the target represented as white pixels. Figure 3 shows the GUI and Figure 4 shows a typical ROI.
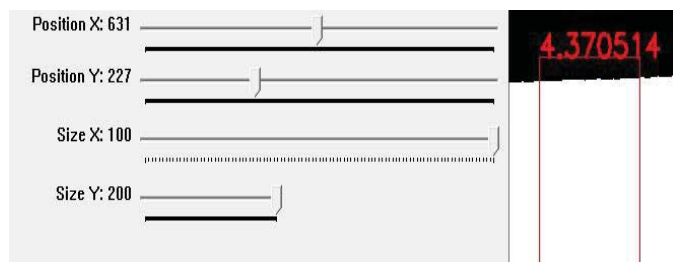
Figure 3. Graphical User Interface of the system

Figure 4. Typical extracted ROI from which the system computes the width of a shape

After selecting the ROI, the user starts the estimation process. On the ROI the system goes through all of the pixels in each column and calculates the number of white pixels in each column. When it reaches the end of ROI it calculates the mean value of the width of the white pixels in the ROI. These values are then used for calibration and later for estimation.

IV. Calibration

To calibrate the system, we first take known widths, and calculate the relationship between the mm and pixels. For setting referent values we used Wurth Digital Venire Caliper

that has an accuracy of 0.02 mm in the range from 0 to 100 mm. A range from 1mm to 4 mm was used as referent values. A sample step of 0.2 mm was used, with an exception of the region from 2.6mm to 3mm, where the sample step was 0.1 mm. This is a special region of interest for further work and that was why the sample step was lowered. Figure 5. shows the gathered data and the fitted line with R^2 of 0.9998 and can be represented as Eq. 1:

$$y = 35.084x + 0.9808 \qquad (1)$$
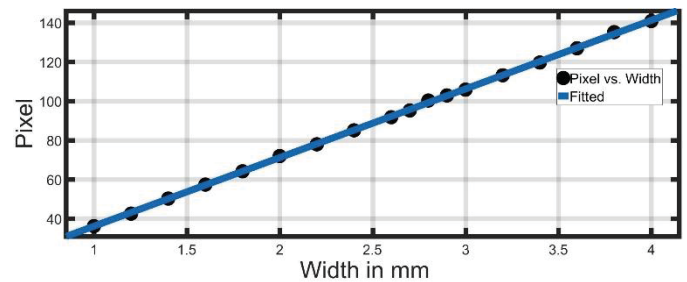
Figure 5. Calibration data and fitted line

By using Eq. (1) to calculate the width we can determine the error of fitting. The results are shown in Figure 6.
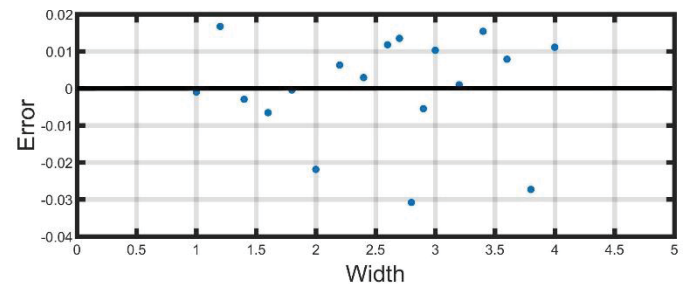
Figure 6. Error of fitting

The average absolute value of the error is 0.011 mm and the maximum positive error is 0.017 mm and the minimal negative error is -0.031 mm. This is satisfactory as a fitting error.

V. Testing

Testing was done using two different test sets. The initial test was done using a range from 1 to 5 mm of width with a step of 1mm. And the second test sample was larger with a smaller step size of 0.3mm. The results from the first test gave and absolute average error of 0.037 mm, which encourage us to test on a larger sample. The error in estimation from the second test is shown in Figure 7.
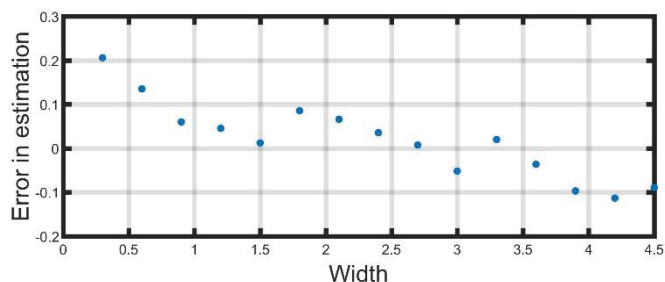
Figure 7. Error in estimation

The error of estimation is in the range from -0.113 mm to 0.206 mm. The error is larger at the beginning of the range and at its end. In the region from 1mm to 4 mm the error is within ±0.1 mm. For testing 10 different points ware used for each measurement, and the mean was used as a final estimation.

Using one arbitrary image the system was tested for continuity along the x-axis of the image. A sliding window of 20 pixels was positioned at 10 different positons along the x-axis and the estimation of width was done. The results are shown in Figure 8.
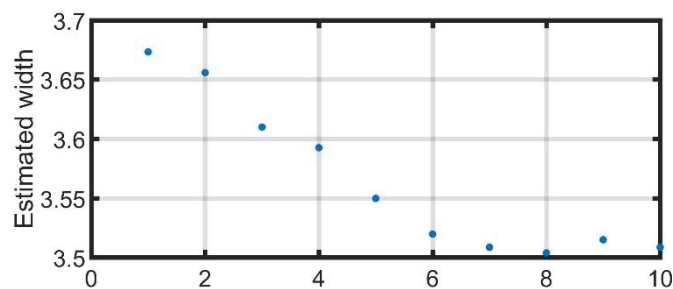
Figure 8. Estimation at different points along the x-axis.

From this we can conclude that the x position plays a role in the estimation error. The error is large when working with the left part of the image. This will be investigated further, but our opinion is that it may be from the calibration. During the calibration of the system the calliper used was positioned so that it is in the right part of the image. For the y-axis all of the pictures taken ware so that the centre part of the image is in line with the target. In this paper we assumed that the camera lens is centred and symmetrical, but the provided data suggest otherwise. This will be further investigated determining intrinsic camera parameters by means of camera calibration [4].

VI. CONCLUSION

In this paper a system for width estimation of an oblong shape has been designed and tested. The system has proven that it has potential to achieve the required estimation error of less than ±0.1mm.

The results from the initial tests show that the system has a small error of estimation, but upon further investigation the error has shown correlation with the position of measurement along the x-axis. This system can be used as a cheap non-invasive on-line width measurement system for processes where the tolerances are not the vital.

For further work other cameras will be tested against the same datasets, and intrinsic camera parameters will be investigated for their role in the process of estimation.

Another direction that we hope to investigate is the use of the proposed method as a basis for roundness estimation, as this is also a parameter that needs to be measured in the production of wire, pipes, filament, and etc.

ACKNOWLEDGEMENT

REFERENCES

[1] P. Cielo, "Optical sensors for on-line inspection of industrial materials", Optical Engineering, 32(9),2130, 1 September, 1993.
[2] M. Sezgin, B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", Journal of Electronic Imaging 13(1), 1 January 2004.
[3] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Trans. on Systems, Man, and Cybernetics, Vol.9, Issue:1, January 1979.
[4] Z. Zhang, "A flexible new technique for camera calibration", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol:22, Issue:11, November 2000.

Local-adaptive Enhancement of Details in Thermal Images

Velin Iordanov[1] and Ivo Draganov[2]

Abstract – **In this paper two algorithms are compared for local adaptive enhancement of contrast of thermal images. They are a modified adaptive gamma correction and local adaptive multiscale combining. The second is recursively applied at different scales preceded by smoothing with a Gaussian kernel and weighted averaging with consequent combination of resulting images. It is believed that this algorithm is also applicable for details highlighting. The analysis is made over a set of thermal images varying in contrast, details and overall content. Positive results are obtained which prove the applicability of the second approach for real-world implementations.**

Keywords – **Sharpness Enhancement, Local-adaptive Operator, Gaussian Kernel, Gamma Correction, Thermal Image.**

I. INTRODUCTION

Contrast enhancement and image sharpening play a crucial role in digital image processing related to various tasks from tracking vehicles to object spotting, from establishing medical diagnosis to forensic analysis and many more. Two major branches may be followed along the development of these techniques – the local and global approach. The first allows better processing of finer details in images but takes longer times to execute while the global algorithms are faster but lowering the finer details after processing.

In [1] Kim et al. present a histogram equalization algorithm which employs block overlapping for a sequence of images contrast enhancement. It is addressed towards security applications, mainly domestic surveillance, in low light conditions. Results are presented in qualitative manner with image comparison among various other methods with visible increase of the quality.

Stark [2] makes a step further into proposing generalization of the histogram equalization. Using a cumulation function in order to make mapping between local histogram the author achieves to get a wide set of degrees in contrast change – from untouched intensities to complete equalization.

Another approach proposed by Chang and Wu [3] includes analysis of the local standard deviation and then applying histogram transformation. This approach is considered especially effective in medical image processing following radiography.

Despite the high quality images obtained by described

algorithms it becomes necessary over time for a number of applications to have implementations fast enough working in or close to real-time [4, 5].

In our study we are comparing two algorithms, originally developed for contrast enhancement – the adaptive gamma correction [6] and local-adaptive multi-scale contrast enhancement [7]. The testing is oriented specifically towards thermography applications and as the results described below reveal the second approach could be very well used for enhancing the details in such thermos images.

In Section II detailed description of both algorithms are given followed by experimental results in Section III and then a conclusion is made in Section IV.

II. COMPARED ALGORITHMS

Input data:

Grayscale valued digital image with P pixels horizontally and Q pixels vertically. The brightness I varies from 0 to 255 for each pixel. It is located by path and filename. The software implementation needs to find P and Q automatically. If the input file contains a color image, only the brightness component (e.g. Y from the $YCbCr$ color space of a JPG file) is used.

Output data:

The program implementation outputs a non-compressed image to external memory (e.g. HDD) as bmp. If initially the processing started with a halftone image, it must also be a bmp. If it is a colored one, the newly obtained array after processing is the Y component while the color-difference components (Cb and Cr) are the same as those in the original file, and the resulting file is bmp.

A. Adaptive Gamma Correction

Algorithm:

Gaussian filtering is applied to the input image - the σ (standard deviation) needs to be set by the program interface.

For every pixel of the filtered image $i_g(x, y)$ the parameter is:

$$\gamma (x, y) = (i_g (x, y) - 128) / 128. \qquad (1)$$

The output image is obtained from the input by the formula:

$$o(x, y) = 255 (i (x, y) / 255).|\gamma (x, y)|. \qquad (2)$$

The difference with the original approach is that in the current realization the gamma is taken as absolute value. Testing proved that negative values close to -1 when $i (x,y)$ is close to 0 leads to overflow of the types used (double) due to the extremely small values of $o(x,y)$. The effect on the image quality at the output is thought to be not less than that of reported results from testing of the initial implementation.

The output image is saved in the same format, e.g. bmp.

[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

B. Local-Adaptive Multiscale Processing

Specific input parameters:

Proportionality factor - α (fractional coefficient in floating point representation) - is set by the user at startup of the program.

Number of processing scales - k (integer = 1, 2, 3, ...) - set by the user at startup of the program.

Cutting threshold (integer) - $Amax = \{0, 255\}$ - set by the user.

Scaling factor - m - integer (2, 3, ...) - set by the user, default is 3.

Algorithm:

The average image brightness is calculated by:

$$M = sum\ (I\ (i,\ j))\ /\ (P.Q),\ i = \{0,\ (P\text{-}1)\},\ j = \{0,\ Q\text{-}1)\}. \ (3)$$

The image is passed k times:

First pass - a 3x3 window is used to scan the entire image starting from its first row and first column, going to the right in the row to the end of the pixel columns, then the window passes to the second row and so on to the end of the last row. For each mask (window) position a calculation is done for the local average brightness of the pixels in the mask, the local standard deviation from the pixel brightness over the same area (the value of each mask pixel is derived from the average mask brightness, all the differences are raised to a second degree and the resulting numbers are added together and the result is divided by the number of pixels in the mask). It is found: $(\alpha.M\ /\ s)$ -1. If this number is greater than $Amax$, it is acknowledged as equivalent to $Amax$. If it is less than zero, it is confirmed as equal to zero. The center pixel's brightness in the mask is subtracted from the average of all masks $'a'$ and multiplied by $((\alpha.M\ /\ s)$ -1). After all the picture's passes, all newly received values for the brightness in a pre-allocated array in the memory are saved.

Second pass - a $3m$ x $3m$ pixel window is used. All processing steps are repeated as in the first pass. All new brightness values are stored in a separate memory array.

Then the scanning continues - the window size is $3m\ \hat{}\ (k\text{-}1)$ x $3m\ \hat{}\ (k\text{-}1)$ pixels. The steps are the same every time and each time the results are saved in a separate array for the new intensities.

Summing the array with original image brightness levels and those from each pass then follows - getting a resultant array of the same size - P x Q pixels.

Default values:

Example values for initial tests: $\alpha = 0.5$, $m = 3$, $k = 4$ (4 windows with 3x3, 9x9, 27x27 and 81x81 pixels), $Amax = 100$. Input images use low-contrast images, including infrared ones.

III. EXPERIMENTAL RESULTS

The test set of images contains 10 IR shots (grayscale) with dimensions of 640x512 pixels. They are part of the FREE FLIR Thermal Dataset for Algorithm Training [8] The intensity resolution is 16 bits per pixel. All images are saved in tiff container using LZW compression.

The interface of the implementation for both algorithms is web-based generated by C# tools. Fig. 1 depicts the entry point for the local adaptive gamma correction.

Fig. 1. Adaptive Gamma Correction GUI

The local-adaptive contrast enhancement algorithm at various scales is shown in Fig. 2.
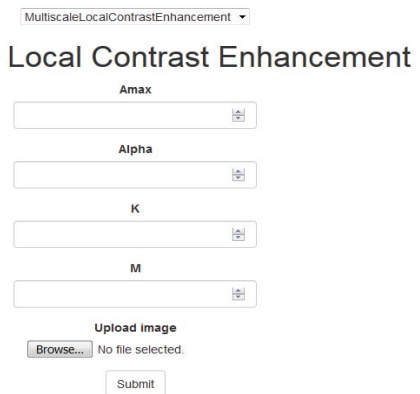
Fig. 2. Multiscale contrast enhancement GUI

Test image 1 is shown in Fig. 3.

Fig. 3. Test image 1 – original

From the results obtained, we notice a multiple increase of the contrast in the algorithm with adaptive gamma correction without major changes in sharpness (Fig. 4).

Fig. 4. Adaptive gamma correction result for image 1

Some images seem to lose part of the details after processing. In the multi-scale processing algorithm, we notice multiple sharpness improvement for all images and contrast enhancement for part of the images (Fig. 5-8).

Fig. 5. Multiscale local-adaptive processing at $k=1$

For image 1, with an adaptive gamma correction, a high illumination of the image is noticeable, but part of the detail is lost - the second car on the right is hard to see, but the details of a darker fan become clearer as a ladder in the distance, for example. After applying the multi-scale pass for image 1, we notice a darker picture than that for the adaptive gamma correction for all passes of 1 - 4. We notice best sharpness when processing with 1 or 2 passes. Better outlines of details in the picture are noticed, for most cases we get the best visibility in 4 passes - some of the pictures can be seen in details that are not visible in the original shots.

Fig. 6. Multiscale local-adaptive processing at $k=2$

Fig. 7. Multiscale local-adaptive processing at $k=3$

Fig. 8. Multiscale local-adaptive processing at $k=4$

The average sharpness over all test images for the tested algorithms is given in Fig. 9.
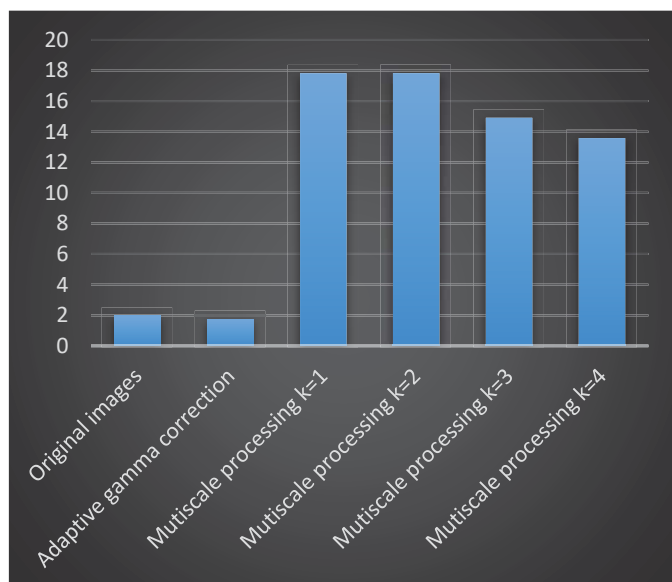
Fig. 9. Resulting image sharpness

For some images, it is clear that the adaptive gamma correction gadget worsens detail, while a multi-step scan algorithm increases it several times.

The RMS contrast for all cases is given in Fig. 10.
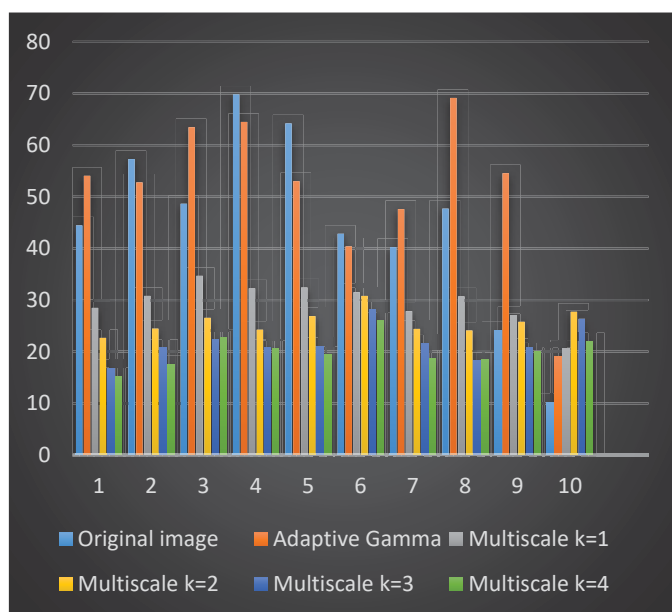
Fig. 10. RMS contrast prior and after processing

From the graph we see that the average image contrast obtained with an adaptive gamma correction algorithm has a higher value than the image after a multi-scale pass.

We could get better values than the multi-scale pass if we provide appropriate mask size, coefficient of proportionality, and cut-off values for each image. There are cases where the default values give the best result – as for image 2 at k=2.

The significant difference in RMS contrast for the images processed by the adaptive gamma correction and the multi-scale local-adaptive algorithm is a result mainly on the fact that the latter operates on a several different scale levels combined at the end. For each of these levels different statistical parameters are found which give good result after applying the transform function – better contrast and details in the same time over the area of the mask. When fusing 2 or more of the areas converted by this approach the details for each are well preserved and leads to even better detailed picture as a final representation but due to interfering effects among areas with varying contrast the overall RMS value tend to be smaller than that of the adaptive gamma correction. A possible solution to this problem is to embed a map of the edges and smaller details in the resulting picture by the multi-scale approach over already enhanced in contrast image by another, possibly globally working, algorithm.

IV. CONCLUSION

In this paper two algorithms for local adaptive contrast enhancement are compared by testing over a set of infra-red images – the adaptive gamma correction and local-adaptive multiscale contrast enhancement. The adaptive gamma correction proves better than the multi-scale approach taking into account the RMS contrast achieved but lots of details are lost from the original image. With the increase of the number of processing scales within the second algorithm the sharpness of the images becomes higher and higher and more and more details become visible. This approach is considered quite promising for large number of applications such as objects tracking, texture extraction, and various patter recognition tasks as a pre-processing step.

REFERENCES

[1] Kim, T. K., J. K. Paik, and B. S. Kang, "Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering." IEEE Transactions on Consumer Electronics 44.1, pp. 82-87, 1998.

[2] Stark, J. A., "Adaptive image contrast enhancement using generalizations of histogram equalization." IEEE Transactions on image processing 9.5, pp. 889-896, 2000.

[3] Chang, Dah-Chung, and Wen-Rong Wu. "Image contrast enhancement based on a histogram transformation of local standard deviation." IEEE transactions on medical imaging 17.4 (1998): 518-531.

[4] Yu, Z., C. Bajaj, "A fast and adaptive method for image contrast enhancement." In Proc. of the 2004 IEEE International Conference on Image Processing, ICIP'04, Vol. 2, pp. 1001-1004, 2004.

[5] Wang, Z., J. Tao. "A fast implementation of adaptive histogram equalization." In Proc. of the 8th IEEE International Conference on Signal Processing. Vol. 2., 2006.

[6] Dijk, J., et al. "Local adaptive contrast enhancement for color images." Visual Information Processing XVI. Vol. 6575. International Society for Optics and Photonics, 2007.

[7] Schutte, K. Multiscale adaptive gain control of IR images. In Infrared technology and applications XXIII, International Society for Optics and Photonics, Vol. 3061, pp. 906-915, August 1997.

[8] FREE FLIR Thermal Dataset for Algorithm Training https://www.flir.eu/oem/adas/adas-dataset-form/, last visited May 20th, 2019.

Simulation Model for Medical Images Watermarking using Wavelet Transform and DCT

Rumen P. Mironov[1], Stoyan Kushlev[2]

Abstract – **A simulation model of algorithm for digital watermarking of medical images using Wavelet transform and DCT is presented. The model is implemented on a signal processor TMS 320C6713 and its experiments indicate it can be used in real-time systems. The developed algorithm for watermarking ensures high transparency of the digital 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 evaluation of mean-squared error and signal to noise ratio of the reconstructed images.**

Keywords – **Medical Image Watermarking, Wavelet Transform, DCT, Unitary Transforms, Matlab Simulation.**

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.

Every system for image watermarking is 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].

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 [8], [9].

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 [10].

In the available literature, there have been various watermarking realizations using Matlab Simulink environment [11], [12], [13], [14] and etc. In all of them two-dimensional simulations of Discrete Cosine Transform (DCT) and / or Discrete Wavelet Transform (DWT) have been examined and individual solutions for different types of attacks have been considered. In these articles are discussed different implementations of watermarking algorithms using MATLAB Simulink environment based on DCT, DWT and combination of DCT&DWT transforms. Simulation results show that DWT is somewhat better than DCT but combination of these two transforms gives much better results than individual one.

In the present work a simulation model on Matlab Simulink environment of developed by the authors' algorithm for digital watermarking of medical images using Wavelet transform and DCT is described [15]. The emphasis in the development has been placed on the selection of the most suitable embodiment directed to applications in the medical field. This is extremely important with regard to the specifics of medical images and their use as was discussed at the beginning of the consideration.

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

---
[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

over the three test medical images are made on the base of evaluation of the 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 with median filters and average filters. The developed on the MATLAB simulation model is experimented by the personal computer with 3.2GHz Core-i5 processor and specialized signal processor board - TMS 320C6713 DSK.

II. SIMULATION MODEL DESCRIPTION

The common block scheme of the new developed algorithm for digital watermarking is discussed in the article [15]. The principal block scheme of simulation model, representing the embedding of the watermark is developed using MATLAB Simulink environment and is presented in Fig.1.
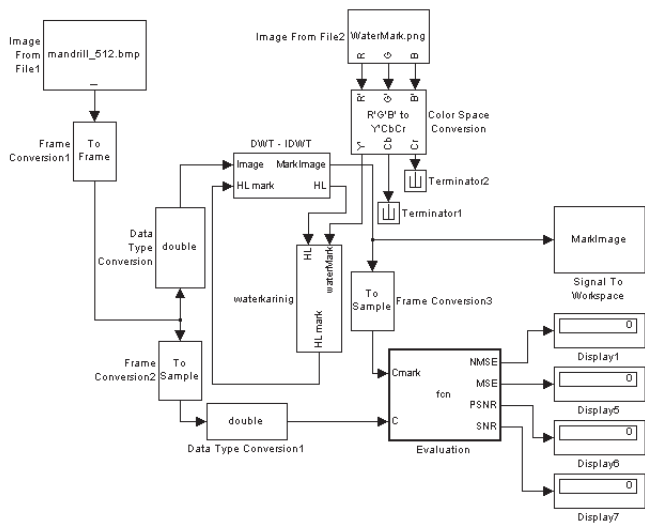
Fig.1. Simulation model, representing the embedding of the watermark

The presented model was developed with standard modules from the Simulink library, which allows it to be used in the implementation on a signal processor boards of the TMS family. The most important blocks are:
- Subsystem "*DWT-IDWT*";
- Subsystem "*Watermarking*";
- Subsystem "*Evaluation*".

In the Subsystem "*DWT-IDWT*" the input image is decomposed into 3 levels via 2D DWT. The transformed by the 2D 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, as is is described in [15].

The Subsystem "*Watermarking*" is presented in Fig.2.

By means of the "*Auto threshold*" and "*Mark correction Block Processing*" blocks, the watermark coefficients are limited to the coefficients of the marker block in order to avoid oversaturation. The "*Resize*" block changes the size of the watermark to be suitable for tagging with the specified block,

then DCT transformation is applied. The resulting values are multiplied by the marker depth coefficient ("*Constant*"). The DCT transformation is applied to the tag block. The two results were combined followed by reverse DCT transformation.
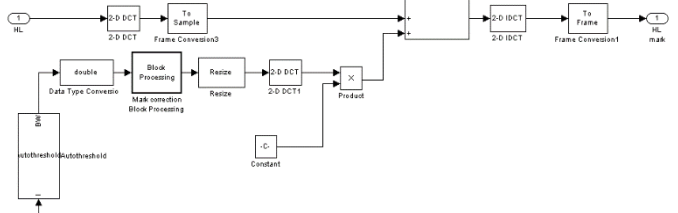
Fig.2. Simulation Subsystem "*Watermarking*"

The efficiency of the developed models for watermarking of medical images is estimated in the Subsystem "*Evaluation*" using the following metrics: 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.

The principal block scheme of simulation model, representing the extraction of the watermark is developed using MATLAB Simulink environment and is presented in Fig.3.
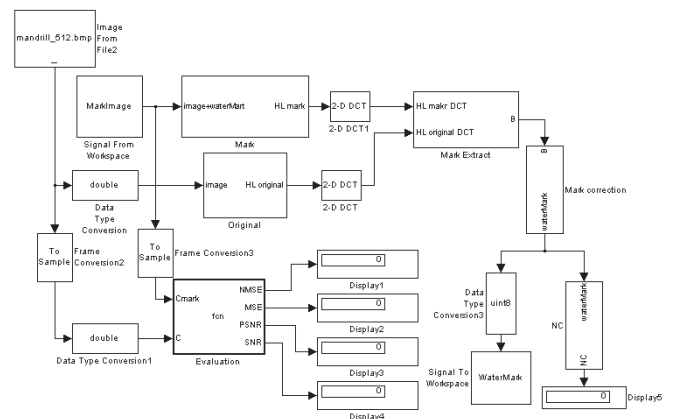
Fig.3. Simulation model, representing the extracting of the watermark

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 watermark is received in the reconstruction unit. Over the data received by the decoder applies inverse discreet wavelet transform. The most important blocks are:
- Subsystem "*Mark*";
- Subsystem "*Original*";
- Subsystem "*Mark Extract*";
- Subsystem "*Evaluation*".

In the Subsystem "*Mark*", analogous to the encoder, a 2D-DWT and extraction of the highlighted block is performed on the tagged images. In the Subsystem "*Original*", a 2D-DWT and extraction of the marker block is performed on a copy of the original image.

The Subsystem "*Mark Extract*" separates the watermark from the tagged fragment of the image using the marker depth coefficient ("*Constant*") and performs 2D-IDCT as is shown in Fig.4.
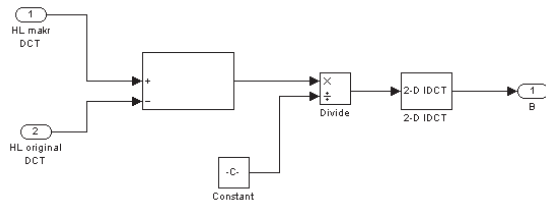
Fig.4. Simulation Subsystem "*Mark Extract*"

In the subblock "*NC*" the normalized cross-correlation is calculated using the formula given in [15]. The block is shown in Fig.5.
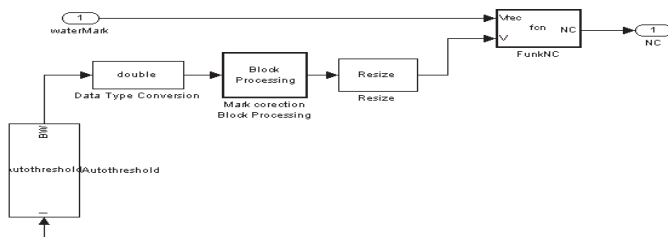
Fig.5. Simulation subblock "*NC*"

III. Experimental Results

For the analyses of efficiency of the developed model for watermarking of medical images three test images, shown in Fig.6a, b, c, with size 512x512 and 256 gray levels are used.

The test images shown in Fig.6 are transformed by the 3 levels DWT and the input watermark (letter K) is embedded into the transformed block (LLH) 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.

To estimating the efficiency of the developed models for watermarking of medical images the following metrics are used: peek signal to noise ratio (PSNR), normalize cross-correlation (NC), mean square error (MSE) and normalized mean square error (NMSE) to determinate how much the watermark image has change compared to the original.

The results obtained from the tests shows that the efficiency of the developed in Simulink model with regard to watermark quality, its invisibility to the user and its resistance to some of the most commonly used attacks is the same as the program implementation developed by the authors and shown in [15]. The differences between the program implementation and the Simulink model appear only in the third decimal place (which represents an average of 0.004%).

Comparing the developed watermarking model with the existing ones in the literature, although the used images are different, the following conclusions can be drawn:

- the embedded watermark is practically invisible for the users and the visual quality of watermarked images is the same in the different publications;
- the coefficients PSNR, NMSE and NC are of the same order of magnitude for the various methods;
- the methods presented in the publications [11],[13] and [14] use frames of video sequences that are of inferior quality compared to those used in the article which demonstrates the advantages of the method described in the work;
- the described hybrid watermarking method (3 stage 2D DWT and 2D DCT) allows better resistance against most commonly used attacks.

Fig.6a. Input X-ray test image "Spine 1".
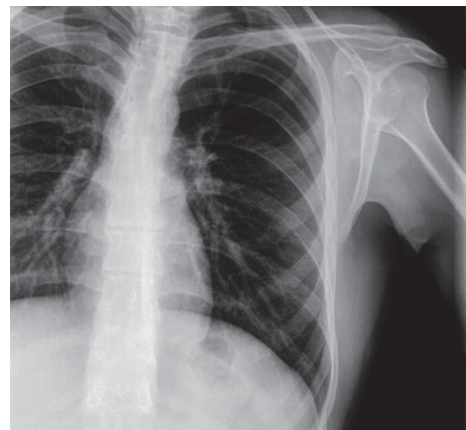
Fig.6b. Input X-ray test image "Spine 2".
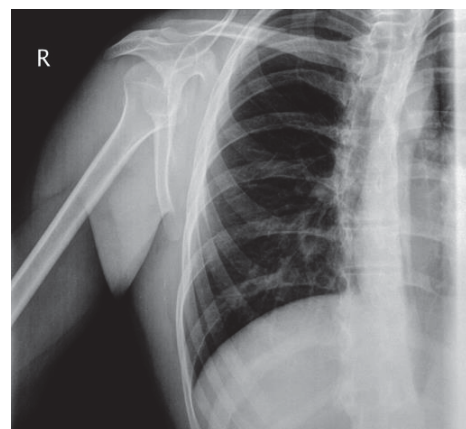
Fig.6c. Input X-ray test image "Spine 3".

The developed in the MATLAB Simulink environment model was tested on a personal computer with 3.2GHz Core-i5 processor and on a specialized signal processor board - TMS 320C6713 DSK.

The results, obtained from the simulation of Matlab, show that time coding for different images varies between 2.7 and 2.79 seconds. The biggest part of it is used for entropy processing and evaluation of the threshold for inserting watermark is about 2 seconds. Simulation on the signal processor TMS 320C6713 takes place over a real time scale, requiring about one second to load each image individually.

IV. Conclusion

A simulation model 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 model 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.

The developed watermarking algorithms and Matlab's simulations are comparable in quality and efficiency to the solutions described in the literature. In addition, experiments were performed with higher quality still images, indicating that the proposed solution could be used for more responsible tasks, for example in medicine, military, agriculture and cultural heritage.

All this leads to the conclusion that the developed models can be used successfully for watermarking not only of medical but on other type of data presented as images.

V. 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] 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.

[9] Neha Singh, Mamta Jain, Sunil Sharma, "A Survey of Digital Watermarking Techniques", *International Journal of Modern Communication Technologies & Research*, August 2013.

[10] 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.

[11] Kavitha KJ, B Priestly Shan. "Video Watermarking Using DCT and DWT: A Comparison". *European Journal of Advances in Engineering and Technology*, 2015, 2(6): 83-87.

[12] Mandeep Singh Saini, Venkata Kranthi B, Gursharanjeet Singh Kalra. "Comparative Analysis of Digital Image Watermarking Techniques in Frequency Domain using MATLAB SIMULINK". *International Journal of Engineering Research and Applications (IJERA)*, Vol. 2, Issue 4, May-Jun 2012, pp.1136-1141. ISSN: 2248-9622.

[13] Prachi V. Powar, S. S. Agrawal. "Design of digital video watermarking scheme using Matlab Simulink". *International Journal of Research in Engineering and Technology (IJRET)*, Vol. 02, Issue: 05, May, 2013, pp. 826-830. ISSN: 2319-1163.

[14] M. Harper, H. Juneja. "Real-Time Digital Video Watermarking". TAMU, *http://www.comm.utoronto.ca/~dkundur/course/real-time-digital-signal-processing/*

[15] R. Mironov, St. Kushlev. "Medical Images Watermarking using Wavelet Transform and DCT". *LIII International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST'2018)*, June 28-30 2018, Sozopol, Bulgaria, Proc. of ICEST'2018, pp.119-122.

Simulation Model Medical Images Watermarking using Complex Hadamard Transform

Rumen P. Mironov[1], Stoyan Kushlev[2]

Abstract – **A simulation model of algorithm for digital watermarking of medical images using complex Hadamard transform is presented. The model is implemented on a signal processor TMS 320C6713 and its experiments indicate it can be used in real-time systems. The developed algorithm for watermarking allow high detection of unauthorized access or attacks on the included watermark. 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, Complex Hadamard Transform, Unitary Transforms, Matlab Simulation.**

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.

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].

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 [8].

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 command 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.

In the class of transparent watermarks they may be further categorize techniques as robust or fragile. A robust mark is designed to resist attacks that attempt to remove or destroy the mark. Such attacks include lossy compression, filtering, and geometric scaling. A fragile mark is designed to detect slight changes to the watermarked image with high probability. The main application of fragile watermarks is in content authentication. Most of the work, as reported in the literature, in watermarking is in the area of robust techniques [4], [8], [9]. Many important applications could benefit from the use of fragile watermarks [10]-[15].

A fragile watermark is a mark that is readily altered or destroyed when the host image is modified through a different image transformation techniques. Fragile marks are not suited for enforcing copyright ownership of digital images - an attacker would attempt to destroy the embedded mark and fragile marks are, by definition, easily destroyed. The sensitivity of fragile marks to modification leads to their use in image authentication [10].

The fragile watermarking techniques in spatial domain mainly include manipulation of the LSB by different methods – vector quantization, chaos theory and etc. [11].

Frequency domain techniques have proved to be more effective than spatial domain techniques in achieving high robustness against attacks and can embed more bits of

---

[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

watermark. Commonly used frequency domain transforms are Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), and Discrete Fourier transform (DFT) and genetic algorithms [12], [13], [14].

The idea to use the Complex Hadamard transformation for watermarking was first presented by Mironov and Kunchev in [16] and is further developed for still images and video sequences in a number of subsequent publications. In the present work a simulation model of developed by the authors' simulation model for digital watermarking of medical images using Complex Hadamard Transform (CHT) is described [15].

The developed model allow high detection of unauthorized access or attacks on the included watermark. The obtained experimental results for some simulated attacks over the 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 with median filters and average filters. The developed on the MATLAB Simulink simulation model is experimented by the personal computer with 3.2GHz Core-i5 processor and specialized signal processor board - TMS 320C6713 DSK.

II. SIMULATION MODEL DESCRIPTION

The common results and properties, obtained from the one dimensional Complex Hadamard Transform (CHT) [16], can be generalized for two-dimensional Complex Hadamard Transform. In this case the 2D signals (images) can be represented by the input matrix [X] with the size NxN. The result is a spatial spectrum matrix [Y] with the same size. The corresponding equations for the forward and the inverse 2D CHT are:

$$\begin{vmatrix} [Y] = [CH_N][X][CH_N] \\ [X] = \dfrac{1}{N^2}[CH_N][Y][CH_N] \end{vmatrix} \quad (1)$$

The Hadamard transformation is simple for implementation, there for it is used for compression and watermarking of information. The proposed complex Hadamard transform matrix has the advantages of having similar structure as the well know real Hadamard matrix. The properties of complex Hadamard transform matrices and its applications in digital image processing are described in detail in article [16].

The algorithms on which are based the developed models for embedding and extraction of watermark are described in article [15]. The principal block scheme of simulation model, representing the embedding of the watermark, developed using MATLAB Simulink environment is presented in Fig.1.

The presented model was developed basically with standard modules from the Simulink library, which allows it to be used in the implementation on a signal processor boards of the TMS family. The difference is only in the blocks where the complex transformations are performed because they do not exist in existing Simulink implementations. The most important blocks are:
- Subsystem "*Embedding*";
- Subsystem "*Marking*";
- Subsystem "*F/M Matrix*";
- Subsystem "*Watermarking*".
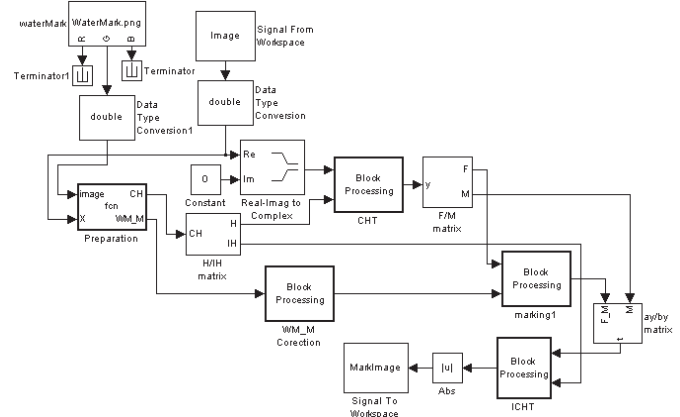
Fig.1. Simulation model, representing the embedding of the watermark using CHT

In the Subsystem "*Embedding*", shown in Fig.2, the input image and the sign image (in subblock "*Preparation*") are preparing for watermarking.
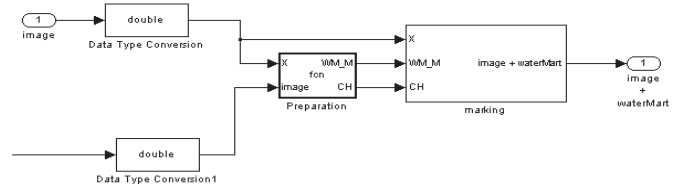
Fig.2. Simulation Subsystem "*Embedding*"

In the Subsystem "*Marking*" show in Fig. 3, the watermarking process is executed.
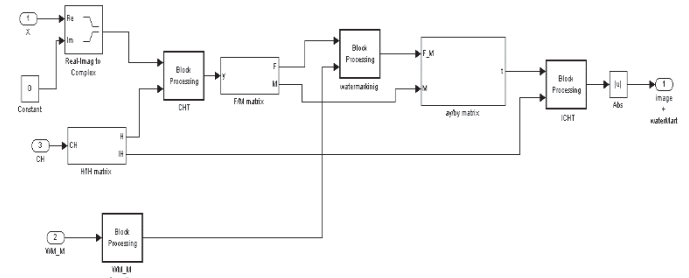
Fig.3. Simulation Subsystem "*Marking*"

The "*H/IH*" subsystem specifies direct Complex Hadamard Matrix and its inverse matrix for the input image. The "*F/M Matrix*" matrix subsystem (shown in Fig.4) defines complex and real parts of amplitude-frequency and phase-frequency components on the base of equations, described in [15].
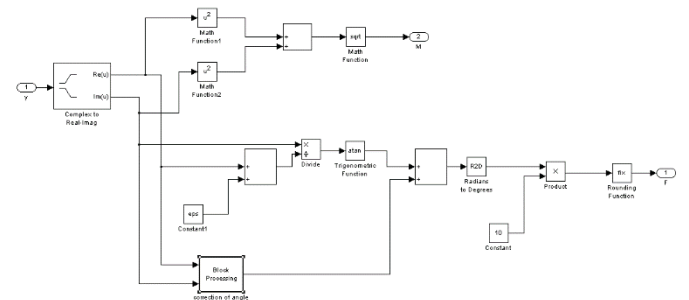
Fig.4. Simulation Subsystem "*F/M Matrix*"

Through the "*WM_M Correction*" block, watermark coefficients are limited to marker block coefficients in order to avoid oversaturation. The "*Watermarking*" subsystem adds the resulting watermark to the converted input image.

In the subblocks "*CHT*" and "*ICHT*", the direct and inverse complex Hadamard matrices are created respectively based on the examinations made in [16].

The principal block scheme of simulation model, representing the extraction of the watermark is developed using MATLAB Simulink environment and is presented in Fig.5.
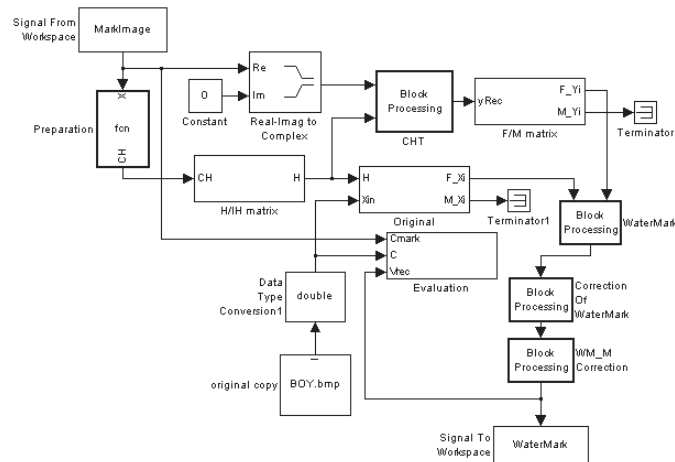
Fig.5. Simulation model representing the extraction of the watermark using ICHT

The marked image is loaded from the workspace using the subblock "*Signal from Workspace*". In the subblock "*Preparation*" the preparation of the complex matrix of Hadamard is carried out. In the Subsystem "*Extracting*", show in Fig. 6, the extraction of watermark is done.
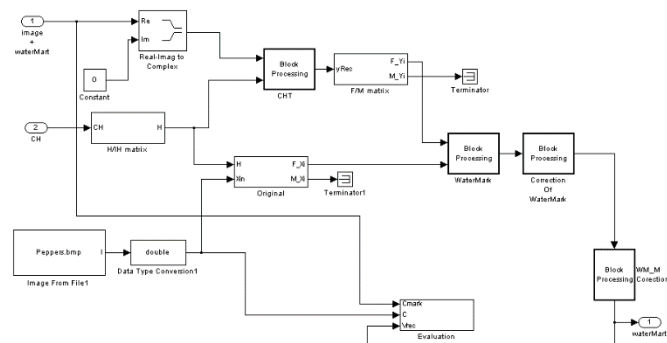
Fig.6. Simulation model of Subsystem "Extracting"

The "*H/IH*" subsystem specifies direct Complex Hadamard Matrix and its inverse matrix for the input image. The "*CHT*" subsystem performs a complete complex transformation of Hadamard onto the tagged image. The "*F/M Matrix*" matrix subsystem (shown in Fig.4) defines complex and real parts of amplitude-frequency and phase-frequency components on the base of equations, described in [15].

The Subsystem "*Original*" performs the functions of "*CHT*" and "*F/M Matrix*" blocks, but applied to a copy of the original image. The "*Water Mark*" subblock performs the extraction of the watermarked image. The "*Correction of Water Mark*" and "*WM_Correction*" subblocks make a retrograde correction of the watermark coefficients that stopped the coder. The

"Evaluation" subsystem calculates the SNR, PSNR, MSE, NMSE and normalized cross-correlation metrics.

III. Experimental Results

For the analyses of efficiency of the developed model for watermarking of medical images three test images, shown in Fig.7a, b, c, with size 512x512 and 256 gray levels are used.

Fig.7a. Input X-ray test image "Spine 1".
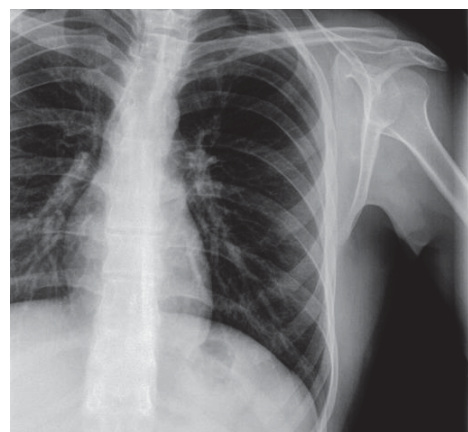
Fig.7b. Input X-ray test image "Spine 2".
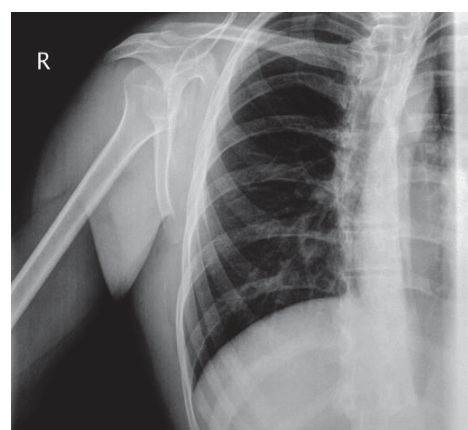
Fig.7c. Input X-ray test image "Spine 3".

These images are transformed by the 2D CHT with kernel 32x32. By this way the input image is divided on 256 sub-images with size 32x32, the input watermark (letter K) is embedded into the phase spectrum of some sub-images.

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.

To estimating the efficiency of the presented simulation model 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.

The results obtained from the tests shows that the efficiency of the developed on Matlab Simulink environment model with regard to watermark quality, its invisibility to the user and its resistance to some of the most commonly used attacks is the same as the program implementation developed by the authors and shown in [15]. There are no differences between the results obtained through program simulation and simulation made with Simulink.

This is due to the fact that in modules using complex Hadamard transform are included software blocks (Math blocks) as there are no standard blocks for this transform in Matlab Simulink Blockset.

The developed in the MATLAB Simulink environment model was tested on a personal computer with 3.2GHz Core-i5 processor and on a specialized signal processor board - TMS 320C6713 DSK.

The results, obtained from the simulation of Matlab, show that time coding for different images varies between 3.1 and 3.36 seconds. The decoding process is about 1.93 seconds.. Simulation on the signal processor TMS 320C6713 takes place over a real time scale, requiring about one second to load and prepare each image individually.

IV. Conclusion

A simulation model for digital watermarking of medical images using complex Hadamard transform is presented. The obtained experimental results for some attacks over the 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 show that the developed model allows high detection of unauthorized access or attacks on the included watermark. 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.

Experimental results obtained from the simulation of signal processor TMS 320C6713 show that the developed model can be implemented and hardware used in systems for the protection of medical imaging systems against unauthorized access in real time.

All this leads to the conclusion that the developed models for watermarking can be used successfully for watermark protection of medical data.

V. 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] Neha Singh, Mamta Jain, Sunil Sharma, "A Survey of Digital Watermarking Techniques", *International Journal of Modern Communication Technologies & Research*, August 2013.

[9] 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.

[10] S. Radharani, M. L. Valarmathi . "A study of watermarking scheme for image authentication". *International Journal of Computer Applications*, 2010, Vol. 2, No.4, pp. 24-32.

[11] Pragya Jain, Anand S. Rajawat. "Fragile Watermarking for Image Authentication: Survey". *International Journal of Electronics and Computer Science Engineering (IJECSE)*, pp. 1232-1237. ISSN 2277-1956.

[12] Huang-Chi Chen, Yu-Wen Chang, Rey-Chue Hwang. "A Watermarking Technique based on the Frequency Domain", *Journal of Multimedia*, 2012, Vol. 7, No. 1, pp. 82-89.

[13] A. Kannammal, K. Pavithra, S. Subha Rani. "Double Watermarking of Dicom Medical Images using Wavelet Decomposition Technique". *European Journal of Scientific Research*, 2012, Vol. 70, No. 1, pp. 46-55.

[14] K. Sreenivas, V. Kamkshi Prasad. "Fragile watermarking schemes for image authentication: a survey". *International Journal of Machine Learning and Cybernetics*, July 2018, Volume 9, Issue 7, pp 1193–1218.

[15] R. Mironov, St. Kushlev. "Medical Images Watermarking using Complex Hadamard Transform". *LII International Scientific Conference Information, Communication and Energy Systems and Technologies (ICEST'2017)*, June 28-30 2017, Niš, Serbia, Proc. of ICEST'2017, pp.52-55. ISSN: 2603-3259 (Print), ISSN: 2603-3267 (Online).

[16] R. Kountchev, R. Mironov. Audio Watermarking in the Phase-Frequency Domain. *XXXIX International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST 2004)*, Bitola, Macedonia, June 16-19, 2004, pp.123-126. ISBN: 9989-786-38-0.

A review of methods for human motion prediction in video sequences with application to 3D telepresence and holoportation systems

Maria Yotova[1], Nikolay Neshov[2], Agata Manolova[3]

Abstract – **Nowadays, the new technologies especially 5G are paving the way of innovative solutions for communication between humans. One very new technology is holoportation, but it also presents a lot of challenges in terms of digital data gathering and transmission. Human motion prediction methods will offer one of the solutions for real time communication in the holoportation context. We present a conceptual model of holoportation architecture for real time communication based on highly accurate 3D modelling of the human face and body, recognition and prediction of human actions and facial expressions to achieve realistic communications. Designed this way the proposed conceptual model of the holoportation system addresses the challenges from the information transmission aspect where real time constraints and narrowband channels are imposed.**

Keywords – **holoportation, immersive telepresence, mixed reality, prediction of human actions.**

I. INTRODUCTION

The way people see the world and interact between themselves is changing; 5G and other new technologies are paving the way of innovative solutions for real time life like long distance communication between humans. Without doubt video communication tools such as Skype and FaceTime are useful for many applications and tasks. However, text, 2D image and voice are not enough of a satisfying alternative to the personal interaction because the real eye contact is missing, touch is not possible between interlocutors and there is a limited feeling of presence (sharing the same space) [1].

These facts bring a sense of incompleteness and dissatisfaction from the communication process to the user. So how the ever evolving technology will help?

Currently in their infancy, augmented reality (AR), mixed reality (MR) and virtual reality (VR) technologies offer great potential to include all five human senses in the communication process to make it more meaningful for all participants. These technologies are already changing the way people interact between each other and with the machines and eventually will radically restructure the work process and the way of living. They will make it easier and more effective for executives, teachers, technologists, people with different professions to communicate in a more natural way anywhere around the globe, with all the advantages of physical presence but without the limitations of the current 2D video communication systems.

One of the new technologies that can enhance the current state of the 2D communication is holoportation. Holoportation is a complete human-computer-machine interface combining AR and VR telepresence thus creating real-life digital scans and realistic 3D avatars of subjects, displayed in MR environment, used for real time communication and interaction between remote users [2]. Holoportation can incorporate all five senses - optic, auditory, olfactory, gustatory, and tactile. This technology opens the door to new heights of interpersonal communication, but it also presents a lot of challenges in terms of digital data gathering and transmission.

In this paper we present a schematic model of a holoportation architecture addressing real time communication needs, based on highly accurate 3D modelling of the human face and body, recognition and prediction of human actions. We will concentrate our efforts to present and review the most current methods for human action prediction – a vital step for the real time holoportation process.

The rest of the paper is organized as follows: the next section presents the current state of the art of immersive telepresence systems. In section III the key features/building blocks of the proposed holoportation system will be described with accent to the human action prediction step. The final section draws the conclusion and suggest the scope of future work.

II. CURRENT TELEPRESENCE SYSTEMS

In the last couple of years many research groups approached in different ways the above noted challenges with system developments for different use case scenarios and contributed to the advancing field of telepresence research. With the increased availability and sophistication of consumer RGB-D sensors and VR/AR glasses, there is currently an exponential emergence of 3D telepresence systems.

One of the most comprehensive reviews and classification of the current state of the immersive telepresence is done in [3]. The author presents an "Immersive Group-to-Group Telepresence" system by implementing a 3D capture and reconstruction pipeline which generates and distributes realistic 3D user representations in real time. One of the functionalities of the systems is the use of two large multi-user projection displays which offer individual stereoscopic perspectives for up to six co-located users. The distributed VR

[1]Maria Yotova, Agata Manolova and Nikolay Neshov are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: E-mail: m.yotova08@gmail.com, nneshov@tu-sofia.bg amanolova@tu-sofia.bg.

framework delivers gaze tracking and eye contact between participants. It suffers from fairly coarse user's representation due to the low resolutions of the utilized RGB-D sensors. The framework was tested over a local area network so no data for more remote connection is yet available.

One of the very recent papers describes a low-cost, low-bandwidth telepresence system capable of rendering people and objects in 3D through data fusion and reconstruction without the use of head-mounted displays or any other wearable devices [4]. The visualization is done on a quadrangular acrylic pyramid by projecting the images using a video projector. This approach albeit requiring low bandwidth suffers from a serious resolution decrease from 1920×1080 to 640×480 with limited size holographic display and no high fidelity audio transmission and synchronization.

SLAMCast [5] is a very promising multi-client real time telepresence framework for remote collaboration which enables efficient remote exploration of quasi-static scenes by multiple independent users while at the same time being able to observe the other user's interactions with the environment. An oriented towards interpersonal communication cost-effective telepresence framework based on 3D data streaming and real time 3D reconstruction is described in [6]. The authors use skeleton data extracted from the depth sensors to animate a 3D human model including rigging the face to express facial expressions. The initial experimental setup show encouraging results to accomplish real time shared person space but no latency information or compensation of data loss is supplied; the sensors used do not allow transferring complex hand movements and multi-user real time telepresence is not assured.

AVATAREX [7] follows the principles of MR and connects users that are simultaneously occupying the same space in the real world and its virtual replica. This framework deals with the idea of how users experience the co-presence in AR and VR. VR users share the same gaming area, but they are physically in another room than the AR users. Even though AVATAREX covers both aspects of shared space it does not currently deal with distance communication so data transfer and latency are not considered in this research.

Many more research papers on the topic of immersive telepresence including different senses can be cited but there is no universal solution. The proposed approaches' success depends on the sensors used, the type of data gathered and way of data processing. The immersive telepresence systems are built on several complex software and hardware technologies, so it is very important to pinpoint what are the current challenges and requirements for real time holoportation.

III. SCHEMATIC MODEL OF THE HOLOPORTATION SYSTEM

A. Online and offline communication

A holoportation system is composed of multiple consecutive steps in order to receive an avatar object.

The first step is the object scanning, where we aim to capture the body of the person in order to estimate its shape, skinning, and pose. There are multiple ways to implement a high quality 3D body scanner, but the most common one is by using high quality cameras in a controlled environment. The scanning process can be performed either online or offline, but in both cases there are shortcomings.

The online scanning system must have a very high performance in order to reconstruct the body. For this scenario a Kinect is used, which captures the raw data, then this data needs to be processed in order to be received correctly.

On the other hand –the offline scanning is performed only once and therefore it is more difficult to animate the avatar in real-time. However, the offline scanning creates more accurate and noiseless avatar, because the cameras used for the scan have very high quality.

B. Data processing

The next logical step is the data processing. The captured images are used to create a 3D model of the human facial characteristics and body with real textures such as skin, facial expressions and clothing. Creating such model based on images is usually done in multiple stages [8]. Capturing and representing accurately the facial characteristics and expressions especially when the parties in the communication process wear large physical devices such as VR/AR glasses, which occlude the majority of the face, is very challenging. So in order to identify the user and capture in detail the facial expressions, the facial data will be gathered and transmitted separately in the communication channel. We must parametrize the facial characteristics of the participants during the teleconference [9]. The facial key points transmitted in real time are used to reanimate the avatar's face.

The second task for the proposed holoportation will be the real time avatar animation visualized at the remote site, based on the metadata captured at the home site. The created 3D model needs to be rigged with the captured skeleton hierarchy created by the Kinect sensors and appropriate texture maps. Skeleton data is transmitted separately in the communication channel. A skeleton based animation strategy is employed for robustly and accurately fitting the avatar to the skeleton and then large scale deformations and movements are applied in real time. Thanks to the multiple Kinect sensors employed, there are no occlusions of joints.

Additionally, we will employ a method for recognizing human activities, to perform short term prediction of the skeleton movements to compensate for network latency.

C. Motion prediction

To provide a real time experience in a two way communication, between the participants, all the data, including body movement, speech and facial expressions, must be transferred without any lag, because it will lead to inconsistency of the speech-movement relation, for example. There is no need to transfer the static objects, because they can be created in both systems – such objects are furniture, walls, etc.

In a short distance, transferring a 3D human model a real time communication isn't an issue, but unfortunately, in a long distance communication there are restrictions due mainly to the speed of the data transfer, because a real-time transfer on a LAN network reaches a speed between 1-2Gbps, and to achieve a real time communication, we need to transfer hundreds of megabits. To avoid the network delay, a human motion prediction algorithm can be applied. This way, by scanning the avatar, we can send a metadata, which will be received by the recipient, containing the predicted position of the sender. Then the skeleton, presented by a number of joints that represent key body parts, animates the avatar, avoiding the long-distance data-transfer delay.

In this section we will make an overview of the recent methods that are used to make a short-term human motion prediction. To make a motion prediction in a video sequence there are several steps to be completed. First must be performed an action detection, then this action must be tracked and after that, we may say that this action is recognized. After the recognition of the action, comes the classification and then it is possible to make a short term prediction on what the next body position will be. Traditional approaches use Markovian assumptions [10, 11], but the latest works on this problem are based on different methods which use convolutional neural networks(CNN) or recurrent neural networks, or more specifically on Long short term memory (LSTM) and Gated Recurrent Unit (GRU). The LSTMs are widely used in action and speech recognition [12] thanks to their ability to learn long-term feature relationships by processing overlapping sequences of consecutive frames.

To solve the motion prediction task, Fragkiadaki et al. [13]. prose two architectures: LSTM-3LR (3 layers of Long Short-Term Memory cells) and ERD (Encoder-Recurrent-Decoder). Both are based on concatenated LSTM units, but the latter adds non-linear space encoders for data pre-processing. The authors also note that, during inference, the network is prone to accumulate errors, and quickly produces unrealistic human motion. Therefore, they propose to gradually add noise to the input during training which forces the network to be more robust to prediction errors. This noise scheduling makes the network able to generate plausible motion for longer time horizons, especially on cyclic walking sequences. Jabri et al. [14] have shown competitive performance on VQA with a simple baseline that does not take images into account, and state-of-the-art performance with a baseline that is trained to exploit the correlations between questions, images and answers.

In their paper Martinez et al. [15] propose a different approach than the LSTM networks. Their work is based on the GRU [16], with the help of which, they manage to drop the spatial encoding layer. This allows them to train their model on 3.6M Dataset for a few hours, and there is no need to train a different model for each different action. They analyse the motion continuity as velocity, rather than a set of poses, which allows them to model only one velocity, as opposed to presenting all possible human poses. To achieve this, they base their work on Sequence-to-Sequence architecture [13] with the help of which during training, the ground truth is fed to an encoder network, and the error is computed on a decoder

network that feeds its own predictions. The decoder also has a residual connection, which effectively forces the RNN to internally model angle velocities.

Pavllo et al. [17] propose similar approach as Martinez et al. for short term predictions using a quaternion network, we consider predicting either relative rotation deltas (analogous to angular velocities) or absolute rotations. We take inspiration from residual connections applied to Euler angles, where the model does not predict absolute angles but angle deltas and integrates them over time. For quaternions, the predicted deltas are applied to the input quaternions through quaternion product. First they create a recurrent architecture, where they also use a GRU, because of its simplicity. The idea is that at each step, the model receives the previous state and features, in order to make a next pose estimation. After further work, they create a convolutional based architecture, where they replace the GRU and the linear layer with convolutional layers.

In Table 1 are summarized the most recent papers on motion prediction on the Human3.6M dataset (http://vision.imar.ro/human3.6m/description.php).

IV. CONCLUSION

The main goal of this paper is to provide a comprehensive survey of the most recently published motion prediction methods that can be employed in immersive telepresence systems. And also to present a schematic model of a holoportation system and to outline some of its key concepts. The benefit of the proposed holoportation system is that it allows users to interact practically in real time - anywhere, anytime, with anybody – either in a virtual or integrated way offering the feeling of personal interactivity and the feeling of shared space. Holoportation offers the opportunity to address the current limitations in 2D communication permitting immediate, 3-D visual, auditory, tactile and emotional interaction between remote users. But a lot of work is still needed to make this system an everyday reality for the society.

ACKNOWLEDGEMENT

REFERENCES

[1] Cohen, Aviva, et al. "Sustaining a caring relationship at a distance: Can haptics and 3D technologies overcome the deficits in 2D direct synchronous video based communication?." Virtual System & Multimedia (VSMM), 23rd International Conference on. IEEE, 2017.

[2] Orts-Escolano, Sergio, et al. "Holoportation: Virtual 3d teleportation in real-time." Proceedings of the 29th Annual

Symposium on User Interface Software and Technology. ACM, 2016.

[3] Beck, Stephan. "Immersive Telepresence Systems and Technologies." (2019)., PhD thesis Fakultät Medien der Bauhaus-Universität Weimar.

[4] Córdova-Esparza, D.-M., Terven, J. R., Jiménez-Hernández, H., Herrera-Navarro, A., Vázquez-Cervantes, A., & García-Huerta, J.-M. (2019). Low-bandwidth 3D visual telepresence system. Multimedia Tools and Applications. doi:10.1007/s11042-019-7464-0.

[5] Stotko, P., Krumpen, S., Hullin, M. B., Weinmann, M., & Klein, R. (2019). SLAMCast: Large-Scale, Real-Time 3D Reconstruction and Streaming for Immersive Multi-Client Live Telepresence. IEEE transactions on visualization and computer graphics.

[6] Huynh, D., King, S. A., & Katangur, A. K. (2018). A framework for cost-effective communication system for 3D data streaming and real-time 3D reconstruction. Proceedings of the 3rd International Workshop on Interactive and Spatial Computing - IWISC '18. doi:10.1145/3191801.3191804

[7] Koskela, T., Mazouzi, M., Alavesa, P., Pakanen, M., Minyaev, I., Paavola, E., & Tuliniemi, J. (2018, February). AVATAREX: Telexistence System based on Virtual Avatars. In Proceedings of the 9th Augmented Human International Conference (p. 13). ACM.

[8] Zanfir, Mihai, Alin-Ionut Popa, Andrei Zanfir, and Cristian Sminchisescu. "Human appearance transfer." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5391-5399. 2018.

[9] Manolova, Agata, Nikolay Neshov, Stanislav Panev, and Krasimir Tonchev. "Facial expression classification using Supervised Descent Method combined with PCA and SVM." In International Workshop on Biometric Authentication, pp. 165-175. Springer, Cham, 2014.

[10] McGhan, Catharine LR, Ali Nasir, and Ella M. Atkins. "Human intent prediction using markov decision processes." Journal of Aerospace Information Systems 12, no. 5 (2015): 393-397.

[11] Lehrmann, Andreas M., Peter V. Gehler, and Sebastian Nowozin. "A non-parametric bayesian network prior of human pose." In Proceedings of the IEEE International Conference on Computer Vision, pp. 1281-1288. 2013.

[12] Liu, Jun, Amir Shahroudy, Dong Xu, and Gang Wang. "Spatio-temporal lstm with trust gates for 3d human action recognition." In European Conference on Computer Vision, pp. 816-833. Springer, Cham, 2016.

[13] Fragkiadaki, Katerina, Sergey Levine, Panna Felsen, and Jitendra Malik. "Recurrent network models for human dynamics." In Proceedings of the IEEE International Conference on Computer Vision, pp. 4346-4354. 2015.

[14] Jabri, Allan, Armand Joulin, and Laurens Van Der Maaten. "Revisiting visual question answering baselines." In European conference on computer vision, pp. 727-739. Springer, Cham, 2016.

[15] Martinez, Julieta, Michael J. Black, and Javier Romero. "On human motion prediction using recurrent neural networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2891-2900. 2017.

[16] Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014).

[17] Pavllo, Dario, Christoph Feichtenhofer, Michael Auli, and David Grangier. "Modeling Human Motion with Quaternion-based Neural Networks." arXiv preprint arXiv:1901.07677 (2019).

Modeling Human Motion																
Action	Walking				Eating				Smoking				Discussing				Base architecture and NN
milliseconds	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	
Zero-velocity (Martinez et al., CVPR 2017)	0.39	0.68	0.99	1.15	0.27	0.48	0.73	0.86	0.26	0.48	0.97	0.95	0.31	0.67	0.94	1.04	Sequence-to-sequence architecture, RNN, GRU
ERD (Fragkiadaki et al., CVPR 2015)	0.93	1.18	1.59	1.78	1.27	1.45	1.66	1.80	1.66	1.95	2.35	2.42	2.27	2.47	2.68	2.76	ERD architecture, RNN, LSTM (with non-linear space encoders)
LSTM-3LR (Fragkiadaki et al., CVPR 2015)	0.77	1.00	1.29	1.47	0.89	1.09	1.35	1.46	1.34	1.65	2.04	2.16	1.88	2.12	2.25	2.23	RNN, LSTM
QuaterNet abs. (Pavllo et al., BMVC 2018b)	0.26	0.42	0.67	0.70	0.23	0.38	0.61	0.73	0.32	0.52	0.92	0.90	0.36	0.71	0.96	1.03	Quaternion-based neural architectures, RNN, GRU,CNN
QuaterNet vel. (Pavllo et al., BMVC 2018b)	0.21	**0.34**	0.56	**0.62**	0.20	0.35	0.58	0.70	0.25	0.47	0.93	0.90	0.26	0.60	0.85	0.93	Quaternion-based neural architectures,RNN, GRU,CNN
QuaterNet vel. TF	**0.20**	0.37	0.64	0.76	0.19	0.34	0.61	0.78	**0.24**	0.48	0.90	0.99	0.25	0.64	0.97	1.07	Quaternion-based neural architectures, RNN, GRU,CNN

Table 1: Results under the standard protocol (Fragkiadaki et al., 2015), with 4 samples per sequence. The mean angle error for short-term motion prediction on Human 3.6M for different actions is presented. Bold indicates the best result.

Speech Intelligibility in Serbian Orthodox Churches and IEC 60268 – 16 Standard

Violeta Stojanović[1], Zoran Milivojević[2] and Momir Praščević[3]

Abstract – **This paper presents an analysis of the speech intelligibility in Serbian Orthodox churches "St. George" in Žitni Potok and "St. Prokopije" in Katun, which was based on the measured values of the objective acoustic STI parameter and IEC 60268-16 Standard. The first part of the paper includes tables and graphs presenting results of the calculated values of STI parameter (obtained on the basis of practical measurements of acoustic impulse responses in the churches) and the corresponding percentages values of the intelligibility of sentences, PB words and CVC logatoms set by the Standard. The second part of the paper includes correlation and regression analyses of the results, as well as the conclusion.**

Keywords – **Objective acoustic parameter STI, IEC 60268-16 Standard, Correlation and regression.**

I. INTRODUCTION

Speech is the main means of communication among people. In many situations, the voice signal degrades while transmitting the channel between the speaker and the listener. This results in a decreased speech intelligibility at the location of the listener.

One of the methods of assessing the intelligibility of the listener's speech is the method of measuring the objective Speech Transmission Index *STI*. This method has been the subject of continuous development and refinement since it was introduced in the 1970s, as confirmed by many revisions of IEC 60268-16 Standard [1]. Houtgast and Steeneken described in 1973. an objective method for estimating the speech intelligibility in rooms by calculating a physical index from the modulation transfer function (*MTF*) [2]. This physical index, called the Speech Transmission Index (*STI*) is calculated at discrete frequencies, weighted, summed and normalized to yield a single index of speech intelligibility. Houtgast and Steeneken modified the original *STI* model in 2002. [3]. They established a relationship between the revised Speech Transfer Index (*STIr*) and subjective speech intelligibility. The revised Speech Transmission Index (*STIr*) is obtained by a weighted summation of the modulation transfer indices for all octave bands and the corresponding corrections of excessive repetition. The redundancy correction is related to the contribution of adjacent frequency bands [3].

The relationship between the objective acoustic parameter *STI* and the speech intelligibility (expressed as percentage) is shown in Table I [1]. This correlation was made using intelligibility of [2]: a) syllables, b) words from the so-called "Harvard List" and c) sentences, using the Speech Reception Threshold (SRT) method. Fig. 1. shows the relationship between the revised Speech Transmission Index (*STIr*) and subjective intelligibility: a) sentences, b) PB words, and c) CVC logatoms [1].

TABLE I
RELATIONSHIP BETWEEN STI, QUALITY SPEECH INTELLIGIBILITY TO IEC 60268-16, AND THE SPEECH INTELLIGIBILITY SIABLES (SI_S), WORDS (SI_W) AND SENTENCES (SI_{SENT}).

STI	Quality according to: IEC 60268-16	SI_s (%)	SI_w (%)	SI_{sent} (%)
0 ÷ 0.3	bad	0 ÷ 34	0 ÷ 67	0 ÷ 89
0.3 ÷ 0.45	poor	34 ÷ 48	67 ÷ 78	89 ÷ 92
0.45 ÷ 0.6	fair	48 ÷ 67	78 ÷ 87	92 ÷ 95
0.6 ÷ 0.75	good	67 ÷ 90	87 ÷ 94	95 ÷ 96
0.75 ÷ 1	excellent	90 ÷ 96	94 ÷ 96	96 ÷ 100
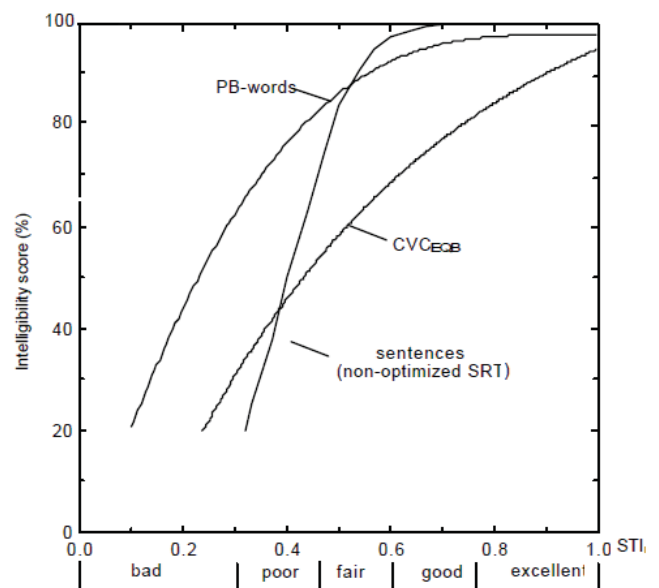
Fig. 1. Classification of the relationship of speech intelligibility and *STI*.

In this paper, we analyzed the speech intelligibility for two Serbian Orthodox churches and this was based on measured values of objective acoustic parameters *STI* at central frequencies f_c = {500 Hz, 2 kHz} and the quality of speech intelligibility according to IEC 60268-16 Standard. The values of objective acoustic parameters *STI* were determined

[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.

according to the measurement of acoustic impulse response and EASERA software package. The classification of speech intelligibility in acoustically examined churches was determined by the relationship between the value of *STI* parameters and the percentage of speech intelligibility of sentences SI_{sent}, PB words SI_{PBw} and CVC logatoms SI_{CVC}, set by the Standard. The relationship between these quantities was determined by correlation and regression analyses, using Matlab software package.

The paper is organized as follows: Section II explains the experiment and presents the experimental results; Section III shows the analysis of the results and Section IV is the conclusion.

II. EXPERIMENT

The process of measuring impulse response was carried out in two Serbian Orthodox churches: 1) the church "St. George" in Žitni potok and 2) the church "St Prokopije" in Katun. Analysis of acoustic objective parameters in these churches is explained in detail in [4].

The church "St. George" in Žitni Potok (church 1) has a volume of $V = 2163$ m^3 and inner area is $S = 167$ m^2. The church "St Prokopije" in Katun (church 2) has a volume of V $= 1659.68$ m^3 and inner area is S $= 646.68$ m^2. The interior walls and the ceilings in the churches covered with plaster (the coefficient of absorption $\alpha = 0.02$). The floors with the ceramic tiles (the coefficient of absorption $\alpha = 0.015$).

For the purpose of the analysis in this paper, we took the acoustic impulse responses measured at MP = 9 measuring points in both churches. Their arrangement is shown in Fig. 2. and Fig. 3. The database includes 63 wav files that were obtained by recording the acoustic impulse response using the software package EASERA. For each measuring point, 7 measurements were made. Sound source LS is placed near the altar.

The equippment used for the experiment as follows: (a) an omnidirectional microphone (PCB 130D20), having a diaphragm diameter of 7mm; (b) a B&K omnidirectional sound source type 4295 (dodecahedron loudspeaker); (c) a B&K audio power amplifier, rated at 100W RMS, stereo, type 2716-C; (d) a laptop, incorporating a Soundmax Integrated Digital Audio sound card from Analog Devices. Measuring of the impulse response is carried out using incentive log sweep signal with the duration of 5 s sampling frequency is fS = 44.1 kHz. The procedure for recording and calculation of acoustic parameters was performed in accordance with ISO 3382.

The mean values of *STI* parameters at all measured MP positions in the churches, at $f_c = \{500$ Hz, 2 kHz$\}$, were calculated using the EASERA software package. Based on the classification of the speech intelligibility and the *STI* parameter set by the Standard (Fig. 1.), the values of the speech intelligibility were determined as follows: sentences SI_{sent}, PB words SI_{PBw} and CVC logatoms SI_{CVC} for each MP position in the churches. Next, mean values $\overline{SI_{sent}}$, $\overline{SI_{PBw}}$ and $\overline{SI_{CVC}}$ were calculated for all measured MP positions in
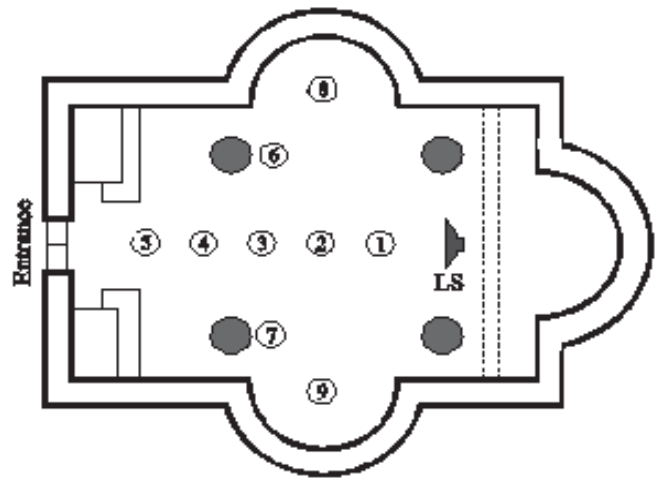
Fig. 2. The position of the measuring points MP and sound source LS in the church "St. George" in Zitni Potok.
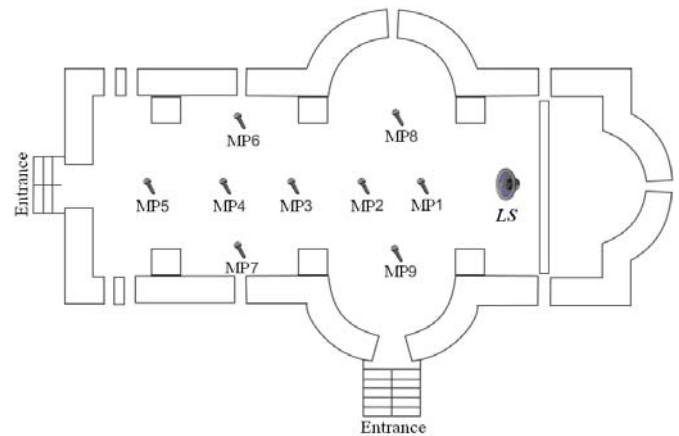
Fig. 3. The position of the measuring points MP and sound source LS in the church "St Prokopije" in Katun.

church 1 and church 2. The assessment of the quality of speech in the churches was performed according to the Standard (Table I). The strength of the relationship between the objective *STI* parameter and the qualifying values of speech intelligibility qualities: SI_{sent}, SI_{PBw} and SI_{CVC} for both churches individually, at $f_c = \{500$ Hz, 2 kHz$\}$, was determined by regression and correlation analyses.

A. Resultats of the experiment

Tables II and III show the values of objective *STI* parameters and the values of subjective speech intelligibility SI_{sent}, SI_{PBw} and SI_{CVC} at the central frequencies $f_c = \{500$ Hz, 2 kHz$\}$ for church 1 and church 2, respectively. Table IV shows the mean values of these measures for both churches. Figs. 4. -6. show regression lines for measures *STI* and SI_{sent}, *STI* and SI_{PBw} and *STI* and SI_{CVC} at central frequencies $f_c = \{500$ Hz, 2 kHz$\}$ for church 1. Figs. 7. – 9. show lines regression of these measures, at the same frequencies, for church 2.

TABLE II

THE VALUES *STI*, *SI_sent*, *SI_PBw* AND *SI_CVC* AT 500 Hz AMD 2kHz FOR THE CHURCH 1

	f = 500 Hz								
	MP								
	1	2	3	4	5	6	7	8	9
STI	0.66	0.53	0.51	0.4	0.38	0.37	0.34	0.37	0.45
SI_sent (%)	99	90	62	50	40	36.5	27	36.5	67
SI_PBw (%)	95.2	89	81	76	73	72.5	68	72.5	82
SI_CVC (%)	74.5	62	52.5	46	43	41	37	41	52.5
	f = 2kHz								
STI	0.77	0.62	0.5	0.46	0.4	0.37	0.35	0.39	0.58
SI_sent (%)	100	98	84	70	50	36.5	31	44	96
SI_PBw (%)	97	93.5	87	82	76	72.5	70	74.5	92
SI_CVC (%)	82	70.1	58	54	46	41	38.5	43	68.5

TABLE III

THE VALUES *STI*, *SI_sent*, *SI_PBw* AND *SI_CVC* AT 500 Hz AND 2kHz FOR THE CHURCH 2

	f = 500 Hz								
	MP								
	1	2	3	4	5	6	7	8	9
STI	0.57	0.49	0.44	0.38	0.37	0.42	0.43	0.44	0.48
SI_sent (%)	95	81	64	40	36.5	57	61	64	77
SI_PBw (%)	91	85.05	81	73	72.5	78.5	80	81	84
SI_CVC (%)	66.5	58	51.5	43	41	48.5	50	51.5	56.5
	f = 2kHz								
STI	0.69	0.55	0.54	0.52	0.5	0.57	0.57	0.58	0.56
SI_sent (%)	100	97	96	87.5	90	95	95	96	94
SI_PBw (%)	96	90	89.5	87.5	89	91	91	92	90.1
SI_CVC (%)	74.5	64	62.5	61	62	66	66	67	65

TABLE IV

THE MEAN VALUES $\overline{STI}$, $\overline{SI_{sent}}$, $\overline{SI_{PBw}}$ AND $\overline{SI_{CVC}}$ AT 500 Hz AND 2kHz FOR BOTH CHURCHES

	Church 1		Church 2	
f (Hz)	500	2000	500	2000
$\overline{STI}$	0.44 poor	0.45 poor-fair	0.45 poor-fair	0.57 fair
$\overline{SI_{sent}}$ (%)	64 bad	70 bad	66.5 bad	95 fair-good
$\overline{SI_{PBw}}$ (%)	81 fair	82 fair	81.5 fair	91 good
$\overline{SI_{CVC}}$ (%)	51.5 fair	54 fair	52.5 fair	66 fair
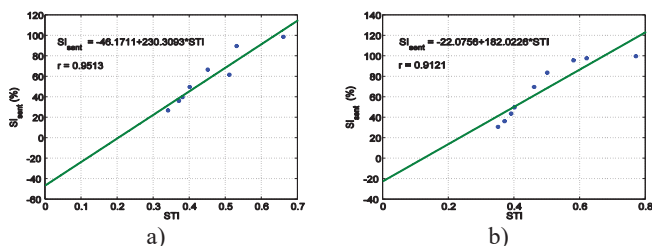
Fig. 4. The regression line for *SI_sent* and *STI* at: a) 500 Hz and b) 2kHz for the Church 1.
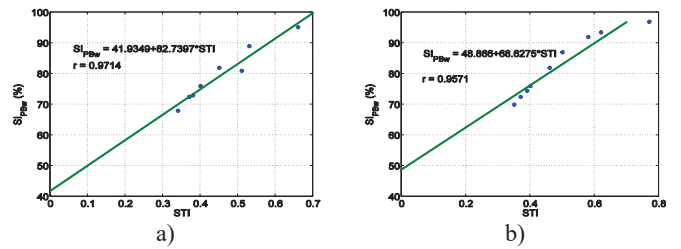
Fig. 5. The regression line for *SI_PBw* and *STI* at: a) 500 Hz and b) 2kHz for the Church 1.
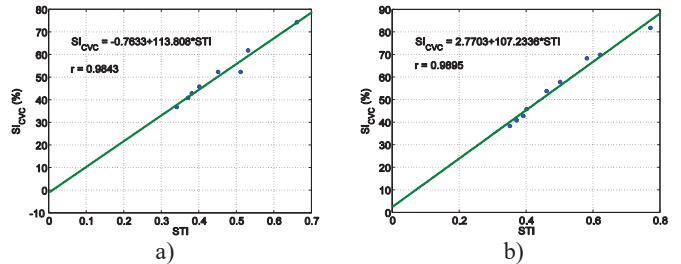
Fig. 6. The regression line for *SI_CVC* and *STI* at: a) 500 Hz and b) 2kHz for the Church 1.
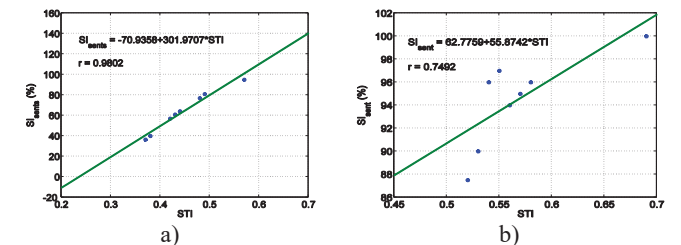
Fig. 7. The regression line for *SI_sent* and *STI* at: a) 500 Hz and b) 2kHz for the Church 2.
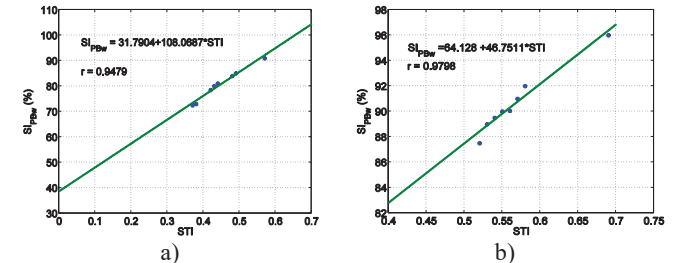
Fig. 8. The regression line for *SI_PBw* and *STI* at: a) 500 Hz and b) 2kHz for the church 2.
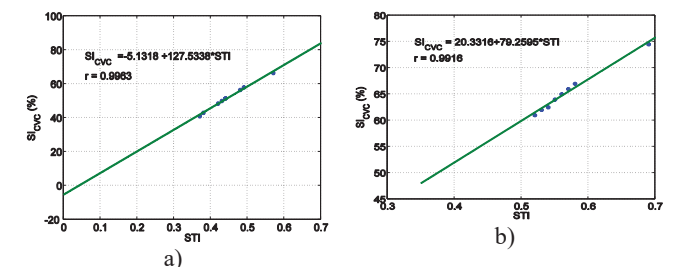
Fig. 9. The regression line for *SI_CVC* and *STI* at: a) 500 Hz and b) 2kHz for the church 2.

B. The results analysis

Based on the results shown in the Tables II - IV and the Figs 4. – 9. the following can be concluded:

1) for church 1: a) At f = 500 Hz mean value of the objective parameter $\overline{STI}$ = 0.44 classifies poor intelligibility of speech. According to the criteria for the subjective intelligibility of speech set by the Standard, the mean $\overline{SI_{sent}}$ = 64% determines poor intelligibility of sentences while the values $\overline{SI_{PBw}}$ = 81% and $\overline{SI_{CVC}}$ = 51.5% determine fair intelligibility of PB words and CVC logatoms. b) At f = 2 kHz is $\overline{STI}$ = 0.45 and determines the intelligibility of speech that is at the border line between poor and fair. The intelligibility of sentences is poor, and the intelligibility of PB words and CVC logatoms is fair ($\overline{SI_{sent}}$ = 70%, $\overline{SI_{PBw}}$ = 82% i $\overline{SI_{CVC}}$ = 54%).

2) for church 2: a) At f = 500 Hz speech comprehension is at the border line between poor and faire because the mean value of the objective parameter is $\overline{STI}$ = 0.45. And here the poor intelligibility have sentences, $\overline{SI_{sent}}$ = 64% and faire intelligibility of PB words and CVC logatoms, $\overline{SI_{PBw}}$ = 81.5% and $\overline{SI_{CVC}}$ = 52.5%. b) Compared to previous results concerning the speech intelligibility, church 2 at 2 kHz shows the best rating of speech intelligibility- faire ($\overline{STI}$ = 0.57). The intelligibility of sentences is at the border line between faire - good, $\overline{SI_{sent}}$ = 95%, because $\overline{SI_{sent}}$ = 64% The intelligibility of PB words is good ($\overline{SI_{PBw}}$ = 91.%), whereas the intelligibility of CVC logatoms is acceptable ($\overline{SI_{CVC}}$ = 66%).

3) for church 1, the correlation and regression analysis gave the following connection of the investigated quantities with the corresponding Pearson coefficients at a) f = 500 Hz and b) f = 2 kHz:

a) SI_{sent} = -46.1711 + 230.3093 STI, r = 0.9513;
$SI_{\check{S}Bw}$ = 41.9349 + 82.7397 STI, r = 0.9714;
SI_{CVC} = -0.7633 + 113.808 STI, r = 0.9843;

b) SI_{sent} = -22.0756 + 182.0226 STI;
$SI_{\check{S}Bw}$ = 48.666 + 68.6275 STI, r = 0.9571;
SI_{CVC} = 2.7703 + 107.2336 STI, r = 0.9895.

This analysis shows that there is a statistically positive, strong, linear connection between the objective parameter STI and the subjective intelligibility of speech SI_{sent}, SI_{PBw} i SI_{CVC}

4) for church 2, the correlation and regression analysis gave the following connection of the investigated quantities with the corresponding Pearson coefficients at a) f = 500 Hz and b) f = 2 kHz:

a) SI_{sent} = -70.9358 + 301.9707 STI, r = 0.9802;
$SI_{\check{S}Bw}$ = 31.7904 + 108.0687 STI, r = 0.9479;
SI_{CVC} = -5.1318 + 127.5338 STI, r = 0.9963;

b) SI_{sent} = 62.7759 + 55.8742 STI;
$SI_{\check{S}Bw}$ = 64.128+ 46.7511 STI, r = 0.9798;
SI_{CVC} = 20.3316 + 79.2595 STI, r = 0.9916.

There is also a statistically positive, strong, linear connection between the objective STI parameter and the subjective speech intelligibility SI_{sent}, $SI_{\check{S}Bw}$ and SI_{CVC}.

III. CONCLUSION

In this paper, for two Serbian Orthodox churches: 1) the church "Sveti Đorđe" in Žitni potok and 2) the church "Sveti Prokopije" in Katun, we analyzed the intelligibility of speech sentences SI_{sent}, PB words SI_{PBw} and CVC logatoms SI_{CVC}, based on the calculated values of STI acoustic parameters for central frequencies f_c = {500 Hz, 2 kHz}, and the quality of comprehensibility of speech according to the Standard IEC 60268 - 16.The relationship between these acoustic parameters was analyzed using regression and correlation parameters.

The mean values of the objective acoustic STI parameter for both churches classify poor and fair of speech intelligibility: $\overline{STI}$ = {0.44, 0.45} i $\overline{STI}$ = {0.45, 0.57} for churches 1 and 2, respectively. Such quality of speech intelligibility results in: 1) poor intelligibility of sentences in church 1: $\overline{SI_{sent}}$ = {64%, 70%} and in church 2 at f_c = 500 Hz, $\overline{SI_{sent}}$ = 66.5%, whereas the intelligibility is faire - good in church 2 at f_c = 2 kHz, $\overline{SI_{sent}}$ = 95%; 2) faire intelligibility of PB words in church 1: $\overline{SI_{PBw}}$ = {81.5%, 82%} as well as in church 2 at f_c = 500 Hz, $\overline{SI_{PBw}}$ = 81.5%, and good intelligibility in church 2 at f_c = 2 kHz, $\overline{SI_{PBw}}$ = 91% and 3) faire intelligibility of CVC logatoms in both churches: $\overline{SI_{CVC}}$ = {51.5%, 54%} and $\overline{SI_{CVC}}$ = {52.5%, 66%}.The correlation and regression analyses confirmed a statistically positive, strong, linear connection (r $\geq$ 0.7) between the acoustic objective Speech Transmission Index STI and the subjective intelligibility of speech analyzed according to the comprehensibility of sentences, PB words and CVC logatoms set by the Standard.

REFERENCES

[1] International Electrotechnical Commission IEC 60268-16 – International Standard, "Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index", Switzerland: IEC, 2011.

[2] Houtgast, T. and Steeneken, H.J.M., "The Modulation Transfer Function in room acoustics as a predictor of speech intelligibility", Acustica, 28, 66–73 (1973).

[3] Steeneken, H.J.M. and Houtgast, T., "Validation of the STIr method with the revised model", Speech Communication 38, 2002, pp. 413 – 425.

[4] V.Stojanović, Z. Milivojević, "Analysis of Acoustic Parameters in Serbian Orthodox Churches", 26th International Conference Noise and Vibration, December 6 – 7, 2018., Niš, Serbia, pp. 93 – 98.

Signal Analysis with Application of k-Nearest Neighbors Method

Ivelina Balabanova[1], Georgi Georgiev[2] and Stela Kostadinova[3]

Abstract – **The paper presents the results of synthesizing k-Nearest Neighbors (k-NN) models for identification the presence of noises. Signals with Gaussian White and Periodic Random noises are analyzed. The models are evaluated by resubstitution and cross-validation procedures. Classifiers at Euclidean, Minkowski, Cityblock and Chebychev metric distances are examined. Classification model with 97.375% accuracy, parameter k = 5 and Minkowski distance is selected.**

Keywords – **noise identification, k-NN models, metric distance, resubstitution, cross-validation.**

I. Introduction

The wide use of the k-Nearest Neighbors (k-NN) method in the processing of biomedical signals is confirmed by a number of studies.

The method is particularly useful in assessing the quality levels of ECG data from wireless sensor systems for continuous monitoring the conditions of patients, where the possibility of interference from the human body movement is especially large [1].

The k-NN is used for diagnosing the condition of the human cognitive system, for detecting disturbances in the normal brain function as well as exploration of the human motor system on the basis of EEG signals [2-4].

Voice recognition in background noise environment or speech segmentation for identification of the minimal degraded speech fragments in the transmission through a co-channel interface, are investigated in some studies [5, 6].

There is interest in detecting anomalies in the echo distribution of meteorological radar data, as well as the classification of objects based on 2D scans of separate radar routes [7-9].

In the report applicability and method efficiency are studied for identification of electrical signals with Gaussian White Noise (GWN) and Periodic Random Noise (PRN).

II. Experimental data

Sine, square and triangle waveforms with GWN and PRN are simulated in LabVIEW environment. The simulation is conducted at an amplitude level "one unit" and a frequency of the signals "10.1 Hz", as well as specified noise parameters -

"0.05" for Standard deviation and Spectral amplitude, respectively for GWN and PRN.

The taken oscillograms are shown on Fig. 1 and Fig. 2.
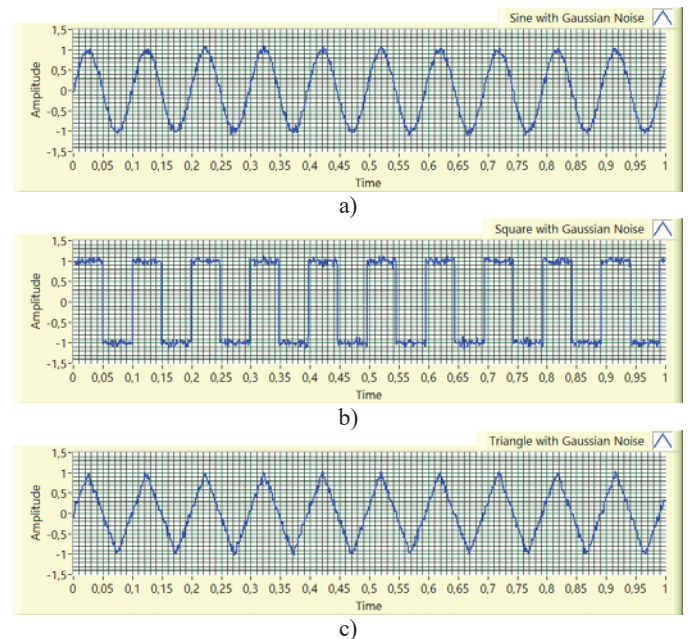
Fig. 1. Sine a), square b) and triangle c) waveforms with GWN

Fig. 2. Sine a), square b) and triangle c) waveforms with PRN

[1]Ivelina Balabanova is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadji Dimitar Str., Gabrovo 5300, Bulgaria, E-mail: ivstoeva@abv.bg.

[2]Georgi Georgiev is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadji Dimitar Str., Gabrovo 5300, Bulgaria.

[3]Stela Kostadinova is with the Faculty of Computer Sciences and Automation at Technical University of Varna, 1 Studentska Str., Varna 9000, Bulgaria, E-mail: stela.kostadinova@tu-varna.bg.

Fig. 3. Sine a), square b) and triangle c) waveforms with PRN

Sample from 2000 information samples for GWN (Class №1) and PRN (Class №2) signals defined as target identification groups is acquired based on data from experimental signals. Graphical interpretation of the investigated signals as points in a three-dimensional classification space is given in Fig. 3.

III. SYNTHESIS OF THE MODELS FOR NOISE IDENTIFICATION BY к-NN METHOD

Classification models based on the k-Nearest Neighbors method are created for the following metric distances:

- Euclidean (№1);
- Minkowski (№2);
- Cityblock (№3);
- Chebychev (№4).

The study is performed with a successive increase of k-neighbors (points in the classification space) in the range of 5 to 100 neighbors. Key indicators defining the quality of classifiers are the errors in the technical approaches "resubstitution" and "cross-validation".
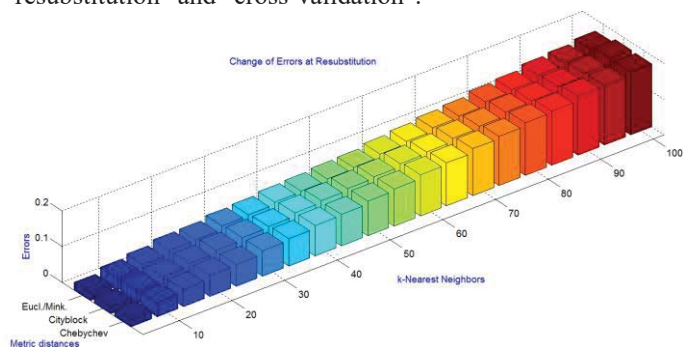
Fig. 4. Resubstitution errors at a) Euclidean/Minkowski, b) Cityblock and c) Chebychev distances
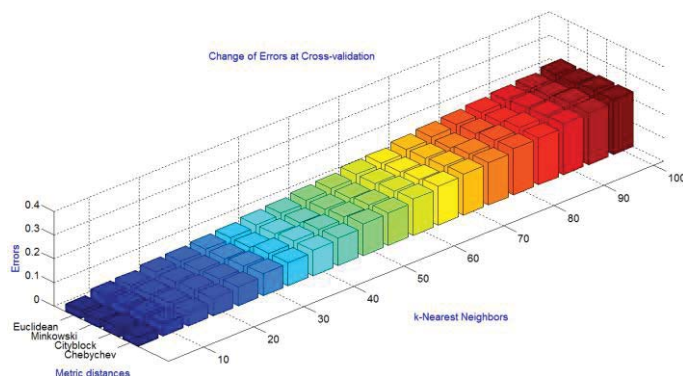
Fig. 5. Cross-validation errors at a) Euclidean/Minkowski, b) Cityblock and c) Chebychev distances

3D representations of the resulting errors of model analysis are shown in Fig. 4 and Fig. 5. The following variational limits were obtained:

For resubstitution:

- Euclidean / Minkowski: from 0.0210 to 0.1665;
- Cityblock: from 0.0195 to 0.1785;
- Chebychev: from 0.0215 to 0.1775.

at cross-validation:

- Euclidean: from 0.0325 to 0.2225;
- Minkowski: from 0.0315 to 0.2225;
- Cityblock: from 0.0375 to 0.2310;
- Chebychev: from 0.0365 to 0.2315.

Their accuracy levels are calculated according to faults (error "0" is equal to 100.00 and "1" to zero precisions). The accuracy results for signal identification are contained in Table 1 and Table 2.

TABLE I
ACCURACIES AT RESUBSTITUTION, %

k-neighbors	Distances №1 and №2	Distance №3	Distance №4
5	97.900	98.050	97.850
10	96.450	96.300	95.950
15	95.550	95.350	95.250
20	94.400	94.350	94.150
25	94.200	93.600	93.800
30	93.300	92.800	92.800
35	92.900	92.600	92.200
40	92.200	91.600	91.350
45	91.950	90.950	91.150
50	91.400	90.400	90.250
55	91.000	89.900	89.600
60	90.100	89.100	88.450
65	89.350	88.600	88.000
70	88.400	87.650	86.850
75	87.900	87.000	86.250
80	86.750	85.900	85.200
85	86.150	85.550	84.700
90	84.900	84.550	83.350
95	84.350	83.550	83.250
100	83.350	82.150	82.250

TABLE II
ACCURACIES AT CROSS-VALIDATION, %

k-neigh.	Dist. №1	Dist. №2	Dist. №3	Dist. №4
5	96.750	96.850	96.250	96.350
10	94.850	94.950	94.900	94.700
15	94.550	94.400	94.100	94.200
20	93.150	93.400	92.850	92.700
25	92.800	92.750	92.450	92.000
30	92.000	92.100	91.350	90.650
35	91.300	91.550	90.150	90.250
40	90.150		89.250	
45	89.500	89.650	88.550	88.400
50	88.300	88.250	87.650	
55	87.550	87.750	86.750	86.250
60	85.900		85.200	84.650
65	85.400	85.250	84.600	83.550
70	83.850	83.700	83.100	82.600
75	82.800	83.600	82.200	82.050
80	81.900	81.800	80.700	80.450
85	81.000	81.250	79.650	80.200
90	80.000	79.400	78.850	78.450
95	79.300	79.000	77.750	78.100
100	77.750		76.900	76.850

There are achieved, regarding the assessment of the quality of classification in the procedures of resubstitution:

- ✓ minimum accuracy 83.350%, 85.150% and 82.250% at Euclidean/Minkowski, Cityblock and Chebychev distances;
- ✓ Highest value of the 98.050% for criterion for Cityblock, followed by 97.900% for Euclidean/Minkowski and 97.850% for Chebychev.

In test validation are established:

- ✓ Highest criteria of success rate in a sequential order of 96.850% for Minkowski, 96.750% for Euclidean, 96.350% for Chebychev and 96.250% for Cityblock distances;
- ✓ Lowest accuracy, as follows: 77.750% for Euclidean/Minkowski distances, 76.900% for Cityblock and 76.850% for Chebichev.
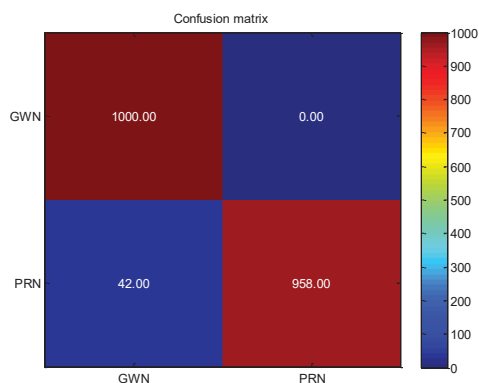
Fig. 6. Confusion matrix at Resubstitution
Euclidean/Minkowski distances for k = 5
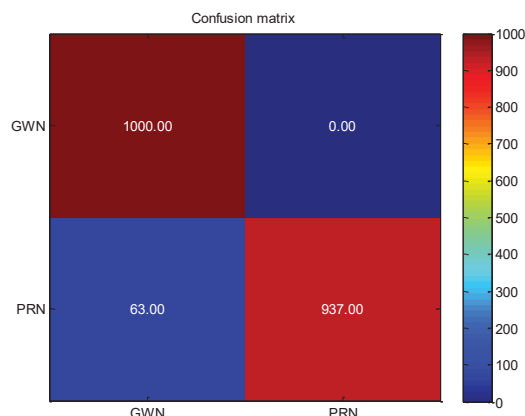
Fig. 7. Confusion matrix at Cross-validation
at Minkowski distance for k = 5

In terms of results from an analysis of classifiers concerned, matrices of correct and incorrect classifications are composed. Matrices with minimum parameter k = 5 in Fig. 6 and Fig. 7 informs about the distribution of the prototypes by groups, for Minkowski distance respectively:

- the first element of the first matrix line represents those with a properly defined affiliation of Class №1;
- The second element of the second matrix line represents the correct patterns of Class №2.

TABLE III
APPROXIMATELY EXPECTED ACCURACIES, %

k-neigh.	Dist. №1	Dist. №2	Dist. №3	Dist. №4
5	97.325	97.375	97.150	97.100
10	95.650	95.700	95.600	95.325
15	95.020	94.975	94.725	
20	93.775	93.900	93.575	93.425
25	93.500	93.475	93.025	92.900
30	92.650	92.700	92.075	91.725
35	92.100	92.225	91.375	91.225
40	91.175		90.425	90.300
45	90.725	90.800	89.750	89.775
50	89.850	89.825	89.025	88.950
55	89.275	89.375	88.325	87.925
60	88.000		87.150	86.550
65	87.375	87.300	86.600	85.775
70	86.125	86.050	85.375	84.725
75	85.350	85.800	84.600	84.150
80	84.325	84.275	83.300	82.825
85	83.575	83.700	82.600	82.450
90	82.450	82.150	81.700	80.900
95	81.825	81.675	80.650	80.675
100	80.550		79.525	79.550

The forecasts for approximate estimates of accuracy values in connection with identification of GWN or PRN impacts to signals that did not participate in k-NN patterns are given in Table 4.3. Models with the most appropriate and lowest

potential application for noise identification are synthesized with 97.375% accuracy at k = 5 and 79.525% k = 100 at Minkowski and Cityblock metric distances.
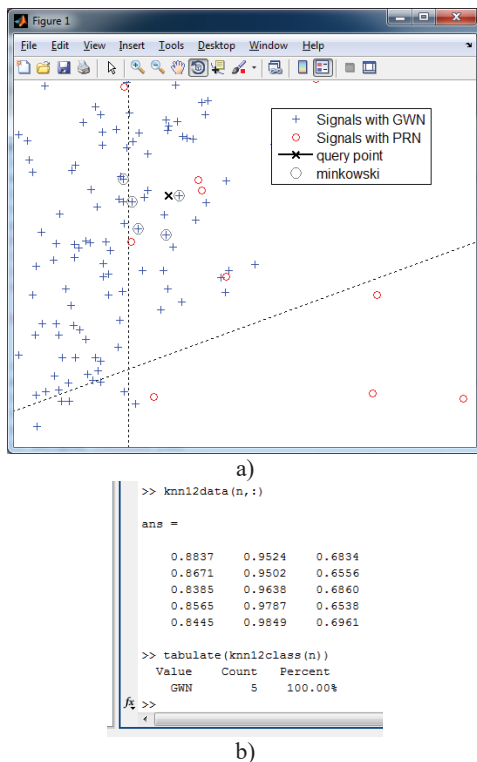
Fig. 8. k-NN search for class №1 at Minkowski distance
for k = 5 in a) graphical and b) numeric types
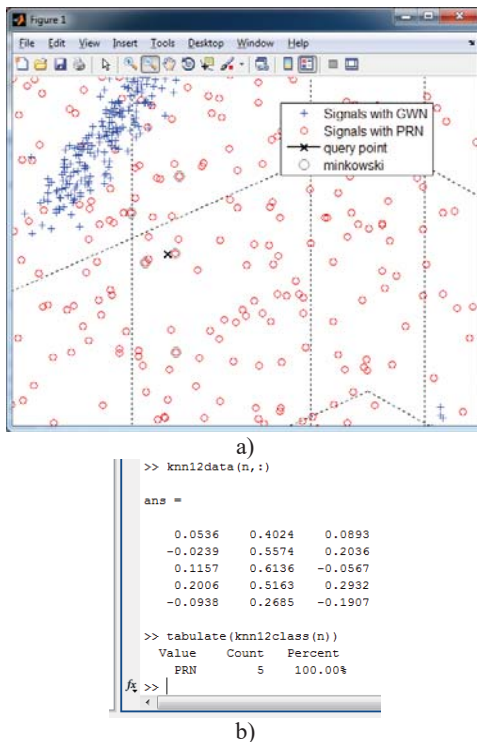
Fig. 9. k-NN search for class №2 at Minkowski distance
for k = 5 in a) graphical and b) numeric types

New points have been claimed in the formed classification space to illustrate the k-neighbors search process through the selected best model. The new patterns are marked with "x", whereas the k-neighbors found in the search areas k = 5 are surrounded by "o" (Fig. 8.a) and Fig. 9.a)). In addition to the graphical, the nearest found prototypes at a defined Minkovski distance are also shown in numerical form (Fig. 8.b) and Fig. 9.b)). The identification process shows 100.00% accuracy in both test groups.

IV. CONCLUSION

The analysis of the background noise to analogue and digital signals using a statistical method k - Nearest Neighbors is assessed with a high degree of efficiency. A drop in accuracy of less than 90.00% is observed when searching for more than 40 points in the vicinity of new patterns. The synthesized classification model can be integrated into various signal processing systems in communications and electronics.

REFERENCES

[1] T. Tanantong , "A kNN Approach for ECG Signal Quality Classification", International Journal of Information and Electronics Engineering, vol. 66, no. 4, pp. 269-272, 2016.

[2] K. Sabanci, M. Koklu, "The Classification of Eye State by Using kNN and MLP Classification Models According to the EEG Signals", International Journal of Intelligent Systems and Applications in Engineering, vol. 3, no. 3, pp. 127-130, 2015.

[3] V. Kuandanya, A. Patil, A. Panat, "Classification of Emotions from EEG Using k-NN Classifier", International Journal of Advances in Science Engineering and Technology, vol. 3, no. 2, pp. 103-106, 2015.

[4] N. Isa, A. Amir, A. Panat, M. Iyas, M. Razalli, "The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on EEG Signal", MATEC Web of Sciences 140, International Conference, on Emerging Electronic, Solutions for IoT, 01024, pp. 1-6, 2017.

[5] Ch. Prakash, V. Gaikwad, R. Singh, Om Prakash, "Analysis of Emotion Recognition System through Speech Signal Using KNN & GMM Classifier", International Journal of Advances in Science Engineering and Technology, vol. 10, no. 2, pp. 55-61, 2015.

[6] T. Priya, N. Raajan, N. Radju, P. Preethi, S. Mathini, "Speech and Non-speech Identification and Classification Using KNN Algorithm", ELSEVIER, International Conference on Modeling Optimization and Computing, vol. 38, pp. 952-958, 2012.

[7] H. Lee, J. Kim, S. Wibowo and S. Kim, "Bagged Fuzzy k-nearest Neighbors for Identifying Anomalous Propagation in Radar Images", INTELLI, The Sixth International Conference on Intelligent Systems and Applications, pp. 62-67, 2017.

[8] L. Lochumsen, "Radar Target Classification using Recursive Knowledge-Based Methods - Dissertation", Aalborg Universitetsforlag. Ph.d.-serien for Det Teknisk-Naturvidenskabelige Fakultet, Aalborg Universitet, pp. 1-152, 2016.

[9] Rankovska, V., S. Rankovski. Multi-functional Demo Module for Microprocessor Development Systems for Educational Purposes. Proc. XXVII International Scientific Conference Electronics - ET2018, September 13 - 15, 2018, Sozopol, Bulgaria, 2018, pp. 55-58. ISBN: 978-1-5386-6691-3. IEEE Catalog Number CFP18H39-CDR.

The Influence of Early Reflections and Babble Noise on the Intelligibility of Speech Signal

Dijana Kostić[1], Zoran Milivojević[2], Zoran Veličković[3]

Abstract – **The first part of the paper describes the effect of the first reflected component on intelligibility. In the second part of the paper, an experiment was described in which the intelligibility of words and sentences spoken on the Serbian language was tested under the conditions of superimposed Babble noise BN8 and in the presence of the first reflection. The analysis was performed for a delay up to 50 ms. Objective (STOI algorithm) and subjective (MOS test) testing were performed. The results are shown tabular and graphically. The analysis of the results determinated the delay, in which obtain the best intelligibility without the influence of the Babble noise, as well as the great destructiveness of the Babble noise to intelligibility.**

Keywords – **Intelligibility, Reflection, Babble noise, SMT.**

I. INTRODUCTION

Sound as a means of communication (speech or music) plays a major role in human life. For successful communication, it is necessary to provide good intelligibility. Intelligibility is one of the essential characteristics of communication systems, but also of spatial acoustics. On its transmission path from the sound source to the listener, the reproduced audio signal may suffer some degradation. Signal degradation refers to the impact of various types of interference: Gaussian Babble, Industrial noise, etc. In speech communication in a spatial acoustic (room, classroom, hall) early reflections also have a significant impact on the speech signal. The effect of the first reflection is particularly significant. In order to assess the intelligibility of speech in conditions of interference, different methods have been developed, and one them is speech audiometry. It involves the use of certain spoken material that is reproduced to the examinee: words (logatoms) or sentences (everyday and matrix), which after reproduction should repeat what they understood.

The words most commonly used in order to obtain objective test results are logatoms. Logatoms represent the meaningless words, composed of consonants (C) and vocals (V), constructing with the logatom type: CCV [1], CVC etc.

The words from everyday speech were created by Plomp and Mimpen in 1979 [2]. They formed a list of 170 sentences spoken in the Dutch language. When forming sentences, it was taken into account, that the words in the sentence do not contain more than 3 syllables.

The matrix sentences test was obtained by forming sentences according to the precisely defined order of the word in the sentence, so-called fixed syntax structure (name, verb, number, adjective, noun). This type of test was first developed for Swedish language by Hagerman [3]. The obtained results this kind of test have proven to be good, so this type of test has been developed for some other languages: Russian [4], Spanish [5], Serbian [6] ...

Respecting the rules set by Hagerman [3], the authors formed the Serbian Matrix Sentence Test - SMST base of 50 words (5 types of words x 10 words of each type) [6]. The authors also took care that the words were phonetically balanced and that they had no more than 3 syllables in their composition. In order to determine whether the words in the SMST base have a good appearance of the phonemes and reflect the spirit of the Serbian language, a comparative analysis was performed with capital literary works written in Serbian language: the novels "Bridge on the Drina" by Ivo Andrić, "Bakonja fra Brne" Sima Matavulj, epics "The Mountain wreath" by Petar Petrović Njegoš and drama "Koštana", author Bora Stanković [6].

In this paper, an experiment was performed to determine the influence of early reflections and the Babble noise on the intelligibility of the speech signal from the SMST base. The experiment is organized in several steps:

a) a speech signal consisting of a words from the SMST base were created, to which the Babble noise is superimposed, with the ratio SNR = {-5, -2, 0} dB;

b) the delay of the speech signal was made for values Δt = {0, 10, 25, 50} ms and

c) The testing intelligibility of speech signal was performing using: MOS and STOI test.

By comparing the results of the experiment with the results shown in [7] for Gaussian and Babble noise, as well as with the International standard [8], a conclusion was made for intelligibility of the speech signal.

The paper is organized in the following way. Section II describes the effect of the first reflexion and the delay time on intelligibility. Section III describes the experiment, experimental results and analysis. Section IV is a conclusion.

II. EFFECT OF THE FIRST REFLECTION

The early reflection is a sound wave reflected from a certain surface (an obstacle encountered in its path): a wall, a ceiling, and a floor and arrives to the listener almost at the same time as a direct sound. The first reflection usually appears very quickly after a direct wave in the range of $0 \div 50$ ms [9]. In order to ensure good intelligibility, it is considered

[1] Dijana Kostić, College of Applied Technical Sciences of Niš, 20. Aleksandra Medvedeva, St, 18000 Niš, Serbia, e-mail: koricanac@yahoo.com

[2] Zoran Milivojević, College of Applied Technical Sciences of Niš, 20. Aleksandra Medvedeva, St, 18000 Niš, Serbia, e-mail: zoran.milivojevic@vtsnis.edu.rs

[3] Zoran Veličković, College of Applied Technical Sciences of Niš, 20. Aleksandra Medvedeva, St, 18000 Niš, Serbia, e-mail: zoran.velickovic@vtsnis.edu.rs

that the first reflection should appear at an interval of 30 ÷ 80 ms after direct sound [10]. Beside that, in [11] it has been shown that early reflection can improve the intelligibility of speech in the absence of background noise.

It has been shown in [10], [11] that the first reflections that reach to the listener for about 35 ms after a direct wave, contribute to increasing the volume of the sound, and in this way can improve the intelligibility of speech. However, in conditions of ambient background noise there is no improvement in intelligibility [12], [13].

In the next part of the paper, an experiment was performed in which the intelligibility was tested in the presence of a strong first reflection (the amplitude of the direct component and the first reflected components are the same) and the superimposed Babble noise BN8.

III. Experimental results and analysis

The experiment was performed to determine the influence of early reflections and the Babble noise on the intelligibility of the speech signal from the SMST base (words and sentences).

A. Experiment

The experiment was realized in the following steps (Fig.1). *Step 1*: By combining words from the SMST base, sentences with a fixed grammatical structure are formed, i.e., a clear speech signal **x** is formed. *Step 2*: The Bubble noise **BN8** are superimposed to the clear speech signal, after the amplification **k**. In this way, the signal y with the specified SNR is formed. *Step 3*: The reflected speech signal x_r is generated with the time of delay Δt of the speech signal **x**. *Step 4*: By superimposed of the speech signal **x**, the Babble noise **BN8** and the reflected speech x_r, a test signal **z** is formed. *Step 5*: Testing intelligibility of whole sentences and individual words by subjective MOS test and objective STOI test.
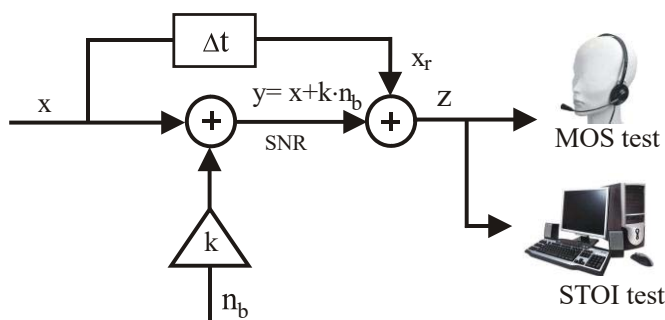
Fig 1. Block diagram of the realization MOS and STOI test. Amplitude of the reflected signal is A_r=1 [14].

The subjective MOS test (Step 5) was realized as follows. The test signal z was reproduced to the examinee trough the handset. After reproducing the test signal, the examinee repeated out loud what he had heard. After the answer the examiner note the accuracy of certain types of words as well as whole sentences. The statistical processing of the results

gives the intelligibility expressed in percentages. The objective test was realized using the STOI algorithm [15]. Test result is intelligibility in percentages. At the end, a comparative analysis of the intelligibility of speech was made in the presence of early reflections and Babble noise with the results of intelligibility of speech in the presence of Gaussian and Babble noise [7].

B. SMST base

For the purposes of experiment, words from the SMST base were used [6]. The words were spoken in Serbian language and they recorded in the studio of "Banker Radio" in Niš (Serbia). The word was read by a professional female speaker. The voice signal is memorized in a **wav** format on a hard disk with Fs = 44.1 kHz, 16 bps.

C. Test group

The test group were formed from students of the College of Applied Technical Science in Niš, with a gender structure: 13 men and 7 women (aged 19 to 45), μ = 22.9 years. Before the test, the examinees said, they are not aware that they had hearing problems.

D. The results

The results of the experiment are shown in the Tables I ÷ III and graphically on Figs.2 ÷ 6. The experiment was realized for SNR = {-5, -2, 0} dB and Δt = {0, 10, 25, 50} ms.

TABLE I INTELLIGIBILITY OF SPEECH WITH SUPERIMPOSED BN8 (SNR = 0 dB), IN THE PRESENCE OF EARLY REFLECTION (Δt)

	Δt (ms)	0	10	25	50	μ
Inteligibility (%)	Name	30	30	40	30	32,50
	Verb	55	35	55	40	46,25
	Number	55	60	65	55	58,75
	Adjective	65	50	30	35	45
	Object	60	35	45	45	46,25
	Sentence	5	0	0	0	1,25
	dSTOI	59,44	58,66	61,32	59,40	59,70

TABLE II INTELLIGIBILITY OF SPEECH WITH SUPERIMPOSED BN8 (SNR = -2 dB), IN THE PRESENCE OF EARLY REFLECTION (Δt)

	Δt (ms)	0	10	25	50	μ
Inteligibility (%)	Name	65	40	50	25	45
	Verb	80	55	45	55	58,75
	Number	90	55	50	30	56,25
	Adjective	70	50	45	35	50
	Object	40	45	25	25	33,75
	Sentence	15	10	10	0	8,75
	dSTOI	54,13	54,05	56,28	55,66	55,01

TABLE III INTELLIGIBILITY OF SPEECH WITH SUPERIMPOSED
BN8 (SNR=-5dB), IN THE PRESENCE OF EARLY REFLECTION (Δt)

	Δt (ms)	0	10	25	50	μ
Inteligibility (%)	Name	50	30	10	20	27,50
	Verb	70	35	25	20	37,50
	Number	75	55	45	35	52,50
	Adjective	40	10	15	30	23,75
	Object	40	20	5	10	18,75
	Sentence	10	0	0	0	2,50
	dSTOI	44,87	45,46	45,70	45,45	45,37

TABLE IV INTELLIGIBILITY OF SPEECH IN PRESENCE OF
GAUSSIAN AND BABBLE NOISE [7]

		Inteligibility (%)		
SNR (dB)		-5	-2	0
Name	Gaussian	46.67	66.67	63.33
	Babble	26.67	40	53.33
Verb	Gaussian	53.33	63.33	70
	Babble	6.67	10	36.67
Number	Gaussian	53.33	63.33	70
	Babble	10	36.67	53.33
Adjectiv	Gaussian	56.67	60	63.33
	Babble	3.33	20	46.67
Object	Gaussian	33.33	53.33	53.33
	Babble	6.67	16.67	30
Sentence	Gaussian	0	20	20
	Babble	0	6.67	6.67
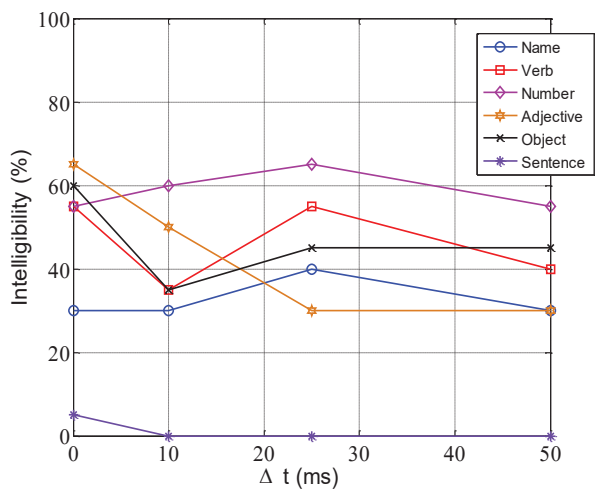
Fig. 2. Intelligibility of speech for: a) sentences and b) word for the
SNR=0dB.
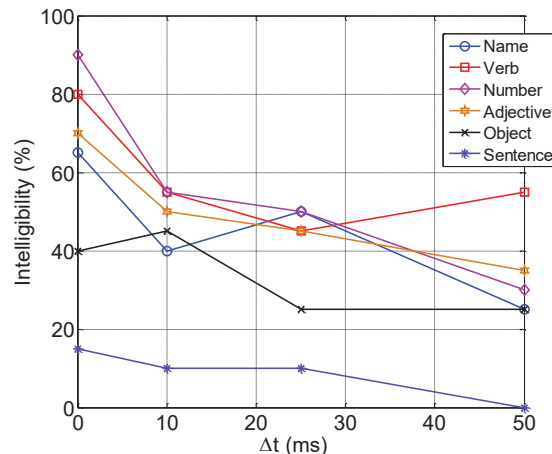
Fig 3. Intelligibility of speech for: a) sentences and b) word word for
the SNR=-2dB.
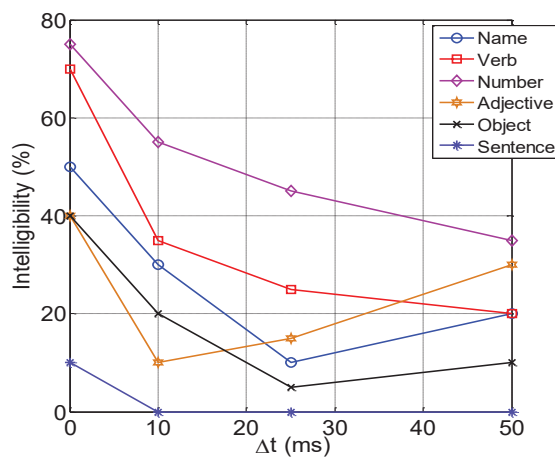
Fig 4. Intelligibility of speech for: a) sentences and b) word word for
the SNR= -5dB.
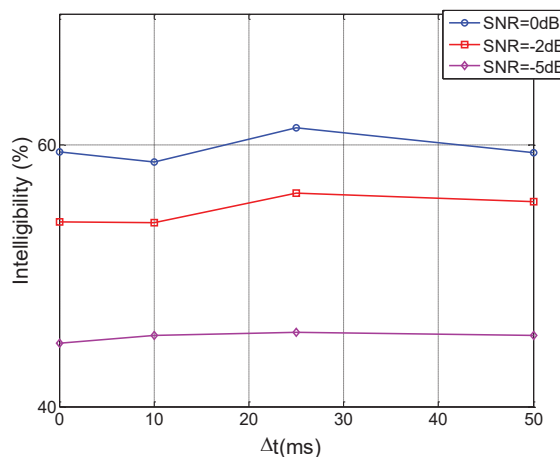
Fig 5. Intelligibility of speech dSTOI coefficient.

A. Analysis the results

Based on the results shown in Table I ÷ III and Fig. 2 ÷ 4 it can be concluded that intelligibility of type of word go in range:

a) for the 0 dB from 30 ÷ 65%,

b) for the -2 dB from 25 ÷ 90%,

c) for the -5 dB from 5 ÷ 75%,

Based on the results shown in Table I ÷ III and Fig. 2 ÷ 4 it can be concluded that intelligibility of word: *'Number'* is the best 90% for -2 dB and $\Delta t = 0$ ms, and the worst is *'Object'* for -5 dB and $\Delta t = 25$ ms.

Based on the results shown in Table I ÷ III and Fig. 2 ÷ 4 it can be concluded that intelligibility of sentence goes in range from 0 ÷ 15%, and it is the best for -2 dB (15%) and $\Delta t = 0$ ms, and the worst is for 0 dB (0%) and $\Delta t = (10, 25, 50)$ ms, for 2 dB (0%) and $\Delta t = 50$ ms, for -5 dB (0%) and $\Delta t = (10, 25, 50)$ ms.

Analysis the results shown in Table I÷III and Fig.5 it can be notice, that is dSTOI coefficient of intelligibility goes in range:

a) for the 0 dB from 58,66 ÷ 61,32%

b) for the -2 dB from 54,05 ÷ 56,28%

c) for the -5 dB from 44,87 ÷ 45,70%.

It can be notice that the greater time of delay have influence on the intelligibility. The great time of delay in combination with parameter SNR, when is strength of noise signal increased relative to strength of speech signal, have as results lower intelligibility of speech.

Observing the results from [7] in Table IV, comparing intelligibility of the speech signal in the presence of Gaussian and in the presence of Babble noise, it can be concluded that the results for Gaussian noise is better, while the results in the presence of the Babble noise, with and without the presence of the early reflections, do not differ significantly.

IV. CONCLUSION

In this paper the experimental results of speech intelligibility testing in the presence of Babble noise and early reflections are presents. The intelligibility test was performed by a subjective (MOS test) and an objective (STOI algorithm) test. Subjective MOS results show that increasing of value SNR = (-5 ÷ 0) dB increases the intelligibility of the word (18,75 ÷ 58,75) %, and sentences (2,5 ÷ 8,75) %. Increasing the time of delay of first reflection $\Delta t = (0 ÷ 50)$ ms leads to a decrease intelligibility of word (90 ÷ 5) %, and sentences (15 ÷ 0) %. Objective STOI results refer to the intelligibility of whole sentences. These results show that with the increasing of the SNR = (-5 ÷ 0) dB and $\Delta t = (0 ÷ 50)$ ms have influence on intelligibility of sentences. Increasing the SNR value the intelligibility of the sentences have been increases (45,37 ÷ 59,70) %, while when Δt increase the intelligibility of sentence decrease (61,32 ÷ 45,45) %.

The test results confirm the conclusion that there is a constant intelligibility of the word [10] up to 20 ms, and that the delay of about 30 ms can raise speech Intelligibility [12].

From the Figs. 2 it can be noticed that after "constant intelligibility" the intelligibility of words and sentences have been improved at $\Delta t = 25$ ms.

According to the standard IEC 60268-16, the obtained results for intelligibility are classified as "bad intelligibility" words (0 ÷ 67) % and sentences (0 ÷ 89) %.

The analysis shows that there is a greater degradation of intelligibility in the presence of Babble noise because of the identical distribution of energy in the spectral domain.

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.

[2] R. Plomp, A.M. Mimpen," Improving the reliability of testing the speech reception threshold for sentences", Audiology, pp 18, 43–52, 1979.

[3] B. Hagerman,"Sentences for testing speech intelligibility in noise", Scand Audio, Vol. 11, pp. 79-87, 1982.

[4] M. Boboshko, A. Warzybok, M. Zokoll, N. Maltseva, RUMatrix test: construction, evaluation and clinical validation. Otorhinolaryngologia Hungarica. Vol. 59, N 2., 2013.

[5] S. Hochmuth, T. Brand, M. Zokoll, F. Zenker, N. Wardenga, B. Kollmeier, A Spanish matrix sentence test for assessing speech reception thresholds in noise. Int. J. Audiol. 51(7) pp. 536-544, 2012.

[6] Z. Milivojević, D. Kostić, Z. Veličković, D. Brodić, "Serbian sentence matrix test for speech intelligibility measurement in different reverberation conditions", UNITEH Gabrovo, 2016.

[7] Z. Milivojević, D. Kostić, D. Brodić,"Performanse razumljivosti Srpskog MST-a u prisustvu akustičkog Gausovog šuma", INFOTEH Jahorina, 2017.

[8] International Electrotechnical Commission IEC 60268-16 - International Standard: Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index, Switzerland: IEC, 2011.

[9] I. Arweiler, J. M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility", J.Acoust.Soc.Am.130 (2), pp. 996-1005, 2011.

[10] L. L. Beranek, T. J. Mellow, "Acoustics: Sounde Fields and Transducers", 2012.

[11] J. Lochner, F. Burger, "The influence of reflections on auditorium acoustics", J Sound Vib. 1 (1964).

[12] E., Parizet, J. D. Polack, "The influence of an early reflection upon speechintelligibility in the presence of a background-noise," Acustica 77, 21-30,1992.

[13] A. K. Nabelek, J. M. Pickett "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners", J Speech Hear. Res. 7 (1974).

[14] A.Warzybok, J. Rennie, T.Brand, S.Doclo, B. Kollmeier, "Effect of spatial and temporal integration of a single early reflection on speech intelligibility", J.Acoust.Soc.Am.133(1), pp. 269-282, 2013

[15] Z. Milivojević, D. Kostić, Z. Veličković, "The optimization of the STOI algorithm parameters in presence of the WGN", ICEST 2018, Sozopol, Bulgaria.

APPLICATION OF THE CONTROL OF CHAOTIC PROCESSES IN THE ELECTRONIC COMMUNICATIONS

Galina Cherneva[1]

Abstract − **This paper deals questions related to the ability to control the chaotic processes in nonlinear systems with dynamic chaos. In the proposed approach is synthesized external impact on the system parameters. These parameters extend the phase trajectory from equilibrium points to achieve the chaotic regime. The approach is applied to dynamic system, described by the equations of Lorenz, as the study is done in an environment of Mathcad.**

Keywords – **nonlinear dynamics, Chaos Shift Keying, chaotic processes**

I. INTRODUCTION

The most important achievements of nonlinear dynamics in recent decades are the theory of the determined chaos. Chaotic processes are quite common and can occur in systems of different nature – economic, physical, and biological, etc. [3, 4, 5]. The chaotic signal is a relatively a new field in communication systems, too. Until the sixties of the last century the term "chaos" is associated with unpredictable and uncontrollable processes. The combination of the terms "control" and "chaos" is considered paradoxical. But in the last decade of the twentieth century this concept is being changed. On the limit of two studies: nonlinear dynamics and the theory of control emerged scientific direction, which explore possibilities to control the chaotic processes. Under this concept is the reforming of process by focusing a little impacts on the system in periodic, quasi-periodic, or in chaotic, but with other properties [1, 2, 3, 4, 5].

The control of the chaotic processes is associated with their sensitivity to the initial conditions [6, 7], as a requirement for performance of chaotic process is their parameters. As it is known [8, 9] small difference in the initial condition can lead to exponential decay of the trajectories, which achieve the aim of control.

Another important factor is that the systems with chaotic behavior are typical many unstable states, for example, between them can happen a little change of the system's parameters.

The main idea of the control methods of the chaotic processes in this paper is to reach needed attractor. It happens

by impact on the parameters of the system.

The idea to obtaining needed chaotic attractor is applied to the chaotic generators in Chaos Shift Keying (CSK) communication system (CS) [10, 11] is proposed an approach for the synthesis of impacts of the chaotic generators parameters. Through which it is increased the amplitude of the deviation of the phase trajectory from the balance points until reaching the chaotic mode. The proposed approach is applied to the system described by the equations of Lorenz [3] and the research was done of environment of Mathcad.

II. REVIEW OF CSK APPROACH

A block diagram of the CS with CSK is shown in figure 1. The chaotic system with parameters μ is located in the
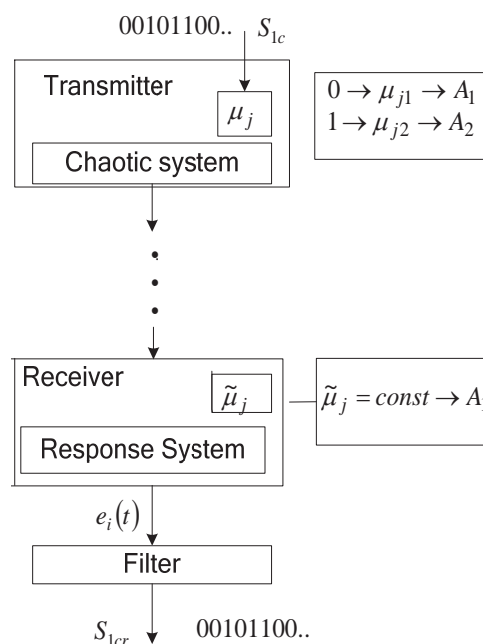
Fig. 1. A principle block diagram of the CS with CSK

transmitter.

The transmitter dynamics are dissipative and chaotic and the transmitter state trajectory converges to a strange attractor.

A message is transmitted by changing one of the parameters μ (μ_j) which results in a change of the transmitter attractor dynamics.

[1] Galina Cherneva is with the Faculty of Telecommunications and Electrical Equipment in Transport at Higher School of Transport of Sofia, 158 Geo Milev Str., Bulgaria, E-mail: galja_cherneva@abv.bg

If value μ_{j1} responds to symbol "0" and value μ_{j2} to symbol "1" of the binary information signal S_{1c}, the transmitter chaotic generator switches over attractors A_1 and A_2.

In the receiver is the response system. In coherent detection [5], the receiver is required to reproduce the same chaotic signals sent by the transmitter, often through a chaos synchronization process.

During transmission, the response system parameter $\tilde{\mu}_{j1} =$ constant. All other parameters of the chaotic system in the transmitter are identical to their corresponding parameters in the response system in the receiver.

So, synchronization will occur only during periods of submission to symbol "0", when there is a complete coincidence between the parameters of both systems.

III. OBTAINING OF THE PARAMETER VALUE

Let the chaotic system in transmitter be defined by a system of differential equations.

$$\dot{x} = f(x, \mu), \qquad (1)$$

where $x = (x_1, x_2, ..., x_n)$ is the multitude of variables describe the state of the system defined in the state space R^n, μ is combination of control parameters.

For the system (1) are defined balanced states $\{\vec{x}^*\}$ with coordinates $x_1^*, x_2^*, ..., x_n^*$.

The stochastic process of the fluctuation, is a solution of (1), is related to the increase of the amplitude on the deviation of the phase trajectory in the surround of the balanced points. This deviation is:

$$R = \sqrt{\left(x_1 - x_1^*\right)^2 + \left(x_2 - x_2^*\right)^2 + ... + \left(x_n - x_n^*\right)^2} \qquad (2)$$

If we express the derivatives (3) and define them about the parameters μ, we will determine the law of change of influence $p(t)$, through which it can be achieved a chaotic mode of fluctuations.

$$\frac{d\left(x_i - x_i^*\right)^2}{dt} \qquad (3)$$

Then they are synthesized external control impacts by the type (4) where K is a weight coefficient

$$P(t) = K.p(t) \qquad (4)$$

They are shown into the system of differential equations (1) so that it has the type:

$$\dot{x} = f(x, \mu) + P(t) \qquad (5)$$

IV. MODEL CONSTRUCTION

Let the transmitter chaotic generator be obtained by Lorenz system equation:

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= xy - \beta z \end{aligned} \qquad (6)$$

where σ, ρ and β are parameters, $x(t)$, $y(t)$, $z(t)$ – variables.

It is determined that at $\rho > 1$ the balanced states of (6) are points with coordinates

$$\begin{aligned} x_{1,2}^* &= \pm\sqrt{\beta(\rho - 1)} \\ y_{1,2}^* &= \pm\sqrt{\beta(\rho - 1)} \\ z_{1,2}^* &= \rho - 1 \end{aligned} \qquad (7)$$

According to the equation (5) in the model of Lorentz is introduced external impact P_ρ on the second equation of the system (7):

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz + P_\rho \\ \dot{z} &= xy - \beta z \end{aligned} \qquad (8)$$

The task is to synthesize the function P_ρ so that the model (8) has a chaotic mode by the new value of parameter ρ.

For this aim is expressed

$$\begin{aligned} \frac{d\left(y - y^*\right)^2}{dt} &= 2\left(y - y^*\right)\frac{d\left(y - y^*\right)}{dt} = \\ &= 2\left(y - y^*\right)\left(\rho x - y - xz\right) \end{aligned} \qquad (9)$$

After definition of (9) towards ρ is prepared

$$\frac{d\left[2\left(y - y^*\right)\left(\rho x - y - xz\right)\right]}{d\rho} = 2\left(y - y^*\right)x . \qquad (10)$$

This equation (10) gives the opportunity influence to be used on the parameter ρ. Which are determined by the sign of the current deviation of the variable y from the balanced coordinate (11), where ρ_0 is the beginning value of the parameter.

$$p_\rho = \rho_0 + sign\left(y - y^*\right)\rho_0 \qquad (11)$$

On the basis of (11) is formed (12), where the weight factor K is selected depending on the desired attractor.

$$P_\rho = Kp_\rho \qquad (12)$$

The described approach is simulated by Mathcad, as they are used parameters with values $\sigma = 10$, $\rho = 10$ and $\beta = \frac{8}{3}$. They are shown in figure 2.
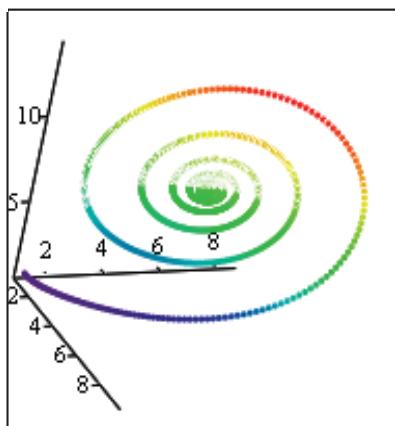
(X, Y, Z)

Fig. 2. Periodical mode

After realization of an effect of the type (12) we received a chaotic mode. When K = 0.1 the attractor look like in figure 3. This is a realization of parameter $\mu_{j1} = \rho_1$.
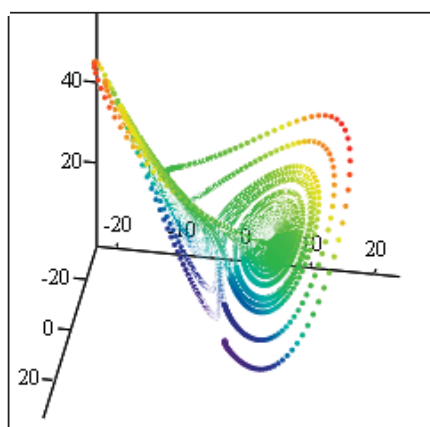
$(X1, Y1, Z1)$

Fig. 3. Chaotic mode at K=0,1

When K = 0.2 is given the classical attracter of Lorentz looks like in the figure 4. This is a realization of parameter $\mu_{j2} = \rho_2$.

From the given three-dimensional phase trajectories shown in figure 3 and figure 4 can be seen how through the influence P_ρ (12) is realized the needful value of the parameter ρ.
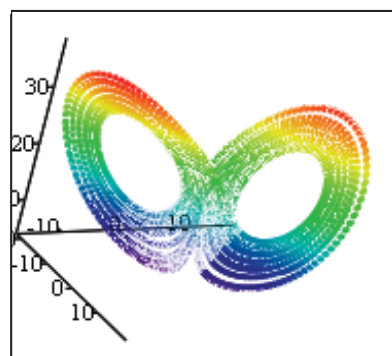
$(X2, Y2, Z2)$

Fig.4.Chaotic mode at K = 0.2

V. CONCLUSION

The studies, made in the paper, show that chaotic system can realize chaotic mode through external influence of its parameters, which are formed of the basis of the deviation of the phase trajectories of the balanced states. Through the changing on the strength of the impact on the parameter, which is expressed by introduce weighting factor, can achieve the wanted chaotic attractor.

The given approach for the control of the chaotic process can be applied to different chaotic system which aim is to achieve the wanted mode of work.

REFERENCES

[1] Zhang H., Derong L., Zhiliang W., Controlling chaos: Suppression, Synchronization and Chaotification, Springer, 2009, ISBN 978-1-84882-522-2.
[2] Banerjee S., Mitra M., Rondoni L., Application of Chaos and Nonlinear Dynamics in Engineering – Vol.1. Springer, 2011, ISBN 978-3-642-21922-1.
[3] Kapitaniak T., Chaos for Engineers. Theory, Applications and Control, *Springer*, 2000, ISBN 978-3-642-57143-5.
[4] Ott E., GrebogI C., Yorke J.A. Phys. Rev. Lett. 64. 1990. p. 1196.
[5] Azar A.T., Vaidayanathan, Advances in Chaos Theory and Intelligent Control, *Springer*, 2016, ISBN 978-3-319-30340-6.
[6] Andrievskii B. R., A. L. Fradkov, Control of Chaos: Methods and Applications. II. Applications. Automation and telemetry, 2003, issue 5, p. 3-45;
[7] Bolotin Y., Tur A., Yanovsky V., Chaos: Concepts, Control and Constructive Use, *Springer*, 2009, ISBN 978-3-319-42496-5.
[8] Chernousko F., Ananieski I., Reshmim S., Control of Nonlinear Dynamical Systems. *Springer*, 2008, ISBN 978-3-540-70784-4.
[9] Chen G., X. Yu, Chaos Control. Theory and Applications. *Springer*, 2003, ISSN 0170-8643.
[10] Ling B., H. Lam (ED.), Control of chaos in nonlinear circuit and systems. *World Scientific Series on Nonlinear Science*, Series A. 2009, Vol. 64, ISBN 13 978-981-279-056-9.
[11] Nagaev, R.F. (ED.), Dynamics of Synchronizing Systems, *Springer,* 2003, ISBN 978-3-540-45761-9.

CB-SVD watermarking algorithm for video protection with reduced cyclic insertion scheme

Zoran Veličković[1], Zoran Milivojević[1], Marko Veličković[1]

Abstract – **This paper presents the RCB-SVD watermarking algorithm for protecting video content from copying. The bit planes of the decomposed monochrome watermark are inserted into the video-frame color channels. A CB-SVD algorithm with a reduced cyclic insertion scheme was used to insert the watermark. Reduced insertion scheme reduces the level of degradation of video while simultaneously extracting a high quality watermark. The RCB-SVD algorithm does not have the lack of detection of a false watermark, as is the case with standard SVD algorithms. The built-in redundancy of the reduced cyclical insertion scheme increases the resistance to "frame-dropping" attacks. Compared to the previously announced results, repair of the extracted watermark was achieved by 9.87%.**

Keywords – ***Bit-plane decomposition; CB-SVD algorithm; Watermarking; Reduced cyclic insertion scheme.***

I. INTRODUCTION

A high performance level of modern communication networks has been achieved by combining new modulation techniques and new generation communication protocols [1]. In addition to the large network flow, modern communication networks also reduce packet latency to just a few milliseconds. Thus, 4G network technologies provided a latency of 50 ms, while from 5G technology it is expected that latency be only 1 ms. With 5G technology, the whole movie in HD format can be downloaded in less than 10 s. By using these technologies, the exchange of digital multimedia content has become a common user activity. Setting up, and then sharing multimedia content on social networks like *Facebook*, *YouTube* or *Instagram* can now be implemented with efficient procedures. By 2022, 82% of network traffic is expected to relate to some kind of video communication [2]. There is a rise in video surveillance, Internet video to TV and Consumer Video-on-Demand traffic.

However, the exchange of multimedia content on the network has led to an increase in security risks and the occurrence of copyright and identity abuse. Multimedia content can be downloaded indefinitely without loss, then modified and used illegally for commercial or other purposes. Especially important are the problems of copyright protection in this

network environment. One way to reduce security risks, before sharing multimedia content, is to perform reliable user authentication [3].

In order to prevent the illegal use of multimedia content, numerous protection techniques have been developed. The problems of protecting multimedia content can be solved by the classic encryption technique. However, classical information-based security systems based on encryption are not adequate for application to multimedia content. The main disadvantages of standard cryptographic techniques can be described through the following facts:

a) *inefficiency*, it is necessary to transfer a huge amount of data further when exchanging encrypted multimedia content,

b) *incomplete protection*, multimedia content is protected only during network transfer,

c) *decrypting during playback*, multimedia content must be decrypted for reception, which potentially makes it unsafe.

Inserting invisible - secret information (watermark) into multimedia content is a frequently used technique to increase information security on a global computer network - the Internet [3] - [14]. In order to improve the performance of individual protection algorithms, *hybrid insertion techniques* are used [4]. These techniques involve the implementation of multiple transformation domains such as DCT, SVD and DWT. The watermark can bee a color picture [5], a monochrome (grayscale) image [6], or a binary image [7].

In this paper, several different binary images obtained by the decomposition of the monochromatic watermark [8], [9] are inserted into the multimedia content – color video sequence. The basic problem with this protection concept is that inserting a watermark leads to interference in multimedia content. A stronger inserted watermark results in the quality decrease of the multimedia content, but it ensures the extraction of a quality watermark. On the other hand, the poorly inserted watermark less degrades the quality of multimedia content, but does not provide high quality watermark extraction. Clearly, these are two opposing requirements that the multimedia content algorithm must reconcile.

In order to protect the video from copying, RCB-SVD watermarking algorithm with reduced cyclic insertion scheme based on [10] is shown in this paper. Prior to insertion, the monochromatic watermark was decomposed into 8-bit planes (binary images) which were then inserted into the color channels [11]. Application of a reduced cyclical insertion scheme removes noticeable video degradation while simultaneously extracting a high quality watermark. The problem of fake watermark detection in standard SVD algorithms is solved, and built-in redundancy increases resistance to frame-dropping attacks. Evaluation of the proposed algorithm was performed in the software package Matlab.

[1]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.

[1]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.

[1]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
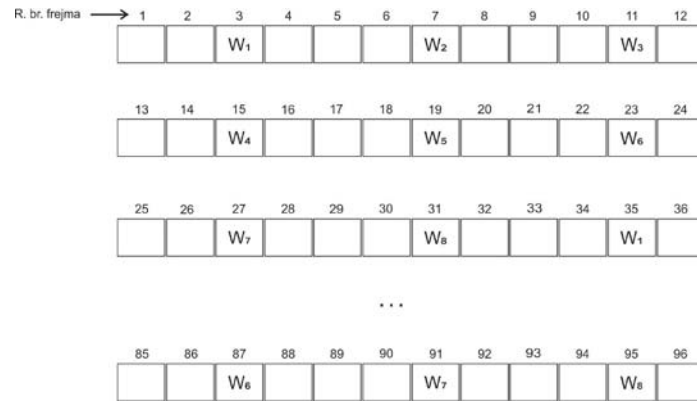
Fig. 1. Reduced cyclical scheme for inserting bit planes $W_1 - W_8$ into frames of unencrypted video.

Below follows a reduced insertion model, that is, the extraction of the decomposed watermark into multiple frames in the color domain of the video. Evaluation of the proposed method was carried out and the obtained results were analyzed. In the last section some conclusions and recommendations were made.

II. RCB-SVD ALGORITHM

The RCB-SVD video protection algorithm is based on the results published in [4] - [11]. Unlike [10] where the linear cyclic scheme of insertion is applied, in this paper a reduced cyclical insertion scheme is proposed. Fig. 1 shows a proposed reduced insertion scheme that reduces the level of interference in the video and provides extraction of the better watermark quality. The RCB-SVD algorithm is based on bit plane decomposition of the grayscale watermark and SVD decomposition of the frame and bit plane plane. The insertion and extraction algorithm is shown in more $\mathbf{I}$ and $\mathbf{E}$ steps respectively.

Step $\mathbf{I_1}$: Perform the bit-plane decomposition of the monochromatic watermark $W_{m \times n}$. In order to perform the decomposition of the watermark in eight bit-planes, first should be present the values of all the pixels of the watermark w_i with the corresponding binary values $b_{i, k}$ as follows:

$$b_{i,k} = \left\lfloor \frac{w_i}{2^{k-1}} \right\rfloor mod\ 2, \quad k = 1, 2, \dots, 8; \quad (1)$$

$$i = 1, 2, \dots, m \times n$$

In the expression (1), the w_i is represented the i-th pixel of the monochromatic watermark. The mark $b_{i,k}$ would represent 8 bits of binary value of the i-th pixel. The bit plane to the watermark $\mathbf{W_k}$ is formed from all k bits of all pixels of the watermark:

$$W^k = b_{i,k} \quad (2)$$

Step $\mathbf{I_2}$: The chrominent component of the frame U from the YUV video format is divided into unconnected blocks $\mathbf{H_{i,j}}$ dimensions of 4×4 pixels.

Step $\mathbf{I_3}$: For each block $\mathbf{H_{i,j}}$ from the $\mathbf{U}$ component, perform the SVD decomposition:

$$H_{i,j} = U_{i,j} \times S_{i,j} \times V_{i,j}^T \quad (3)$$

Step $\mathbf{I_4}$: Modify the elements from the second and third row of the first column of each matrix (elements $u_{2,1}$ and $u_{3,1}$) based on the value of each individual bit w from the corresponding bit level as follows [11] - [13]:

$$if\ \ w = 1, \begin{cases} u_{2,1}^* = sign(u_{2,1}) \times \left(U_{avg} + \frac{T}{2}\right) \\ u_{3,1}^* = sign(u_{3,1}) \times \left(U_{avg} - \frac{T}{2}\right) \end{cases} \quad (4)$$

$$if\ \ w = 0, \begin{cases} u_{2,1}^* = sign(u_{2,1}) \times \left(U_{avg} - \frac{T}{2}\right) \\ u_{3,1}^* = sign(u_{3,1}) \times \left(U_{avg} + \frac{T}{2}\right) \end{cases} \quad (5)$$

$$U_{avg} = \frac{(|u_{2,1}| + |u_{3,1}|)}{2} \quad (6)$$

where T represents the desired threshold in the difference between the elements $u_{2,1}$ and $u_{3,1}$ of $\mathbf{U_{i,j}}$ the matrix. The modified matrix is denoted by $\boldsymbol{U_{i,j}^*}$.

Step $\mathbf{I_5}$: Perform the inverse SVD transformation to obtain a block with the inserted bit from the watermark.

$$H_{i,j}^* = U_{i,j}^* \times S_{i,j} \times V_{i,j}^T \quad (7)$$

Step $\mathbf{I_6}$: Repeat steps 4 and 5 for all bits from the appropriate bit plane watermark.

Step $\mathbf{I_7}$: Repeat steps 2-6 for all bit watermark planes for all video frames using the reduced insertion scheme shown in Fig. 1. In this way, a video with the embedded watermark is obtained. For the extraction of the watermark thus inserted, it is *not necessary* to have the originals of either the video or the watermark, so this algorithm belongs to the class of *blind* algorithms.

A part of the RCB-SVD algorithm for watermark extraction is shown in 8-steps $\mathbf{E}$ [11] - [13].

Step $\mathbf{E_1}$: The component $\mathbf{U}$ of protected frame divide into nonoverlapped blocks $\mathbf{H'_{i,j}}$ of a dimension of 4×4 pixels.

Step $\mathbf{E_2}$: Perform SVD decomposition of all blocks $\mathbf{H'_{i,j}}$ from frame:

$$H_{i,j}' = U_{i,j}' \times S_{i,j}' \times V_{i,j}'^{T} \quad (8)$$

Step $\mathbf{E_3}$: The value of the corresponding extracted watermark w' is obtained using the following terms:

$$w' = 1, \begin{cases} 0, & if\ u_{2,1}' > u_{3,1}' \\ 1, & if\ u_{2,1}' \le u_{3,1}' \end{cases} \quad (9)$$
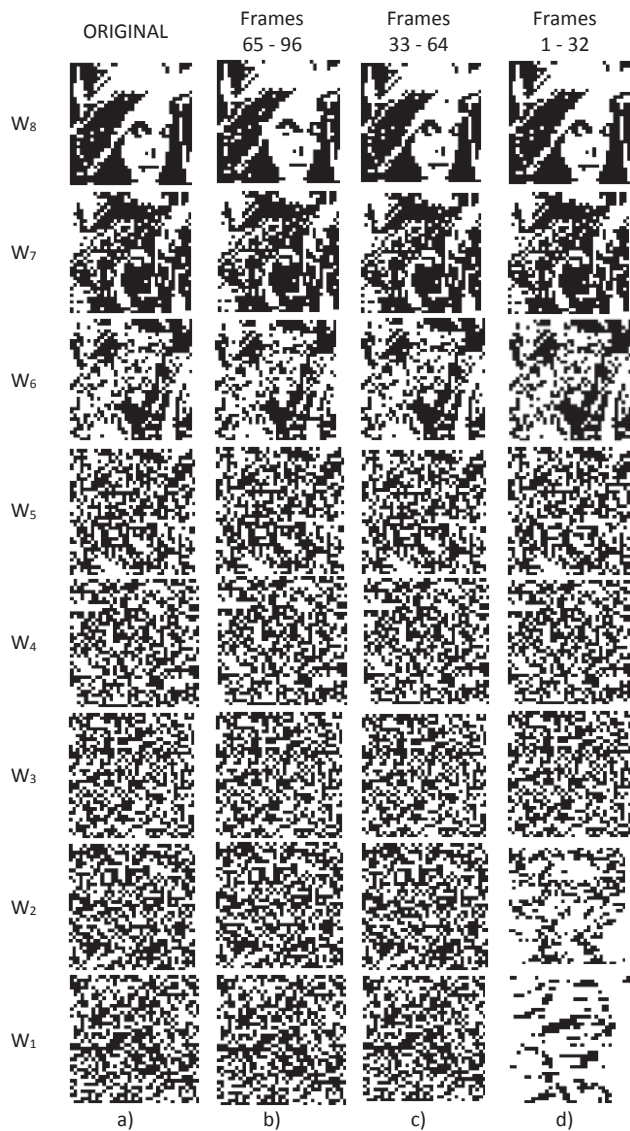
Fig 2. Bit planes W_1 do W_8 a) original watermark, b) extracted from frames 65-96 c) 33-64 d) 1-32 encoded video.

Step E_4: Repeat steps 2 and 3 for all no overlapped blocks of the frame.

Step E_5: Create a binary image of the corresponding bit-plane watermark.

Step E_6: Repeat steps 1-5 for eight frames from the video.

Step E_7: Form the grayscale watermark of 8 binary bit-planes.

Step E_8: Repeat steps 1-7 for all frames in which the bit-plane component of the watermark is inserted into the reduced insertion scheme shown in Fig. 1.

III. EXPERIMENTAL RESULTS

The monochromatic watermark used in the experimental part of this paper is shown in Fig. 3a). The illustrated watermark represents an adapted central part of the famous monochrome Lena.bmp image in a resolution of 36×36 pixels. Before inserting into a frame of uncoded color video - *Foreman.yuv* at a resolution of 288×288 pixels, the watermark was decomposed into 8 bit W_1 - W_8 planes. Bit planes W_1 - W_8 are

obtained by bit-plane decomposition in the manner described in the chapter "RCB-SVD algorithm". In Figure 2a), the images of bit planes of this monochrome watermark are shown. The MSB bit plane of the watermark W_8 is located at the top of the column *original*. Other bit-plane decomposition of watermark W_7-W_1 are shown below in this column. If we observe the displayed bit planes, it can be noticed that with approaching the LSB bit plane W_1, the watermark decomposition takes a stochastic character. This is why in this paper the GMSAT algorithm [14] does not use the scribbling of images as shown in previous works [6]. Only the bit planes W_8, W_7, and W_6 can be protected to increase the security level.

In this paper, a cyclic set of decomposed bit watermark planes are embedded in the first 96 frames of the color video *Foreman.yuv*. The reduced bit-planes insertion scheme is shown in Fig. 1. According to the reduced insertion scheme in the first two frames, no bit planes of the watermark are inserted. The first bit plane W_1 is embedded in the third frame. Then skips three frames and the other bit plane W_2 is embedded in the seventh frame. The process of inserting bit planes continues along the reduced scheme to the last frame to be protected. When all bit levels are inserted, bitwise W_1 is inserted cyclically. The insertion of watermark components according to the considered algorithm is performed in chrominant components of the frame with threshold T = 0.04. This threshold value is commonly used in algorithms of this type [5], [12], [16], but even with a lower threshold value, excellent results are obtained. A reduced insertion scheme provides a certain level of redundancy of watermark components. In this paper, the algorithm for repairing the extracted watermarks described in [13] was applied. If 96 frames of a non-encoded video are inserted bit planes according to the scheme shown, 24-bit planes can be extracted at the reception - 3 complete watermarks.

Insertion is done in a non-encoded video domain, and then encrypted with encrypted video by H264 / AV encoder. Since the H.264 / AV algorithm belongs to the loss coder class, many of the details present in the video are ignored and are forever lost. This will inevitably cause errors in the decoding process that the human eye does not notice. However, errors that occur during decoding will have an impact on the extraction of the inserted watermark. The encrypted video sequence was encoded by the JM reference software of ITU in version 18.4 FRExt [15]. The encoding quality is defined by a set of FRExt parameters: IntraPeriod = 12, NumberReferenceFrames = 5 NumberBFrames = 1. In Figure 2b), 2c) and 2d), the extracted bit planes of the W_1 - W_8 monochromatic watermark from the encoded frames 65 - 96, 33 - 64 and 1 - 32 are shown. Figures 3b), 3c) and 3d) show grayscale watermarks obtained by composing extracted bit planes from frames 65 - 96, 33 - 64, and 1 - 32. Below the figures, the obtained SSIM values of the extracted watermarks are shown.

Errors are evident first in the extracted bit plane (Fig. 2), and then in the composite monochromatic watermark obtained from them (Figure 3). In order to improve the quality of extracted watermarks, an algorithm for its repair was applied [14]. The redundancy provided by the RCB-SVD algorithm enables the extraction of the watermark.

In the shown example, the complete three grayscale
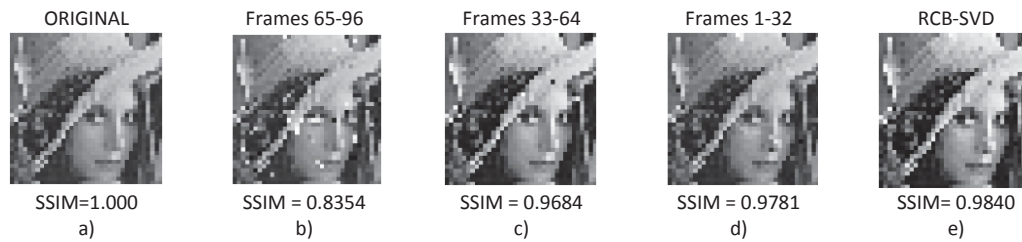
Fig 3. Watermark a) original b) extracted from frames 65-96 c) 33-64 d) 1-32 encoded video e) corrected.

watermarks with various SSIM values of 0.8354, 0.9684 and 0.9781 were extracted. After the repair, a watermark was obtained, the appearance of which is shown in Fig. 3e) with SSIM index 0.9840. It should be noted that for the application of this repair algorithm, the origin of the watermark must be known. An objective assessment of the quality of the extracted watermark measured by the SSIM index for the presented case is 0.9840, which is the highest index compared to each individual index of the extracted watermark. Compared to the previously published results [11], a significant correction of the extracted watermark was achieved by 9.87%.

IV. CONCLUSION

Protecting multimedia content from copying has become an increasingly important activity of the author before their publishing on the Internet. Inserting a watermark into video is a technology that provides copy protection for the entire duration of the existence of video content. In this paper, we present the CB-SVD algorithm with a reduced insertion scheme of the monochromatic watermark in the frames RCB-SVD. Prior to insertion, the monochromatic watermark was decomposed into 8 bit planes (binary images), which were then inserted with a reduced cyclic scheme into the color video channels. Using a reduced cyclical insertion scheme, the extraction of a high quality watermark is possible with the significantly higher quality of the protected video. Evaluation of the proposed algorithm was carried out on the known color video sequence Foreman, while a modified inserted monochrome image of Lena was used as a watermark in a resolution of 32×32 pixels. The proposed reduced insertion model causes significantly less video distortion, which outperforms the performance of standard SVD algorithms. The problem of extracting a false positive watermark was blocked. The built-in redundancy of this algorithm increases the resistance to frame-dropping attacks while at the same time allowing the extracted watermark to be repaired. Compared to the previously announced results, repair of the extracted watermark was achieved by 9.87%. In the continuation of the research, the RCB-SVD algorithm performance will be determined in the presence of different types of attacks. The results presented justify the application of the RCB-SVD algorithm to protect the video from copying.

REFERENCES

[1] M. Jevtović, Z. Veličković, "Protokoli prepletenih slojeva", Akademska misao, Beograd, 2012.

[2] White paper, "Cisco Visual Networking, Index: Forecast and Trends, 2017–2022", 2019.

[3] Z. Veličković, Z. Milivojević, M. Veličković, „The Watermark Applications in Multimodal Biometric Identification", ICIST 2018 Proceedings Vol.1, pp.65-69, 2018.

[4] E. Chrysochos, V. Fotopoulos, M. Xenos, A. N. Skodras, "Hybrid watermarking based on chaos and histogram modification", Sig. Im. Video Proc, pp. 843-857, 2014.

[5] Q Su, G. Wang, S. Jia, al. "Embedding color image watermark in color image based on two-level DCT", SIViP (2015) 9: 991. https://doi.org/10.1007/s11760-013-0534-2.

[6] Z. Veličković, M. Veličković, Z. Milivojević, Improved Gray-Scale Watermark Encryption Based on Chaotic Maps, UNITECH 2016, pp. II-145-150, Gabrovo, 2016.

[7] Z. Tang, J. Song, X. Zhang, R. Sun, "Multiple-image Encryption with Bit-plane Decomposition and Chaotic Maps", Optics and Laserrs Eng. vol. 80, 2016, pp. 1-11, https://doi.org/10.1016/j.optlaseng.2015.12.00.

[8] A. M Joshi, V. Mishra, R. M. Patrikar, "FPGA prototyping of video watermarking for ownership verification based on H.264/AVC", Multimed Tools Appl. (2016) 75: 3121. https://doi.org/10.1007/s11042-014-2426-z.

[9] K. C. Choi, C. M. Pin, "Robust lossless digital watermarking using integer transform with Bit-plane manipulation", Multimed. Tools Appl., 75: 21497, 2016, https://doi.org/10.1007/s11042-015-2596-3.

[10] Z. Veličković, Z. Milivojević, M. Veličković, „Color Blind SVD Watermarking Algorithm for Video Protection with Linear Cyclic Insertion Scheme ", INFOTEH, Jahorina 2019.

[11] Z. Veličković, Z. Milivojević, M. Veličković, „A secured digital video watermarking in chrominance channel", 2018 23rd Int. Sci. Prof. Conference on Information IT, Žabljak 2018, DOI: 10.1109/SPIT.2018.8350858.

[12] Shao-li Jia, "A novel blind color images watermarking based on SVD", Optik-International Journal for Light nad Electron Optics,Vol. 125, No. 12, pp. 2868-2874, June 2014.

[13] Z. Veličković, Z. Milivojević, M. Veličković, „Blind watermarking scheme in chrominance channel based on SVD and bit-plane decompositions", UNITECH 2018, pp. II 67-72, Gabrovo, 2018.

[14] Z. Veličković, Z. Milivojević, M. Veličković, "Digital video protection in the DWT-SVD domain using scrambled watermark by GMSAT algorithm", ETF Jour. of Electrical Engineering, Vol. 23, pp. 36-46, Podgorica, 2017.

[15] "JM reference software version 19.0," iphome.hhi.de, [Online]. Available: http://iphome.hhi.de/suehring/tml/.

[16] Bojan Prlinčević, Zoran Milivojević, Stefan Panić and Zoran Veličković, „Performance of the SVDU Watermarking Algorithm", ICIST 2017, Proceedings Vol. 1, pp. 266-270, Kopaonik, 2017.

Hybrid Method For Image Segmentation And Recoloring Of The Original Grayscale Photographs

Vladan Vučković[1], Sanja Spasić[2]

Abstract – **In this paper we will present hybrid method for recoloring the original black and white photos. Main novelty we attempt to introduce is specific combination between human and machine capabilities in simultaneously image processing. Algorithm covers different manual and automatic techniques in colorization resulting in better achievements. Cross combined methods for image segmentation have a successful image for the result. In this paper combined process methods which are based on Gray scale algorithm, from one side, and analyses of the patent photos in color, from the other side will be also presented. This suggested method requires restoration and segmentation of the photo, as the result of colorization, will give maximum effects. Original colorization algorithm of 2D photo will also be presented here, and that algorithm reduces colorization to the minimum, under the condition that the user obtains enough information about the photo itself during the colorization process. The generality of the proposed method will be demonstrated on the series of original photographs from the Nikola Tesla Museum of Belgrade heritage.**

Keywords – **Image processing, Segmentation, Coloring, Image restoration**

I. INTRODUCTION

Digital image processing belongs to the multidiscipline engineers` sphere which covers different aspects of human life from photos, mathematics, electronic, optics and computer science. In last few decades the number of the applications and techniques for the process of digital images increased. Considering that increase, the demands in relation with developing of technology for the color image process increased, too.

The basic spheres of digital image processing are representing representative. Remodeling image, improving image quality, image restoration, image analysis, reconstruction of image from projection and compression of image in the sphere of image analysis are present in the coloring, as well as in the image segmentation.

The term itself - *segmentation of image* [1] - is related to the group of procedures for dividing an image into regions with similar attributes, in order to get more precise process of coloring [2]. Two basic ways of segmentation which are made from structures of data and transformation, are showed for making groups of pixels [3]. Hence, it is always recommended to start with the method of finding edges and separating regions [1],[3].

After the determination of the regions we use half - automatic techniques of segmentation: class Haf transformation and texture segmentation. Proposed methods of coloring process the monochrome image using appropriate colors so that image primarily shows its existence more closely. To achieve desired results, in the process of colorization, the user the most often uses one automatic technique. In that case application of colors is performed without the possibility of maneuvering nuances, which as the result has different final solutions using variable applications. Scientific contributions to this subject already exist. Welsh [4] widened the automatic system for color transfer from one image to another by changing middle variation of the color on the image. Dalmau [5] suggested the colorization method based on color application on gray scale image wherewith the probabilities of the gray pixel limit values are defined. Calculated probability is used for further utilization through the linear function where the color is channeled through α-channel. Levan [6] has perfomed colorization based on assumption of the neighboring pixels in time/space. He was guided by the probability that pixels with similar intensity should have the same colors. Kumar [7] performs colorization based on the color image which is of the similar texture as the gray one. Afterwards, similarity measures of the texture are calculated and compared. After that the transfer of the colors from the colored image to the gray image pixels is performed. Kumar [8] has also performed image colorization through standard deviations of the data.

However, independently of the technique that is used, object possesses two dimensional values which present the raster of image. Data on the image has different values independently from the image format itself. Grey tones on the image are generated through matrix and appropriate color from the palette of color spectrum is gotten that way. By the use of function for combining 2D vector graphics with the image of the object, extremely big manipulation on the image is not allowed. By using half - automatic coloring we achieve great effect, during which function canalling black and white patterns through RGB canal is performed. Processing of the photos will be achived through algorithm and the result of segmentation and coloring of image will be used on the original image of Nikola Tesla's Long Island Laboratory and the results will be shown in the conclusion of this work.

Thanks to the long cooperation with the "Nikola Tesla Museum" in Belgrade on a project supported by the Ministry of Science of the Republic of Serbia, authors had the opportunity to work with original photographs that were created as part of the Museum heritage. The idea to apply the general methods presented in this paper to the original photos that were obtained from the Museum proved to be very successful, so the Museum participated in the project.

[1] Vladan Vuckovic is with Faculty of Electronic Engineering, ul.Aleksandra Medvedeva 14, Niš, Serbia, E-mail: vladan.vuckovic@elfak.ni.ac.rs

[2] Sanja Spasic is PhD student with Faculty of Electronic Engineering, ul.Aleksandra Medvedeva 14, Niš, Serbia, E-mail: sanjadjor@gmail.com

Nowadays, our recolored Tesla images that are in basic presentation setup of Museum in Belgrade. The results of our work were evaluated both through the basic setup of the Museum which has about visitors 40,000 per year, as well as the successful presentations in many international exhibitions that saw hundreds of thousands of people.

The project of the Museum was also the default at the World Exhibition in China 2010 at the stand of the Republic of Serbia. Precise data about these details may be obtained directly from the Nikola Tesla Museum [9].

II. SEGMENTATION

Segmentation is closely related to the process of coloring [1]. The term „segmentation" of the image is related to the group of actions for dividing image into regions with similar attributes. The attribute which is the most often used is lightness with monochrome images or colors with color images. Segmentation of image comes from the theory of Gestalt psychology which studied the making of the groups of pixels according to their shape, similar to the properties of human observing. The outcomes of this theory are two basic aspects of image segmentation:

Data structure - which is needed for studying the qualities of homogenous groups

Transformations - needed for calculating the quality which means two basic kinds of segments are boundaries and regions. This way of segmentation is presented graphically by the knots, where the pixels are the key and present borders (Fig. 1) [3].
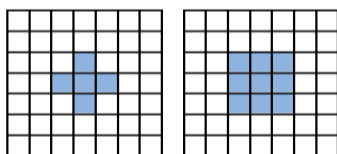
Fig. 1. Classifying regions by boundary pixels.

From here the mathematical definition can be extracted:

Segmentation on of image R is formally defined as separating into not overlapping regions R_i , where are i=1,2,3,....,N. is metrical, not empty, sets.
If P(Ri) is the indicator of uniform of all elements in set R_i, end Ø empty set, then it means:
$U_{Ri} = R_{ni}=1$
For each i and j, for which means i≠j, valves $R_i \cap R_j = \emptyset$
For each i=1,2,..N, must be $P(R_i)=1$
For all pairs i≠j, means $(R_i \cup R_j)=0$. For each i=1,2,..N, R_i is connected region.

This definition is of great importance because according to it the division of image at regions according to the similar attributes is done, as well as combining of the same (in monochrome image: color, lightness, edges, measures for textures, etc.) pixels.

Although the segmentation represents the most important phase in the image analysis, there was only theoretical base for segmentation until nowadays. Most actions of

segmentation which are accepted in practice have neuron character. Apart from this, there is no way for quantitative estimation of how good the action of segmentation is. Hence, starting points are the basic techniques of segmentation for separating (Fig. 2):

Finding the edges- separating the pixels which belong to the rims of the objects,

Separating the regions – separating the whole region is object from the background of the joining pixels which lightness is below the border of background, and the rest to the object and reverse.
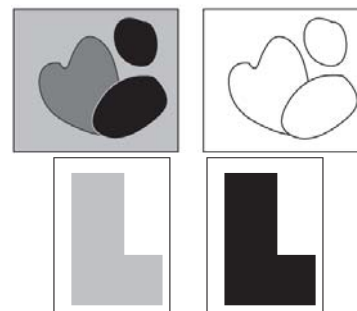
Fig. 2. a) Separating edges; b) Separating object from background.

This kind of use of general techniques is simply not enough, because it would lead to imprecise process of segmentation. Therefore, more precise technique is needed to provide more serious results. The process of combining is specific and very complex. It is needed to coordinate techniques for combining with basic foundation where it is applied. Considering that manual colorization process can take a long time, and automatically can have unwanted results in tuning of colors and impossibility of precise segmentation as a result of the damaged photo – the most precise way of colorization is semi-automatic colorization.

III. THE ALGORITHM FOR SEMI-AUTOMATIC COLORIZATION

Figure shows data-flow diagram of our new algorithm with complete colorization phases (Fig. 3):
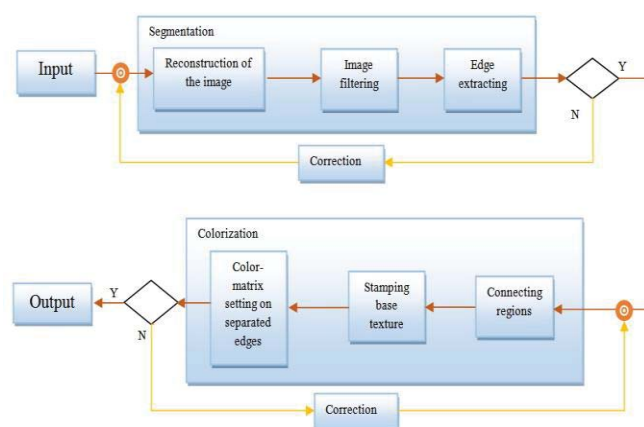
Fig. 3. The algorithm for semi-automatic colorization.

The proposed algorithm consists of the following steps:

Step 1. Upload an image.

Step 2. If damages exist, reconstruct damaged regions of the image. The reconstruction is performed directly. In that case, artificial "cloned" pixels must be generated. The pixel is copied from the closest neighboring, healthy pixel, where the region's borders are very helpful in order not to step out from the mentioned regions.

Step 3. *Image filtering* of scab in the picture, haziness, turbidity. Filters such as: *Dust&Scratches*, *Fibers* and *Sharpe*, must be applied. This way the damage of the photo is minimized. With the help of this tool, 2D matrix which is formed from the basics of the photo is obtained. The product of every damaged pixel with the good pixel on the photo is summarized, and that way resulting pixels which perceive color value of the basic pixel are obtained.

Step 4. *Edge detection* - extracting contours of the image. The edges very often detect important information such as the location of a discontinuity among the shades, color and texture [10]. Moreover, the edge of the figure represents the boundary or contour, where significant changes in some of the physical aspects of the image occur, such as surface reflections, lighting, or distance to the visible surface of the observers. Changes in the physical aspect of the image are manifested in the form of changes in the intensity, color and structure of the image. These edges enrich information that refers to the importance of the object recognition and detection of the edges themselves and refers itself to the process of identification of sharp discontinuities in the analysis.

Step 5. *Finding and connecting regions* with the same basic statistical attributes and *marking the region* by tagging pixels after achieving binary files. The simplest method in this case is to join each pixel tags or index segment. There is one more effective method, which is found in the fact that the closed contour that limits a region is specified, but after that each pixel within the contour must be marked.

Step 6. The choice of texture as a basis for the matrix colorization of individual objects is done by comparing the segmentation method with a threshold. Textures RGB measures and color histogram are calculated for any pixel in a region automatically with following procedures (C++ code):

```
private void panelRed_Paint(object sender, PaintEventArgs e)
    {
base.OnPaint(e);
using (Graphics g = e.Graphics)
        {
int[] hist = this.texture.GetRedHistogram();
int maxHistValue = this.texture.GetMaxHistValue();
this.drawHistogram(g,          ((Panel)sender).ClientRectangle,
Color.FromArgb(255, 0, 0), hist, maxHistValue);
        }
    }
privatevoid panelGreen_Paint(object sender, PaintEventArgs e)
    {
base.OnPaint(e);
using (Graphics g = e.Graphics)
        {
int[] hist = this.texture.GetGreenHistogram();
int maxHistValue = this.texture.GetMaxHistValue();
this.drawHistogram(g,          ((Panel)sender).ClientRectangle,
Color.FromArgb(0, 255, 0), hist, maxHistValue);
        }
    }
privatevoid panelBlue_Paint(object sender, PaintEventArgs e)
    {
base.OnPaint(e);
using (Graphics g = e.Graphics)
        {
int[] hist = this.texture.GetBlueHistogram();
int maxHistValue = this.texture.GetMaxHistValue();
this.drawHistogram(g,          ((Panel)sender).ClientRectangle,
Color.FromArgb(0, 0, 255), hist, maxHistValue);
        }
    }
privatevoid drawHistogram(Graphics g, Rectangle bounds, Color color, int[] histogram)
    {
for (int i = 0; i < histogram.Length; i++)
        {
int height = (int)(0.95f * (float)bounds.Height * histogram[i] / this.maxHistValue + 0.5f);
int width = 3;
int x = i * (width + 1);
int y = bounds.Height - height;
g.FillRectangle(newSolidBrush(color), x, y, width, height);
```

Part of the function belonging to this code calculates and shows color histogram for the starting texture (separating RGB component processing). This algorithm automatically tunes the color components of the manually segmented image. For the purposes of color pre-tuning for these procedures, we use original photo of Tesla's Long Island Laboratory, front view (Fig. 4). This complex, high resolution, original grayscale image has all color components that could be use further on, for automatic colorization.
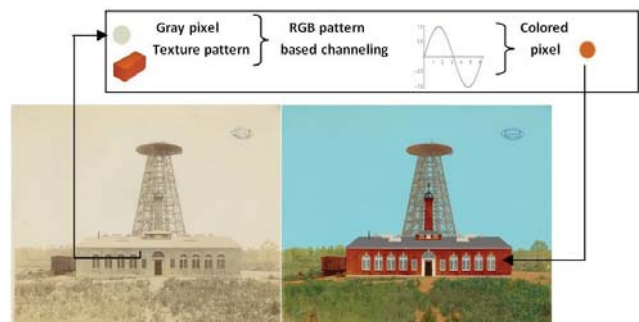
Fig. 4. The algorithm for semi-automatic colorization.

Step 7. Arranging the basic photo, the borders of the photo are separated and that is the base from which we obtain separate regions. After that the color is applied, based on the patent texture which we chose for that region. After the infliction of the colors, the effects (Multiply, Soft Light, Color) are applied because of the authenticity of the image.

Step 8. Upon completion of the question: Are we satisfied with image colorization? *The process is repeated starting with the connecting regions.*

All phases in our proposed algorithm, from original grayscale to colored version are shown on standard test image (pepper_test) (Fig.5):
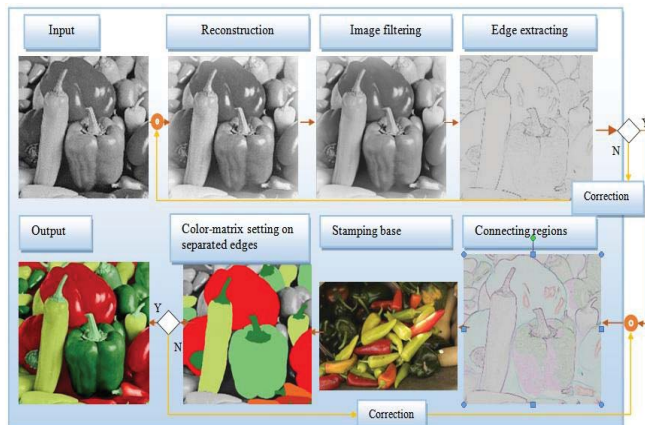
Fig. 5. Standard test image processed by the algorithm for semi-automatic colorization.

Also, there is the deviation using different color environments. Mathematical differences (PSNR values) are negligible, but in visual form that differences are becoming highly visible and that is shown in the Fig. 6 (pepper_test):
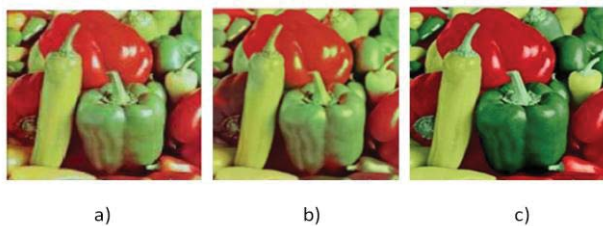
Fig. 6. Comparing of the experimental results of the colored images through PSNR values: a) colored image in RGB environment; b) colored image in YCbCr environment; c) colored image with pattern channeling in RGB environment.

Different experiments show that our method highly depends on manual (human) side. Nevertheless, if the operator is experienced, algorithm clearly improves technique of the colorization, as well as shorter time interval for the image processing itself in comparison to other colorization techniques.

IV. CONCLUSION

In this paper we presented hybrid procedure for recoloring the original black and white photos. We used original procedures and recursive algorithms for segmentation of the grayscale 2D images as well as half-manual colorization. First we performed the photo analysis and did necessary reconstruction of the photo based on copied pixels addition. After that, in the process of the segmentation, the regions were separated and by incorporating procedures the photo was prepared for colorization. Then, we chose original patents and textures, and with the help of the algorithm shown by pseudo-code, we implemented those modifications in the colorization

process. In this way we showed method of coloring in very successful way, where the risk of making mistakes is decreased, because it decreases modifying between the classes of withdrawal. It defines preciseness and exactness which should be appropriate to time it dates from. This direction of research can further be improved in direction of coloring SBS stereo images creating the 3D animation. Based on our insights, an automatic animation algorithm can be developed, giving life to basic original Tesla's images. These methods contribute preserving legacy which can be used as basic foundation for further research.

REFERENCES

[1] Haralick, R.M., Shapiro, L.G. : Image Segmentation Techniques. Computer Vision, Graphics and Image Processing, Vol. 29, No. 1, pp. 100-132, 1985.

[2] Leo Grady and Eric L. Schwartz: Isoperimetric Graph Partitioning for Image Segmentation. IEEE transactions on pattern analysis and machine intelligence, vol. 28, no. 3, pp. 469–475, 2006.

[3] Fu, K. S., Mui, J. K.: A survey on image segmentation. Pattern Recognition, 13, 3-16, 1981.

[4] Welsh T., Ashikhmin M., Mueller K. Transferring color to grayscale images. ACM Transactions on Graphics 21, p.p. 277–280, 2002.

[5] Oscar Dalmau-Cedeno, Mariano Rivera, Pedro P. Mayorga Computing the α-channel with probabilistic segmentation for image colorization. 2007 IEEE 11th International Conference on Computer Vision, 2007. http://www.cimat.mx/~mrivera/conference/dalmau_colorization_icv07.pdf

[6] Levin A., Lischinski D., Weiss Y. Colorization using Optimization. ACM Transactios on Graphics 23, p.p. 289–694, 2004.

[7] Smriti Kumar, Deepak Singh: Colorization Of Gray Image In Lαβ Color Space Using Texture Mapping And Luminance Mapping. International Journal Of Computational Engineering Vol. 2; Issn 2250-3005; p.p. 1272-1278, 2012.

[8] Smriti Kumar, Ayush Swarnkar: Colorization of Gray Scale Images in YCbCr Color Space Using Texture Extraction and Luminance Mapping; Journal of Computer Engineering (IOSRJCE); Vol. 4; Issn: 2278-0661, p.p. 27-32, 2012.

[9] Museum of Nikola Tesla in Belgrade, Serbia: Evaluation data about the project of Tesla images recolorization could be obtained from http://www.nikolateslamuseum.org/web/index.php?l=en

[10] Jianbo Shi , Jitendra Malik: Normalized Cuts and Image Segmentation. IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 8, 2000.

The performance of the modified AGC algorithm for the quality improvement of low-contrast images

Zoran Milivojević[1] and Nataša Savić[2]

Abstract –**The first part of the paper analyzes the AGC algorithm for the improvement of the visual quality of low-contrast images. It has been shown that the quality of images with low luminance and low contrast, after AGC correction is good. However, the quality of images with high luminance and low contrast, after AGC correction is not good. In the second part of the paper, the author has modified the AGC algorithm (MAGC) to improve the visual quality of high-luminance images. The efficiency of the MAGC algorithm was tested experimentally. A comparative analysis of the statistical parameters of images corrected by the AGC and MAGC algorithms shows the effectiveness of the MAGC algorithm. All test images with high luminance and low contrast, after being corrected applying the MAGC algorithm, can be classified as good quality images (GQ images), according to statistical parameters.**

Keywords – **Gama correction, contrast improvement, AGC algorithm.**

I. INTRODUCTION

Digital image processing algorithms can, among other things, perform visual image quality improvement. Improved image quality includes image corrections, such as: luminance, color and saturation correction, sharpness, contrast enhancement, edge and contour emphasizing etc. A number of algorithms have been suggested for improving image contrast [1] - [3]. The algorithms are used intensively, where the corrections are performed by analyzing the histogram of the image, that is, histogram equalization (HE) [4]. HE algorithms have a small numerical complexity. However, HE algorithms do not always give satisfactory results, because they can cause excessive improvement in pixels with luminance intensities that are often repeated, which results in a decrease in contrast intensity with less frequent occurrence [3]. In order to alleviate the problems of excessive increase or decrease in contrast, the following algorithms have been proposed: a) BBHE (bi-histogram equalization algorithms), b) DSIHE (dualistic sub-image histogram equalization algorithms) [5], c) MMBEBHE (minimum mean brightness error bi-histogram equalization algorithms) [6].

In [7] a simple algorithm which uses image classification as well as an appropriate method for each type of image has been proposed. The main goal of the proposed algorithm is to transform the image into a visually high-quality image. This is achieved by: a) increasing contrast and b) correction of luminance. The algorithm is based on adaptive gamma correction (AGC), where the luminance transform function is dynamically determined depending on the input image feature (mean value μ, standard deviation σ).

In [8] the performance of the AGC algorithm for correcting low-contrast images has been determined. The quality of the test image before and after correction with the AGC algorithm was tested. The testing was performed by: a) objective and b) subjective methods. The objective methods (MSE, PSNR, AMBE and SSIM) included a comparative analysis of AGC corrected by good quality (GQ) images. In addition, the image contrasts were analyzed using the following measures: a) RMS (Root-Mean-Square) and b) Ed (Discrete entropy). The subjective picture quality scores were obtained using the MOS (Mean opinion score) method. A detailed analysis has shown that the quality of the AGC corrected images of low contrast and low luminance is good, while the quality of AGC corrected images of low contrast and high luminance is unsatisfactory.

In this paper, the authors have proposed a modification of the AGC algorithm (MAGC algorithm). The modification was performed in the part concerning the correction of high-luminance images ($\mu > 0.5$) and low contrast ($\sigma > 0.1$). The MAGC algorithm was described in detail. After that, an experiment was performed in order to test the efficiency of AGC and MAGC algorithms. An experimental data-base composed of high-luminance images, $\mu = \{0.6, 0.7, 0.8, 0.9\}$ and low-contrast images $\sigma = \{0.02, 0.08\}$ was created. The tables and graphs show the contrast measures (RMS, Ed) and the measures of comparison of MAGC corrected images and GQ images (MSE, PSNR, AMBE and SSIM). The detailed analysis of the results indicates the high quality of MAGC corrected images, which, according to the values of the statistical parameters (μ, σ), are classified as GQ images.

The paper is further organized as follows: Section II describes the AGC and MAGC algorithms; in Section III, the experimental results are presented and the comparative analysis is performed, Section IV is the conclusion.

II. ALGORITHMS FOR REPAIRING A VISUAL QUALITY OF THE IMAGE WITH A SMALL CONTRAST

A. AGC algorithm

In [7], the AGC algorithm for improving the visual image quality has been described. The AGC algorithm, based on the image histogram, determines the mean μ and standard deviation σ. The AGC algorithm first classified the dark ($\mu < 0.5$) and bright ($\mu \geq 0.5$) images. After that, a classification of low-contrast images ($\sigma \leq 1/3\tau$) and high-contrast images ($\sigma > 1/3\tau$) was performed, where $\tau = 3$ [7]. Finally, the parameter γ was calculated to improve the image quality using γ-correction: $I_{out} = c\, I^{\gamma}$, where I is a low-contrast image, I_{out} is γ-corrected image, and c is the transformation parameter. In [9], an

[1]Zoran Milivojevic is with the College of Applied Technical Sciences Nis, Aleksandra Medvedeva 20, Nis 18000, Serbia, E-mail: zoran.milivojevic@vtsnis.edu.rs,

[2]Natasa Savic is with the College of Applied Technical Sciences Nis, Aleksandra Medvedeva 20, Nis 18000, Serbia, E-mail: natasa.savic@vtsnis.edu.rs.

experiment is described in which GQ images are determined by subjective methods (MOS test). The analysis of statistical histogram parameters for GQ images shows μ ≈ 0.5 and σ > 0.1. These results match the results from [7] for high-contrast images.

In [8], the performance of the AGC algorithm has been determined using a) objective and b) subjective measures. The objective testing was carried out using contrast measures (RMS and Ed), as well as the comparative analysis measures (MSE, PSNR, AMBE and SSIM). These measures were obtained by comparing the tested images with the GQ image. The subjective tests were carried out using the MOS test. Fig. 1.a presents test image Lena with low contrast (MOS = 1.364) and its histogram (fig.1.b) ($\mu = 0.2$, $\sigma = 0.02$). By applying the AGC algorithm to the test image (fig. 2.a), a corrected image was obtained (fig. 1.c), with MOS = 3.29, whose histogram is presented in fig. 1.d ($\mu = 0.4943$, $\sigma = 0.1351$). According to the statistical parameters and MOS scores, it has been concluded that the AGC corrected image belongs to the class of GQ images.
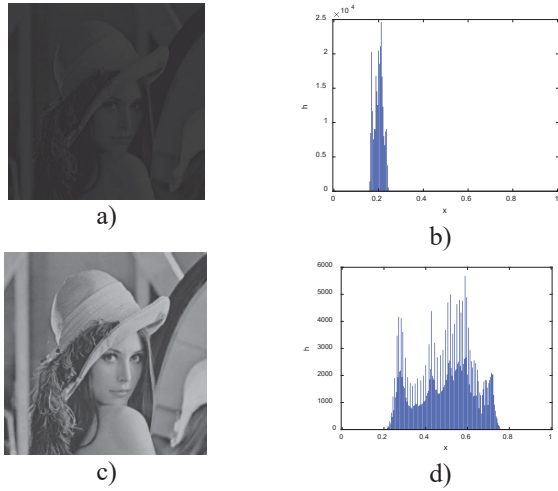
Fig. 1. a) Dark test image Lena with low contrast, b) histogram, c) AGC corrected image, and d) corrected image histogram.

Fig.2.a shows a bright test image Lena with low contrast (MOS = 1.148) and its histogram (fig.2.b) ($\mu = 0.7$, $\sigma = 0.02$). By applying the AGC algorithm to the test image (fig. 2.a), a corrected image was obtained (fig. 2.c), with MOS = 1.034, whose histogram is shown in fig.2.d ($\mu = 0.1350$, $\sigma = 0.0216$). According to the statistical parameters and MOS scores, it has been concluded that the image corrected by AGC is poor in visual quality, and does not belong to the class of GQ images. The results from [8] show that AGC algorithm improved the dark images with low contrast very well, while the improvment of visual quality for bright, low-contrast images was poor. The efficiency, measured by subjective MOS scores, was higher in the dark images compared to the bright ones, and this was MOSdark / MOSbright = 1.4 times. According to the contrast analysis, as one of the objective measures, it could be concluded that the efficiency of the AGC algorithm for dark images was higher RMSdark / RMSbright = 3.74 times compared to the bright images.
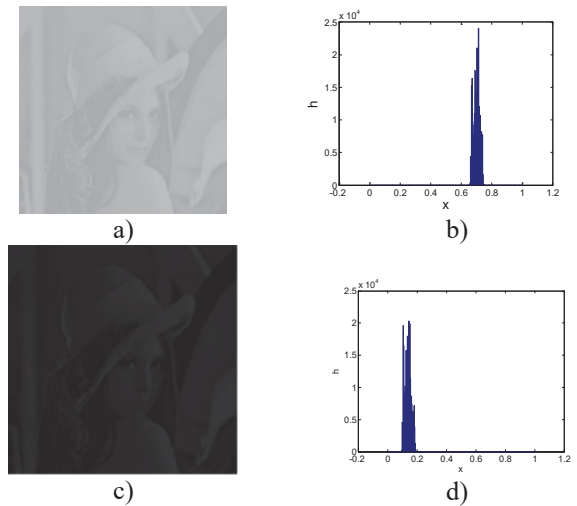
Fig. 2. a) Bright test image of Lena with low contrast, b) histogram, c) AGC corrected image, and d) corrected image histogram.

The authors of this paper have modified the AGC algorithm (MAGC) to repair the visual quality of the images with low contrast and high brightness.

B. MAGC algorithm

The modified AGC algorithm (MAGC) for visual improvement of the images with low contrasts is realized in the following steps:

Input: I_{MxN}
Output: Y_{MxN}
Step 1: Calculating the average brightness:

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j}^{N} I_{i,j}, \tag{1}$$

Step 2: Calculating the standard deviation:

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j}^{N}\left(I_{i,j}-\mu\right)^2}, \tag{2}$$

Step 3: Calculating the γ-factor depending on the contrast:

$$\gamma = -\log_2\left(\sigma\right), \tag{3}$$

Step 4: Classifying the *I* image according to the brightness:

   IF μ<0.5

Step 5: Transforming the dark image *I*:

$$Y = \frac{I^\gamma}{I^\gamma + (1-I^\gamma)\times\mu^\gamma}, \tag{4}$$

   ELSE

Step 6: Transforming the bright picture *I*:

$$I=1-I, \tag{5}$$

$$Y = \frac{I^\gamma}{I^\gamma + (1-I^\gamma)\times\mu^\gamma}, \tag{6}$$

$$Y = |1 - Y|. \tag{7}$$

   END

The MAGC modification in relation to the AGC algorithm is in step 6 (Eq-s. 5, 6, 7). A detailed verification of the efficiency of the MAGC algorithm was performed experimentally.

III. EXPERIMENTAL RESULTS AND THE ANALYSIS

In order to test the efficiency of the MAGC algorithm for improving the visual quality of bright low contrast images, an experiment was performed.

A. Experiment

The experiment was realized within an objective quality testing of low-contrast and high-luminance images was done. It was shown that the visual quality was increased by: a) AGC [7] and b) MAGC (Section II.B) algorithms. The visual quality improvement involved increasing the image contrast. The quality testing of MAGC corrected images was performed using a comparative analysis of the contrast measures (RMS and Ed) and objective measures (MSE, PSNR, AMBE and SSIM) with a) GQ image [9] and b) AGC corrected images [8].

B. Base

A database of the test images was created using the images obtained by modifying the luminance and contrast of the original image. The original test image Lena was used to create low-contrast images with high luminosity ($\mu > 0.5$): I_1 ($\mu = 0.9$, $\sigma = 0.02$), I_2 ($\mu = 0.9$, $\sigma = 0.08$), I_3 ($\mu = 0.8$, $\sigma = 0.02$), I_4 ($\mu = 0.8$, $\sigma = 0.08$), I_5 ($\mu = 0.7$, $\sigma = 0.02$), I_6 ($\mu = 0.7$, $\sigma = 0.08$), I_7 ($\mu = 0.6$, $\sigma = 0.02$) i I_8 ($\mu = 0.6$, $\sigma = 0.08$) (Table I).

C. Results

Table I presents the statistical parameters (μ, σ) and contrast measures (RMS, Ed) of the test images. Table II presents the parameters and contrast measures for the images after applying the AGC algorithm [8]. Table III presents the parameters and contrast measures for the images after applying the MAGC algorithm. The visual effect of applying AGC and MAGC algorithms on the test image I_5 is presented in fig. 3. Figure 4 presents the contrast measures RMS (fig. 4a) and Ed (fig. 4b). Figure 5 presents the objective measures MSE (fig. 5.a), PSNR (fig. 5.b), AMBE (fig. 5.c) and SSIM (fig. 5.d) graphically.

TABLE I

STATISTICAL PARAMETERS AND CONTRAST MEASURES OF TEST IMAGES.

| Image | μ | σ | Contrast | |
			RMS	Ed
I_1	0.9	0.188	$4*10^{-4}$	4.2382
I_2	0.9	0.08	0.0064	5.7216
I_3	0.8	0.02	$4*10^{-4}$	4.2291
I_4	0.8	0.08	0.0064	6.1898
I_5	0.7	0.02	$4*10^{-4}$	4.2382
I_6	0.7	0.08	0.0064	6.1906
I_7	0.6	0.02	$4*10^{-4}$	4.2291
I_8	0.6	0.08	0.0064	6.1898

TABLE II

STATISTICAL PARAMETERS AND CONTRAST MEASURES OF TEST IMAGES AFTER APPLYING AGC ALGORITHM [8].

| AGC Image | μ | σ | Contrast | |
			RMS	Ed
Y_1	0.5553	0.0692	0.0048	5.9884
Y_2	0.7070	0.2230	0.0497	6.7687
Y_3	0.2861	0.0401	0.0016	5.2212
Y_4	0.4648	0.1640	0.0269	7.1261
Y_5	0.1350	0.0216	$4.6594*10^{-4}$	4.3377
Y_6	0.2897	0.1159	0.0134	6.6497
Y_7	0.0568	0.0106	$1.1172*10^{-4}$	3.3653
Y_8	0.1687	0.0779	0.0061	6.0793

TABLE III

STATISTICAL PARAMETERS AND CONTRAST MEASURES OF TEST IMAGES AFTER APPLYING MAGC ALGORITHM.

| MAGC Image | μ | σ | Contrast | |
			RMS	Ed
Y_1	0.5181	0.2354	0.0554	7.2789
Y_2	0.5470	0.3714	0.1380	6.4852
Y_3	0.5063	0.1340	0.0180	6.8524
Y_4	0.5338	0.2797	0.0783	7.1644
Y_5	0.5027	0.0919	0.0084	6.3952
Y_6	0.5191	0.2161	0.0467	7.2841
Y_7	0.5003	0.0700	0.0049	6.0001
Y_8	0.5053	0.1751	0.0307	7.2307

a) I_5 b) AGC I_5 c) MAGC I_5
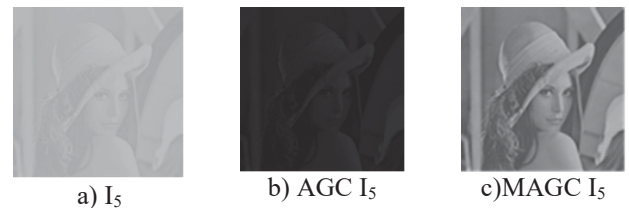
Fig. 3. Examples of images from the test base (I_5): a) the original image, b) the image after AGC correction, and c) the image after MAGC correction.
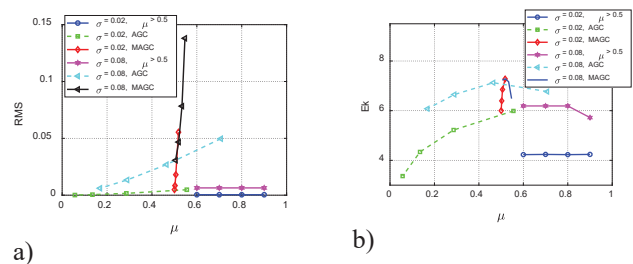
Fig. 4. Contrast measures: a) RMS and b) E_d.

a)   b)
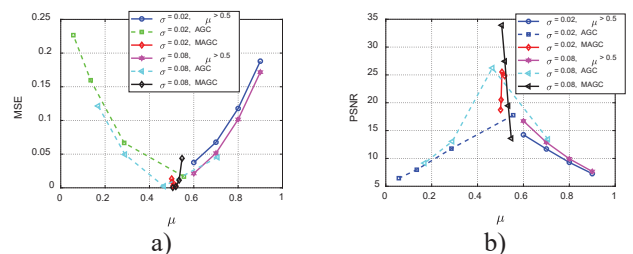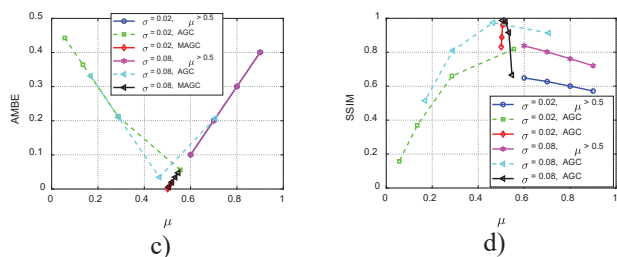
Fig. 5. Objective measures of the image quality: a) MSE, b) PSNR, c) AMBE and d) SSIM.

IV. Conclusion

The paper has been described the AGC algorithm for improving the quality of images with low contrast. A detailed analysis has shown that the quality of AGC corrected dark images with low contrast are of good quality, and that they can be classified as GQ images. However, the quality of the corrected bright images with low contrast is of unsatisfactory quality. The authors suggested the MAGC algorithm obtained by modifying the AGC algorithm. The experimental analysis of the application of the MAGC algorithm on the base of the high-brightness and low-contrast test images, showed that the visual quality of all images from the base, after MAGC correction, was excellent. According to the statistical contrast measures (RMS and E_d) and comparisons with GQ images (MSE, PSNR, AMBE and SSIM), the MAGC corrected images can be classified as GQ images.

D. Analysis of the results

According to the results presented in Tables I, II and III and in Fig. 6 and 7, it has been concluded that:

1. for bright images ($\mu > 0.5$, $\sigma = 0.02$ the increase of the objective parameters in the test images after applying MAGC algorithm: a) $\mu \rightarrow 0.25 / 0.0068 = 36.7647$, b) $\sigma \rightarrow 0.1328 / 0.02 = 6.64$, c) MSE $\rightarrow 0.1028 / 0.0071 = 14.4789$, d) PSNR $\rightarrow 10.6229 / 22.4012 = 0.4742$, e) AMBE $\rightarrow 0.2506 /0.0075 = 33.622$, f) SSIM $\rightarrow 0.6119 /0.9125 = 0.6706$, g) RMS $\rightarrow 4.0000*10^{-04} /0.0217 = 0.0185$ i h) $E_d \rightarrow 4.2336 / 6.6316 = 0.6384$ times,

2. for bright images $(\mu > 0.5$, $\sigma = 0.02)$ the increase of the objective parameters after applying MAGC algorithm in relation to the applying of AGC [8] algorithm: $\mu \rightarrow 0.2693 / 0.0068 = 39.015$, b) $\sigma \rightarrow 0.1328 / 0.0354 = 3.7514$, c), MSE $\rightarrow 0.1174 / 0.0071 = 16.5352$, d) PSNR $\rightarrow 10.9832 /22.4012 = 0.4903$, e) AMBE $\rightarrow 0.2691 / 0.0075 = 35.88$, f) SSIM $\rightarrow 0.5015 / 0.9125 = 0.5496$, g) RMS $\rightarrow 0.0017 / 0.0217 = 0.0783$ i h) $E_d \rightarrow 4.7282 / 6.6316 = 0.71298$ times,

3 for bright images ($\mu > 0.5$, $\sigma = 0.08$) the increase of the objective parameters in the test images after applying MAGC algorithm: a) $\mu \rightarrow 0.25 / 0.0263 = 9.5057$, b) $\sigma \rightarrow 0.2606 / 0.08 = 3.257$, c) MSE $\rightarrow 0.0865 / 0.0143 = 6.0489$, d) PSNR $\rightarrow 11.7980 / 23.6075 = 0.4998$, e) AMBE $\rightarrow 0.2506 / 0.0269 = 9.3073$, f) SSIM $\rightarrow 0.7807 /0.8877 = 0.8795$, g) RMS $\rightarrow 0.0064 / 0.0734 = 0.0872$ i h) $E_d \rightarrow 6.0730 / 7.0411 = 0.8625$ times,

4. for bright images ($\mu > 0.5$, $\sigma = 0.08$) the increase of the objective parameters after applying MAGC algorithm in relation to the applying of AGC [8] algorithm: a) $\mu \rightarrow 0.0199 / 0.0263 = 0.0757$, b) $\sigma \rightarrow 0.2606 /0.1452 = 1.7948$, c), MSE $\rightarrow 0.0547 /0.0143 = 3.8252$, d) PSNR $\rightarrow 15.4755 / 23.6075 = 0.6555$, e) AMBE $\rightarrow 0.1956 / 0.0269 = 7.2714$, f) SSIM $\rightarrow 0.8030 / 0.8877 = 0.9046$, g) RMS $\rightarrow 0.0240 / 0.0734 = 0.32697$ i h) $E_d \rightarrow 6.6559 / 7.0411 = 0.94529$ times.

The obtained results indicate the high efficiency of the MAGC algorithm in relation to the AGC algorithm.

References

[1] H. Cheng, X. Shi, "A simple and effective histogram equalization approach to image enhancement". Digital Signal Process. 14(2), 158–170 (2004.

[2] D. Coltuc, P. Bolon, J-M Chassery, "Exact histogram specification," Image Process. IEEE Trans. 15(5), 1143–1152 (2006).

[3] M. Kim, M. Chung, "Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement," Consum. Electron. IEEE Trans. 54(3), 1389–1397 (2008)

[4] R. Gonzalez, R. Woods, "Digital image processing", ears on /Prentice Hall, Upper Saddle River, NJ, 3rd ed. (2008).

[5] Y. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization". Consum. Electron. IEEE Trans. 43(1), 1–8 (1997).

[6] S. Chen, A. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement." Consum. Electron. IEEE Trans. 49(4), 1310–1319 (2003)

[7] S. Rahman, M. Rahman, M. Abdullah-Al-Wadud, G. Dastegir Al-Quaderi, M. Shoyaib, "An adaptive gamma correction for image enhancement", EURASIP Journal on Image and Video Processing, 2016 2016:35.

[8] N. Savić, Z. Milivojević, "Performanse adaptivnog Gama korektora za popravku kvaliteta slika sa malim kontrastom'', XVIII međunarodni simpozijum INFOTEH-JAHORINA 2019.

[9] Z. Milivojević, N. Savić, B. Prlinčević, "Parametri slike dobrog kvaliteta pogodnih za digitalnu obradu", Međunarodno savetovanje na temu Upravljanje znanjem i informatika, 8-9 Januar, Kopaonik, 2019.

Robustness of the SD watermarking algorithm for γ-corrected images with superimposed Gaussian noise

Bojan Prlinčević[1] and Zoran Milivojević[2]

*Abstract –***First part of the paper describes the algorithm for γ-correction, algorithm for superposition of Gaussian noise and the algorithm for inserting a watermark based on Schur decomposition (SD algorithm). In the second part of the paper, an experiment was described in which the insertion and extraction of the digital watermark in the image was performed. Gaussian noise is superposed to watermarked image, after which it is deformed by a change in contrast using an γ-correction algorithm. The γ coefficient is varied in range γ = 0.25 ÷ 4. A watermark is extracted from the γ-corrected image. The quality control of an extracted watermark was performed using objective quality measure-MSE, and a visual comparison of the extracted and original watermark.**

Keywords – **Schur decomposition, SD algorithm, γ-correction, Gaussian noise, watermark.**

I. INTRODUCTION

The widespread use of digital media makes it possible to manipulate the images. Image manipulation (filtering, contrast changes, re-scaling ...) are used to improve visual image quality [1]. The most widespread image manipulation known as γ-correction or γ-transformation is used to change the contrast of the image and is determined by a single parameter [2]. On other side manipulation with the images can lead to damages which are caused by noise. In the process of digital image processing (transmission, acquisition, equipment characteristic ...) images can be corrupted with Impulse noise and Gaussian (AWGN) noise [2], [3]. Salt and pepper replaces certain pixel images with random values. The AWGN, affects on the entire set of pixels of an image. AWGN noise afects pixels randomly between the minimum and the maximum value of the pixel [4], resulting in the fact that the image looks hazy and blurred. At small noise densities of AWGN, the image restoration can be satisfactorily corrected with gamma correction.

In the process of digital image processing, another very important feature is the protection of copyrights. Copyright protection is best done by invisible watermarks. An invisible watermark is inserted into the image in order not to destroy visual characteristics of the image and may be later extracted for the purpose of proving copyrights [5], [6]. The imperfection of invisible watermark is that they can be insufficiently robust and resistant to disturbances.

[1]Bojan Prlinčević is with the Higher Technical School Zvecan, St. Nusiceva No 6, 38225 Zvecan, Serbia, E-mail: b.prlincevic@vts-zvecan.edu.rs.

[2]Zoran Milivojević is with the College of Applied Technical Sciences, St. A. Medvedeva No 20, Nis, Serbia, E-mail: zoran.milivojevic@vtsnis.edu.rs.

The authors of this paper set the question: How does γ-correction affect an extracted watermark, from a noised (with AWGN) watermarked image? In order to obtain the answer, authors conducted an experiment in which: a) a good quality (GQ) image [7] was formed, b) watermark (with the SD algorithm [8], [9], [10]) in GQ image was inserted, c) Gaussian noise with SNR = [10:2:30] on the watermarked image was superimposed, d) watermarked image was deformed by contrast change using γ-correction (γ = 0.25 ÷ 4) [11], [12], and e) watermark was extracted and its quality checked. The quality of the extracted watermark was checked using objective measurements of the quality: Mean Square Error (MSE), as well as visual comparison of the extracted and original watermark.

The paper is organized in the following way. Section II describes: SD algorithm for inserting and extracting watermark and algorithms for γ-correction and superposition of Gaussian noise. Section III describes an experiment. Section IV presents results and analysis of the results. Section V is a conclusion.

II. ALGORITHMS

The SD algorithm based on Schur decomposition [9], [13] was used for inserting a digital watermark. The algorithm is executed in the next steps:

Input: original image $A_{M \times N}$, binary watermark $W_{Mz \times Nz}$, block dimensions $M_b \times N_b$.

Output: watermarked image X_W.

Step 1: Original matrix is divided on $X \times Y$ blocks, $H_{Mb \times Nb}$, where $X = \lceil M / M_b \rceil$ and $Y = \lceil N / N_b \rceil$.

Step 2: Applying Schur decomposition on blocks H:

$$H_{i,j} = U_{i,j} \times D_{i,j} \times U_{i,j}^T , \qquad (1)$$

where U is unitary matrix, D upper triangular matrix and $1 \leq i \leq \lceil M / M_b \rceil$ and $1 \leq j \leq \lceil N / N_b \rceil$.

Step 3: The elements $u_{2,1}$ and $u_{3,1}$ in each block of matrix U are modified in order to obtain a modified block U', according to the information about the embedded binary watermark W. The insertion of the digital watermark is performed in accordance with the rule shown in (2) and (3), by modifying the elements $u_{2,1}$ and $u_{3,1}$:

$$\text{if } w_{i,j} = 1, \begin{cases} u'_{2,1} = sign(u_{2,1}) * (U_{avg} + T/2) \\ u'_{3,1} = sign(u_{3,1}) * (U_{avg} - T/2) \end{cases} \qquad (2)$$

$$\text{if } w_{i,j} = 0, \begin{cases} u'_{2,1} = sign(u_{2,1}) * (U_{avg} - T/2) \\ u'_{3,1} = sign(u_{3,1}) * (U_{avg} + T/2) \end{cases} \qquad (3)$$

where *sign(x)* denotes sign of *x,* and $U_{avg} = (|u_{2,1}| + |u_{3,1}|)/2$, $|x|$ denotes absolute value of *x.*

Step 4: Reconstruction of the block with inserted watermark:

$$H'_{i,j} = U_{i,j} \times D'_{i,j} \times U^T_{i,j} , \qquad (4)$$

Step 5: Forming watermarked image A_w from blocks H'.

SD algorithm for watermark extraction is performed in following steps:

Input: watermarked image A_w, blocks dimensions $M_b \times N_b$.

Output: Reconstructed binary watermark $W'_{Mz \times Nz}$.

Step 1: Matrix (A_w) dividing is performed on $X \times Y$ blocks, $H'_{Mb \times Nb}$, where $X = \lceil M / M_b \rceil$ and $Y = \lceil N / N_b \rceil$.

Step 2: Applying Schur decomposition on blocks H':

$$H'_{i,j} = U'_{i,j} \times D'_{i,j} \times (U'_{i,j})^T , \qquad (5)$$

where U' denotes unitary matrix, D' upper triangular matrix and $1 \le i \le \lceil M / M_b \rceil$ and $1 \le j \le \lceil N / N_b \rceil$.

Step 3: Extracting one bit, *bw'*, of watermark from matrix D':

$$w'_{i,j} = \begin{cases} 0, & if \quad |u'_{2,1}| < |u'_{3,1}| \\ 1, & if \quad |u'_{2,1}| \ge |u'_{3,1}| \end{cases} \qquad (6)$$

Step 4: Forming watermark W' from extracted bits $w_{i,j}'$.

Deforming image, by changing the contrast, using γ-correction by doing γ-transformation of an image, to the desired mean luminance value μ.

$$A_\gamma = A^\gamma . \qquad (7)$$

where A denotes original image and A_γ denotes γ-corrected image.

Gaussian noise is superimposed to the image:

$$A_{N(i,j)} = A_{(i,j)} + \sigma * N . \qquad (8)$$

where *A*-denotes original image and A_N-image with superimposed Gaussian noise, *N*- denotes Gaussian noise and σ-variance.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experiment

For the purpose of testing the effect of γ-correction on the digital watermark inserted using SD algorithm, the following experiment is performed:

Step 1: On the original image *A* γ-correction is applied so as to obtain: a) $A_{0.5}$ - GQ image, ($\mu = 0.5$); b) $A_{0.25}$ - darken image, ($\mu = 0.25$); and c) $A_{0.75}$ - lighten image, ($\mu = 0.75$).

Step 2: Image with the watermark A_w is obtained as it is in original image *A,* which is divided onto blocks $M_b \times N_b = 4 \times 4$, applying Schur decomposition inserted binary watermark *W*, dimensions $M_w \times N_w = 128 \times 128$, with insertion coefficient *T.* In every block one bit of the watermark is inserted.

Step 3: Gaussian noise (SNR=[10:2:30]) is superimposed to the image with inserted digital watermark $A_{\mu W}$,

Step 4: Watermarked image $A_{\mu W}$ is γ-corrected, with γ-coefficient varied in range $0.25 \div 4$. The watermarked γ-corrected image is obtained $A^\gamma_{\mu W}$.

Step 5: Watermark W_e is extracted from γ-corrected image $A^\gamma_{\mu W}$.

As a quality measure of extracted watermark, a visual comparison of the extracted watermark and original watermark, and MSE are used:

$$MSE = \frac{\sum_{ij} (x_{ij} - y_{ij})^2}{M \times N} , \qquad (9)$$

where: x_{ij}, y_{ij}, i,j-the pixel of original and proceeded elements of the image, $M \times N$-dimensions of the image, d - maximal value of pixel.

Images Lena, Girl and Babon dimensions (512 × 512), fig. 1-a,b,c are used in experiment. As watermark is used image fig. 1-d dimension (128 × 128). The velues of coefficient insertion are varied, $T = \{0.0025, 0.005, 0.0075, 0.01, 0.015, 0.02, 0.025, 0.03\}$.

a)   b)

c)   d)
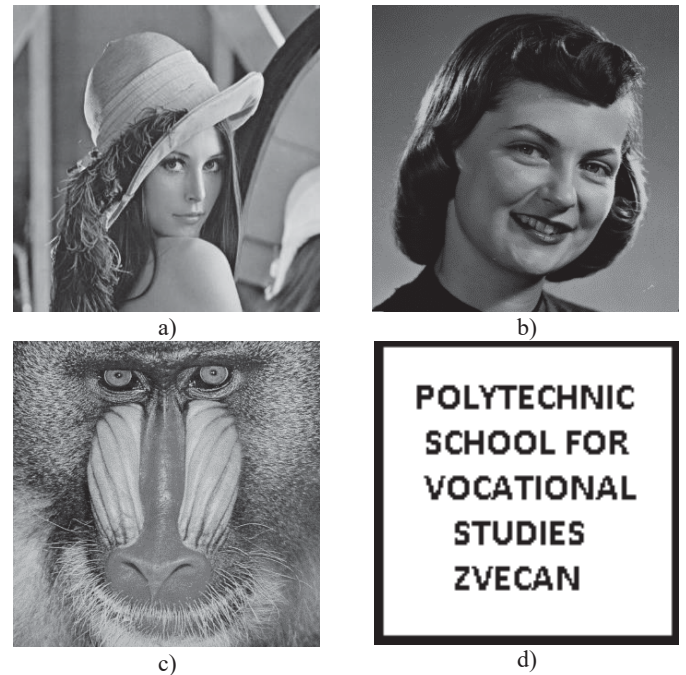
Fig. 1. Images used in the paper: a) Lena, b) Girl, c) Baboon and d) Watermark

B. Results

Appearance of the GQ image after watermark insertion ($T = 0.01$) and superimposed noise with $SNR=\{12, 18, 24, 30\}$ is shown in fig. 2.

Fig. 3 shows appearance of extracted watermark, from the image in which it is inserted with $T = 0.01$ (averaged for all values of γ-coefficient) and: a) $\mu = 0.25$ and $SNR = 18$ dB, b) $\mu = 0.25$ and $SNR = 24$ dB, c) $\mu = 0.5$ and $SNR = 18$ dB, d) $\mu = 0.5$ and $SNR = 24$ dB, e) $\mu = 0.75$ and $SNR = 18$ dB, f) $\mu = 0.75$ and $SNR = 24$ dB. Fig. 4. shows diagram for MSE, for watermark extracted from images, for different levels of luminance $\mu = \{0.25, 0.5, 0.75\}$, depending on SNR ($T = 0.01$). Fig. 5. shows diagram for MSE for extracted watermark depending on γ-coefficient ($T=0.01$, $\mu = 0.5$) for value of $SNR=\{12, 18, 24, 30\}$. Fig. 6. shows diagram for MSE for extracted watermark depending on SNR, for three levels of insertion coefficient $T = \{0.0025, 0.01, 0.025\}$ (averaged for all values of γ-coefficient).

Fig. 2. Appearance of images after inserting digital watermark (in GQ image, $\mu = 0.5$ and $T=0.01$) and superimposing Gaussian noise with: a) $SNR = 12$ dB, b) SNR = 18 dB, c) $SNR = 24$ dB, and d) $SNR = 30$ dB.
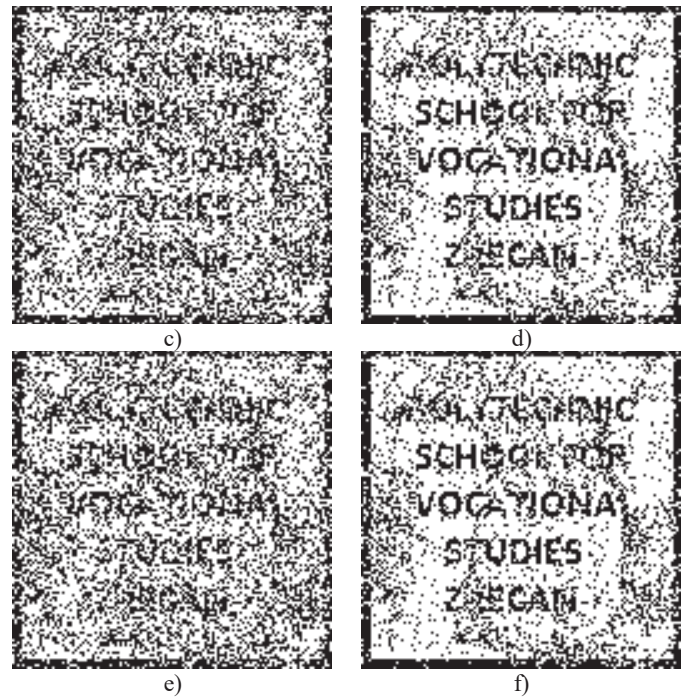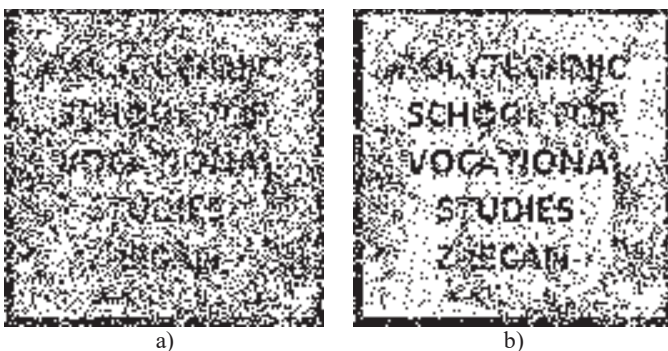
Fig. 3. Appearance of extracted watermark from image Lena ($T = 0.01$) for: a) $\mu = 0.25$ and $SNR = 18$ dB, b) $\mu = 0.25$ and $SNR = 24$ dB, c) $\mu = 0.5$ and $SNR = 18$ dB, d) $\mu = 0.5$ and $SNR = 24$ dB, e) $\mu = 0.75$ and $SNR = 18$ dB, f) $\mu = 0.75$ and $SNR = 24$ dB.
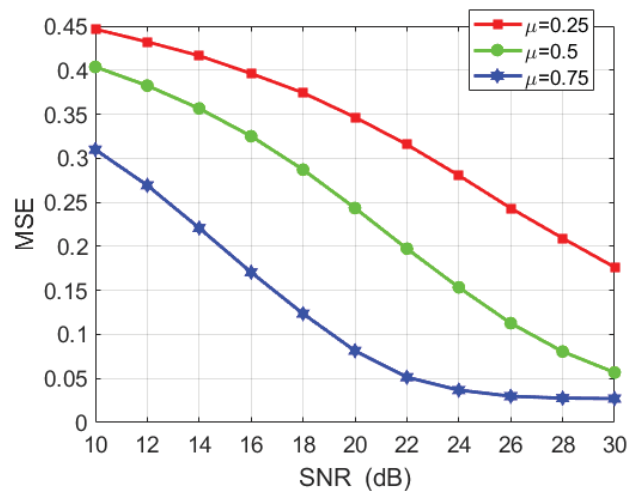
Fig. 4. The MSE of extracted watermark depending of value of SNR for different values of luminance μ ($T = 0.01$).
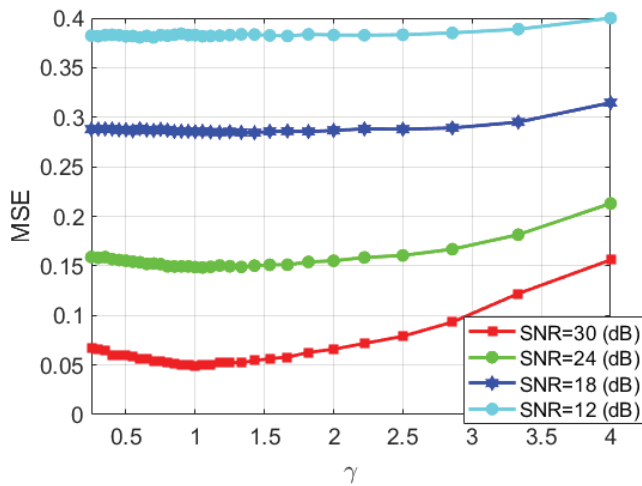
Fig. 5. The MSE of extracted watermark depending of γ-coefficient, for different values of *SNR* (μ=0.5, $T = 0.01$).
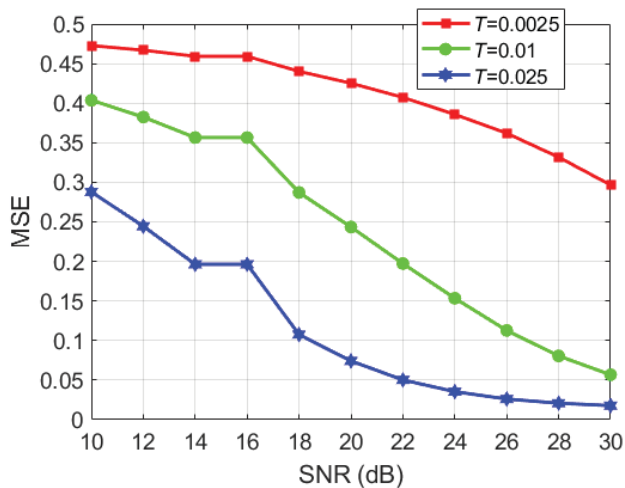
Fig. 6. The MSE of extracted watermark (averaged for all values of γ-coefficient) depending on inserting coefficiente T, (μ=0.5).

C. Analysis

Based on the results shown in fig. 3 and the diagrams shown in figs. 4 - 6. it can be concluded that:

a) The watermark extracted (inserted with $T = 0.01$), from image with luminance $\mu = 0.75$ is with the best appearance. When value of SNR increases, the quality of extracted watermark rises. When ratio of the SNR = 30 dB, the extracted watermark is with the best quality. The extracted watermark is visible at SNR = 18 dB (Fig. 3);

b) The MSE of extracted watermark (inserted with $T = 0.01$) has the smallest image value with luminance $\mu = 0.75$ (lighted image) (fig. 4.);

c) The MSE of extracted watermark (inserted with $T = 0.01$) has the smallest image value when $SNR = 30$ dB. The influence of γ-correction on the MSE extracted watermark is decreasing with decreasing of SNR value. For value $SNR < 18$ dB influence of γ-correction on the quality of extracted watermark is negligible (fig. 5.);

d) The MSE of extracted watermark proportionally decreases with increasing insertion coefficient T (fig. 6.).

IV. CONCLUSION

In this paper testing of influence of the γ-correction on the extracted digital watermark which is inserted in the image is performed, applying SD algorithm, to which Gaussian noise is superimposed. Based on the detailed analysis of the results obtained from the experiment it can be concluded that extracted watermark from the noised image (Gaussian noise) can be visible when value of SNR > 18 dB. Extracted digital watermark from the noised image (Gaussian noise) is with better visible appearance after extraction from image which is preprocessed, to be lighten, and luminance set on value $\mu = 0.75$. With the increasing ratio of *SNR* quality of extracted watermark is getting more visible, and MSE is decreasing with increasing value of *SNR*. With the increase of insertion coefficient quality of extracted watermark is getting better and proportionally increases with increasing value of *SNR*. Based on the above stated it can be concluded that Gaussian noise with relatively small value of SNR significantly influences the quality of extracted watermark.

REFERENCES

[1] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Prentice Hall, 2007.

[2] P. Wang, F. Liu, C. Yang, X. Luo, "Parameter estimation of image gamma transformation based on zero-value histogram bin locations", Signal Processing: Image Communication 64 pp. 33–45, 2018.

[3] T. Rabie, Adaptive hybrid mean and median filtering of high-ISO long-exposure sensor noise for digital photography, J. Electron. Imaging 13 (2) (2004) 264–277.

[4] A. Awad, "Standard deviation for obtaining the optimal direction in the removal of impulse noise", IEEE Signal Process. Lett. 18(7), pp. 407–410, 2011.

[5] I. J. Cox, M.L. Miller and J.A. Bloom, "Digital Watermarking", Morgan Kaufmann Publishers, 2002.

[6] R. Liu, T.Tan, "A SVD based watermaking scheme for protecting rughtful ownership", IEEE Trans. Multimedia 4 (1) (march 2002) 121-128.

[7] Z. Milivojević, N. Savić, B. Prlinčević, "Parametri slike dobrog kvaliteta pogodnih za digitalnu obradu", 5. Međunarodno savetovanje Upravljanje znanjem i informatika, Kopaonik, 8-9. Janura 2019, pp. 129 -137.

[8] Q.Su, Y.Niu, X.Liu, Y.Zhu, "Embedding color watermarks in color images based on Schur decomposition", Optics Communications, 285 (2012) p.p. 1792-1802.

[9] G.H. Golub, C.F. Van Loan, Matrix computations, Johns Hopkins University Press, Baltimore, 1989.

[10] B. Prlinčević, Z. Milivojević, P. Spalević, D. Brodić, "Performanse SD algoritma za insertovanje vodenog žiga u slici baziranog na Šurovoj dekompoziciji", International Scientific Conference INFOTEH 2015, pp. 640-644, 18-20 March 2015, Jahorina, Bosnia nad Hecegovina.

[11] M. Meriama, B. Mohamed, "New Mean-Variance Gamma Method for Automatic Gamma Correction", I.J. Image, Graphics and Signal Processing, , Vol. 3, pp. 41-54, 2017.

[12] P Babakhani, P Zarei, "Automatic gamma correction based on average of brightness", ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 6, No.18, pp. 156-159, November 2015.

Modifying a Kinect v2 sensor to be able to connect to a Windows PC without the use of a Microsoft OEM adapter

Ivaylo Vladimirov[1], Desislava Nikolova[2] and Zornitsa Terneva[3]

Abstract – **This document gives the steps and procedure for modifying a Kinect v2 sensor in order to connect to a USB port on a Windows Personal Computer without using the official Microsoft OEM adapter. For this adaptation, the following items have been used: Kinect v2, Power supply adapter, F144 DC chassis jack, USB 3.0 cable type A-B, single conductor cable and a soldering iron. The idea behind this alteration is to find an alternative form to conjoin Kinect and PC after the not so recent announcement that Microsoft has discontinued the manufacturing of Kinect v2 adapters.**

Keywords – **Kinect v2 Sensor, Modification, Microsoft OEM adapter, USB 3.0 cable type A-B, Power supply adapter, F144 DC chassis jack.**

I. INTRODUCTION

The Kinect device is a motion-sensing sensor made by Microsoft as an accessory for the Xbox gaming console, but it also works with PCs. The various cameras and sensors that the Kinect has, enable it to sense movement just as well in the dark as to when it is well lit. It's due to the 3 constantly illuminating infra-red projectors emitting over the observed scene [1] [2].

The first sensor is an RGB camera that is able to capture full HD quality images up to resolutions of 1920 x 1080 p [1].

The second one is an infrared depth camera that unlike the color camera sees information in the 3-Dimensional plane, It has a resolution of 512 × 424 pixels and can sense high fidelity depth information using infrared sensors and recreate a 3-D view of what the Kinect senses [1]. The Kinect uses this information to perform body tracking from the depth data that it captures. It is smart enough to track up to six people with 25 joints on each person [3]. This information can then be used in applications that require human body interactions for example in healthcare applications for physical therapy, or in interactive games [4].

The third and last sensor is a microphone array consisting of four microphones. If a person wants to interact with the device, using his voice the array detects the direction of the sound source and amplifies the sound from one direction while suppressing the noise coming from other directions. Another use of these microphones is to find the location of a person and estimate his/her coordinates [5].

[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]Desislava Nikolova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: desislava.v.nikolova@gmail.com
[2]Zornitsa Terneva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: z.terneva@abv.bg

II. THE PROBLEM

The usage of the Kinect sensor has progressed way beyond conventional methods when it was only used to make applications and games solely for the Xbox console. The introduction of the Kinect to the PC platform by Microsoft in 2011 and the release of the software development kit (SDK), paved the way for developers to start thinking out of the box to create more interesting, innovative and ground-breaking ways to use the sensor [1].

Today, many businesses and education institutions have taken on the challenge of experimenting and implementing Kinect driven systems into their business processes, projects and work, in order to overcome certain barriers and pave the way for new developments, innovations and inventions [6].

The adapter (fig.1) consists of two main components: the USB adapter itself and the power supply. The USB adapter is a 1.0 by 1.6 by 3.4 inch black block. It has a Kinect port on one end and an USB 3.0 type B and power ports on the other. The Kinect plugs into the Kinect port, and the USB port connects via the included USB 3.0 B-to-A cable to the PC or Xbox One S. The separate 1.0 by 2.0 by 3.4 inch power supply plugs into the power connector with its own attached cable, and the included power cable connects that to any power outlet [7].

But in April 2015 Microsoft decided to stop manufacturing Kinects v2 sensors [8] and respectively in October 2017 - their adapters. The Kinect adaptor is already out of stock at almost every major retailer, leaving users still hoping to use a Kinect at the mercy of the second-hand market, where the adapter runs for between $150 to $300, making it an incredibly expensive proposition for anyone hoping to pick up a Kinect for their Xbox One or ever make use of one currently collecting dust in a closet [9].
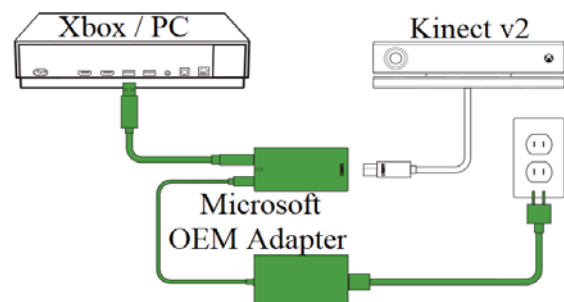
Fig.1 Diagram showing how to connect the adapter.

And so, with Microsoft's official confirmation on stopping the production of Kinect Adaptors, it seems that it's time to close the book on the motion gaming era once and for all - at least until someone tries to resurrect the idea [9].

III. The Solution

The decision of Microsoft Corporation resulted in us trying to find a way to connect a Kinect device to a PC without an original OEM adapter. After an initial research [10] [11] [12] we decided to modify the Kinect sensor using a USB 3.0 Type A-B cable, where the Type A connector is for the attachment to the PC and the Type B is for the Kinect and a DC power supply adapter 12V/3A with a F138N DC PLUG 2.5x5.5. See figure 2 for a general scheme of the modification.
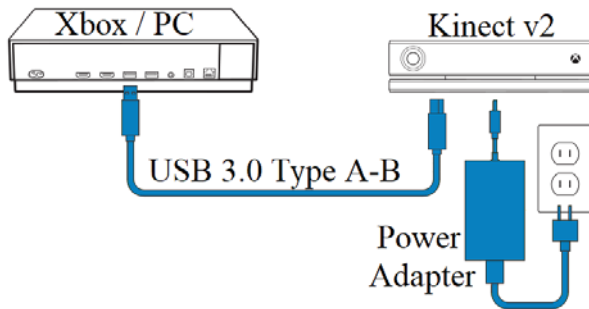
Fig.2 Diagram showing our solution.

The power supply was needed to be soldered via a soldering iron to two pins on the printed circuit board – the ground (GND) and the +12V pin. Figure 3 presents the theoretical approach on the left and the practically realized one on the right.
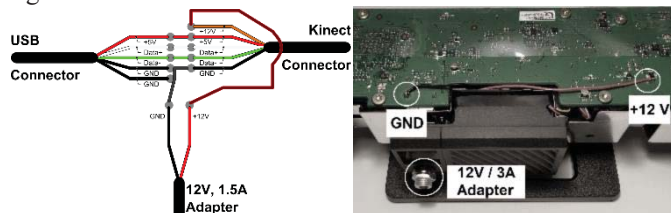
Fig.3 Connection schemes for the power supply.

Another adaptation that needs to be made is the addition of an F144 DC chassis jack. This jack gives the power supply the capability to be easily removed or plugged in and also makes the Kinect sensor more mobile in case of a need for a repair or just to be moved. The position of the F144 connector is secure, safe and it doesn't touch the protective screen. Figure 3 shows: on the left – the F144 jack before installation; in the middle – a view of the connector from the inside-out; on the right – the jack after its installation on the Kinect chassis.

Fig.4 The F144 DC chassis jack

The result of the whole process is presented on figure 4.

IV. Conclusion

The given solution is not only time effective but it also saves money, the prices of the used parts in dollars are: USB 3.0 cable - 9\$, power supply adapter - 10\$, F144 jack - 1\$. When comparing the total price with the price of an OEM adaptor (20\$ to about 100\$), we can see that the reduction of the cost is about 80%. The device is tested and no problems have been found.

Fig.5 The final result of the modification

References

[1] Kinect for Windows. Available online: https://microsoft.com/en-us/kinectforwindows/ default.aspx
[2] E. Lachat, H. Macher, T. Landes, P. Grussenmeyer,"Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling",2015
[3] E. Lachat, H. Macher, M.-A. Mittet, T. Landes, P. Grussenmeyer, "First experiences with Kinect v2 sensor for close range modelling". INSA Strasbourg 25-27 Febuary 2015
[4] PALIYAWAN Pujana, "Intelligent Middleware for Healthy Motion Gaming", Doctoral Dissertation, September 2018
[5] Ling Pei1, Liang Chen, Robert Guinness, Jingbin Liu, Heidi Kuusniemi, Yuwei Chen, Ruizhi Chen, Stefan Söderholm, "Sound positioning using a small-scale linear microphone array", IPIN Conference, 28-31 October 2013
[6] S. Abolhoseini, A. Sadeqi-Niaraki and O. R. Abbasi, "Overview of Microsoft Kinect applications in ubiquitous GIS," 2015.
[7] Will Greenwald, "Microsoft Kinect Adapter for Windows", https://www.pcmag.com, August, 2016
[8] EMIL PROTALINSKI, "Microsoft stops producing Kinect for Windows v2 sensor, will focus on Kinect for Xbox One and Windows apps", https://venturebeat.com, April, 2015
[9] Chaim Gartenberg, "Microsoft has discontinued the Kinect Adapter for newer Xbox One consoles", https://www.theverge.com, January, 2018
[10] "How to Hack Xbox One Kinect to Work on Windows 10 PC" available online: https://youtube.com/watch?v=l0rWWT24TNE
[11] "Connect Xbox One Kinect to PC Tutorial" available online: https://www.youtube.com/watch?v=KZunnyw_RKQ
[12] "DIY Kinect Adapter with Disassembly for Xbox One S, X and PC" available online: https://youtube.com/watch?v=7lQ93qzggo

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES
&
INFORMATICS AND COMPUTER SCIENCE

A Non-asymptotic Space Complexity of a Backtracking Algorithm for the N-queens Problem

Adrijan Bozinovski[1] and Stevo Bozinovski[2]

Abstract – **The paper explores a space complexity function for a backtracking algorithm solving the N-queens problem in embedded applications where application memory is limited. In such a case it is necessary to plan the needed memory space and to observe the space complexity function in some input domain. In this paper the detailed space complexity analysis is provided for N < 40.**

Keywords – **space complexity, back-tracking algorithm, N-queens problem.**

I. INTRODUCTION

Backtracking algorithms are one of the solutions of the constraint satisfaction problem [1]. Such an algorithm searches for a path in a tree, keeping in memory unexplored promising segments of paths. This raises a question of memory requirements, known as the space complexity problem, which is the focus of this paper. Using a backtracking algorithm for solving a problem is a standard approach in the theory of algorithms [2] and Artificial Intelligence [3], in cases when no information is given about the tree to be searched. Basically it uses a generate-and-test method, in which a tree expansion routine generates the next set of promising nodes, and a tester routine tests them for constraint satisfaction. The backtracking class of algorithms has been studied since the 1960's [4] and basic techniques are well known[5][6]. A backtracking algorithm has exponential worst-case time complexity, but there are studies showing that under certain conditions their average time complexity may not be exponential [7][8]

Space complexity of a backtracking algorithm has attracted less attention than time complexity. Newer theoretical reasoning has pointed out that space complexity is possibly polynomial [1]. That points out possibility of using such an algorithm in embedded applications where application memory is limited. Usually those applications require best-approximation rather than worst-case asymptotic approximation of needed space.

A prominent example of a problem which is being solved using a backtracking algorithm is the N non-attacking queens, in short the N-queens problem The N-queens problem is stated as: given a NxN chessboard, arrange N queens such that they do not attack each other . The problem has long history, starting in 1850 as a chess problem [9][10].

[1]Adrijan Božinovski is with School of Computer Science and Information Technology, University American College Skopje, Macedonia, email: bozinovski@uacs.edu.mk
[2]Stevo Bozinovski is with the Department of Mathematics and Computer Science, South Carolina State University, Orangeburg, SC, USA. Email: sbozinovski@scsu.edu

The N-queens problem is a benchmark problem and is used to test various types of algorithms. Among many approaches and applications let us mention integer programming [11], multitasking programming [12], linear congruence equations [13], dynamic programming [14], genetic algorithms [15], neural networks [16] and deductive database [17]. Currently, fast search algorithms in polynomial time are known for large N [18][19].

Bactracking can be considered as an operation defined over tree objects. Other operation are search, traversing, etc, many of the described in literature (e.g., [20]). Recently new operations on trees such as the Binary Roll Tree [21-24] have been proposed.

The solutions of the N-queens problem produce a particular pattern on a NxN matrix. An example of an N-queens pattern for N = 26 is given in Fig. 1 [25].
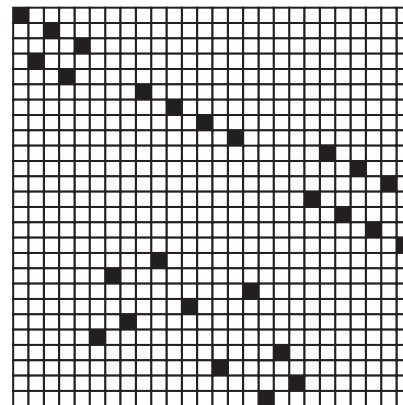
Fig. 1. An N-queens pattern for N=26

There are two variants of the problem: existence variant, in which only one solution is sufficient, and optimization variant, in which all feasible solutions need to be found and the optimal one to be chosen according to some optimality criterion. The numbers of solutions of the optimization variant for a given N are known, and for the first 10 values of N they are 1, 0, 0, 2, 10, 4, 40, 92, 352, 724 respectively. The numbers grow exponentially and for N=23 the number of possible solutions is over 24 trillion. Since the space complexity is the focus of this paper, the existence variant of the problem will be explored.

A. Scope of the Paper

This paper assumes an embedded system which has limited memory space for an application that runs a backtracking

algorithm solving an N-queens type of problem. Since the memory is limited, we explore the space complexity for N < N_{limit} , in our case N_{limit}=40. N_{limit} will depend on particular capacity of a embedded system.

II. METHODOLOGY

Firstly we provide a terminology that is we use in our research and we believe useful for both educational and research purposes.

A current path in a tree is a string of nodes passed up to the current node;

A promising path is the one whose last node is promising;

A node is promising if it can produce a next feasible node;

A feasible node is one that satisfies some constraints;

Given N, the promising path of length k is a string e[1]e[2]...e[k] where e[k] is a feasible node;

The promising path can be extended (toward a possible solution) or shortened (backtracked to a previous feasible path);

If k = N, a solution path of the problem has been found.

Our approach is to look for polynomial bounds for a limited N, rather than finding asymptotic complexity for an unlimited N.

In a space complexity analysis the crucial issue is what should be kept in memory. For a backtracking algorithm it is the set of all promising paths. A promising path should be kept in memory either until it is explored or until a solution is found. A promising path in a search tree of the N-queens problem can be represented in various ways, and we used a string representation. For keeping in memory all the promising paths we used a stack data structure. The stack data structure is chosen because it implicitly resolves the issue of which promising path will be explored next – it is the one on the top of the stack.

The Java program written for this research is based on the following pseudocode:

Non-recursive backtrack algorithm:

```
input N
expand the parent node
obtain the first N successor nodes
stack them in reverse order
repeat
    pop a node from the stack
    expand and test the expanded node
        if(feasible AND solution_length=N)
            then solution found, print it
        if(feasible) push it into stack
until empty stack
```

The solution is represented as a string which grows in length (solution-length) as the procedure progresses. At each step, a new element is appended from a set of possible expansions. The expansion set is tested for feasibility, and, if a node is not feasible, its subtree is pruned and the next element of the expansion set is considered. The length of the final

solution is known to be N, and, once N is achieved, a test for a possible solution gives a solution of the problem.

III. RESULTS

In our experimental research, we defined a space complexity function SpaceQns(N), as the length of the stack upon finding a solution to the N-queens problem. The values of SpaceQns(N) for N < 40, alongside the values of the functions N^2, $2N^2$, and $3N^2$, are shown on Fig. 2.
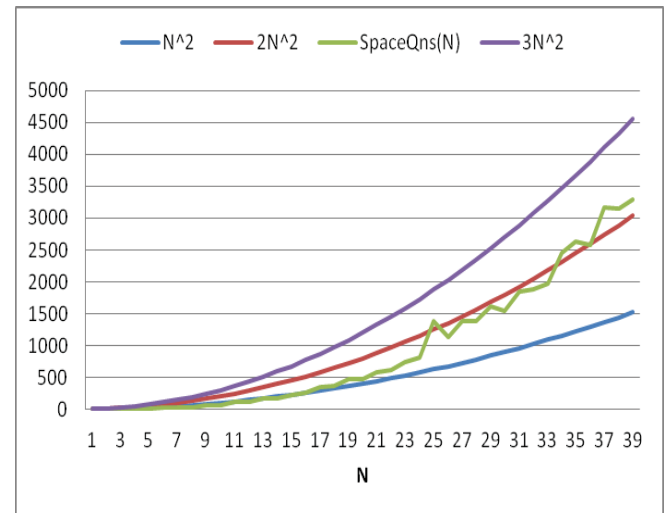
Fig. 2. Space complexity function SpaceQns(N) in range N <40.

More detailed observation gives the following description of space complexity of the N-queens problem, as represented by the SpaceQns(N) function:

$$
\text{SpaceQns(N)} \approx
\begin{cases}
\geq N^2, & N > 17, \text{ i.e. } \Omega(\text{SpaceQns(N)}) = N^2 \\
\leq 3N^2, & 4 < N < N_3, \text{ i.e. } O(\text{SpaceQns(N)}) = 3N^2
\end{cases}
$$

where $N_3 = \min\{N \mid \text{SpaceQns(N)} \geq 3N^2\}$ is the point when SpaceQns(N) passes the $3N^2$ bound. The value N_3 is larger than 40 and is outside our research scope, limited to N < 40. Within the experimental range of 1<N<40, it can be seen that $\Omega(\text{SpaceQns(N)}) = N^2$ (since SpaceQns(N) > N^2 for the entire range) and $O(\text{SpaceQns(N)}) = N^2$ (since SpaceQns(N) < $3N^2$ for the entire range), which then leads to the conclusion that $\Theta(\text{SpaceQns(N)}) = N^2$ (as per [27]), i.e., the space complexity of the backtracking algorithm for solving the N-queens problem is tightly quadratic. On Fig. 2 it can be seen that the SpaceQns(N) function can be approximated with the function $2N^2$.

IV. CONCLUSION

The paper provides a definition of terms used in backtrack programming and uses those terms in an experimental investigation on space complexity of a backtracking algorithm for solving N-queens problems. The scope of solution is

considered for N < 40, which is a non-asymptotic case, a case oriented toward an application of an embedded system.

The experimental investigation shows that in that scope the space complexity is a quadratic function, more specifically it can be approximated with the function $2N^2$. A more detailed observation in various segments of N < 40 is also provided. A precise knowledge for space complexity function (in terms of $2N^2$ rather than $O(N^2)$) is especially needed in embedded applications, where memory management is part of application design.

REFERENCES

[1] P. van Beek, "Backtracking Search Algorithms". In F. Rossi, P. van Beck, T. Walsh (eds.) Handbook of Constraint Programming, Elsevier, pp. 85-134, 2006

[2] G. Brassard, P. Bratley. *Fundamentals of Algorithmics.* Prentice Hall, 1996

[3] E. Rich. *Artificial Intelligence*, McGraw-Hill,1983

[4] S. Golomb, L. Baumert, "Backtrack programming". Journal of the ACM 12, pp. 516-524, 1965

[5] J. Fillmore, S. Williamson, "On backtracking: A combinatorial description of the algorithm". SIAM Journal of Computing 3, pp. 41-55, 1974

[6] J. Bitner, E. Reingold, "Backtrack programming techniques". Communications of the ACM 18, pp. 651-656, 1975

[7] D. Nicol, "Expected performance of m-solution backtracking". SIAM Journal on Computation 17(1), pp. 114-127, 1988

[8] H. Stone, P. Sipala, "The average complexity of depth-first search with backtracking and cutoff." IBM Journal of Research and Development 30(3), pp. 242-258, 1986

[9] F. Nauck , "Schach." Illustrierter Zeitung 361:352, 1850

[10] E. Pauls, "Das Maximalproblem der Damen auf dem Schachbrete." Deutsche Schachzeitung, Bd. 29, pp. 129-134, 257-267,1874

[11] L. Foulds, D. Johnson, "An application of graph theory and integer programming: Chessboard nonattacking puzzles." Mathematics Magazine 57(3), pp. 95-104, 1984

[12] R. Clapp, T. Mudge, R. Volz, "Solutions to the N-queens problem using tasking in Ada." SIGPLAN Notices 21, pp. 99-110,1986

[13] C. Erbas, M. Tanik, Z. Aliyaziciogly, "Linear congruence equations for the solution of the N-queens problem." Information Processing Letters 41, pp. 301-306, 1992

[14] L. Rivin, R. Zabih, "A dynamic programming solution to the n-queens problem" Information Processing Letters 41, pp. 253-256, 1992

[15] K. Crawford, "Solving the n-queens problem using genetic algorithms." Proc ACM/SIGAPP Symposium on Applied Computing, Kanzas City, pp.1039-1047, 1992

[16] O. Shagrir, "A neural net with self-inhibiting units for the N-queens problem." International Journal of Neural Systems 3, pp. 249-252, 1992

[17] J. Han, L. Liu, T. Lu, "Evaluation of declarative n-queens recursion: A deductive database approach." Information Sciences 105, pp. 69-100, 1998

[18] R. Sosic, J. Gu, "Fast search algorithms for the N-queens problem." IEEE Transactions on Systems, Man, and Cybernetics 21 (6), pp. 1572-1576, 1991

[19] R. Sosic, "A parallel search algorithm for the N-queens problem." Proc Parallel Computing and Transputer Conference, Wollongong, pp. 162-172, 1994

[20] Knuth D. *The art of Computer Programming*. Addison Wesley, 1997

[21] A. Božinovski, N. Ackovska, "The Binary Tree Roll Operation: Definition Explanation and Algorithm", International Journal of Computer Applications*, vol. 46, no. 8, pp. 40-47, 2012

[22] A. Božinovski, G. Tanev, B. Stojčevska, V. Pačovski, N. Ackovska, "Time complexity analysis of the binary tree roll algorithm", Journal of Information Technology and Applications, vol. 6, no. 2, pp. 53-62, 2016

[23] A. Božinovski, G. Tanev, B. Stojčevska, V. Pačovski, N. Ackovska, "Space complexity analysis of the binary tree roll algorithm", Journal of Information Technology and Applications, vol. 7, no. 1, pp. 9-19, 2017

[24] G. Tanev, A. Božinovski, "A linear time algorithm for rolling binary trees", Proceedings of the 17th IEEE International Conference on Smart Technologies IEEE EUROCON 2017, Ohrid, Macedonia, pp. 255-260, 2017

[25] A. Bozinovski, S. Bozinovski, "N-queens pattern generation: An insight into space complexity of a backtracking algorithm", Proceedings of the 3rd International Symposium on Information and Communication Technologies. Las Vegas, Nevada, USA, pp. 281-286, 2004

[26] The On-Line Encyclopedia of Integer Sequences, published electronically at https://oeis.org, 2010, Sequence A000170

[27] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein. *Introduction to Algorithms*. *Third Edition*. The MIT Press, 2009

Developing a B2C e-Commerce Graph Data Model from a Relational Schema

Ilija Hristoski[1], Tome Dimovski[2], and Violeta Manevska[2]

Abstract – **The emerging NoSQL databases have capabilities that traditional relational databases do not possess. In this paper, we are focusing on NoSQL graph databases, and propose a graph data model that corresponds to a generic traditional relational database model found within the majority of today's B2C e-Commerce applications. The act of mapping relational databases to graph databases provides a solid basis for carrying out profound analyses of the implicit relationships among entity types' instances and gaining significant insights that cannot be incurred otherwise.**

Keywords – **E-Commerce, Relational schema, Graph data model, Mapping, Neo4j.**

I. INTRODUCTION

Nearly 50 years after Edgar F. Codd postulated the mathematical principles of the relational data model back in 1970, relational databases are still playing the role of a main workhorse in the contemporary digital landscape. There are many reasons for such an observation, e.g. relational databases are keeping data redundancy at a minimum; they are transaction friendly and highly consistent during update operations. Among major disadvantages of relational databases, which include the necessity to rebuild relationships with JOINs and other inexact techniques, their sensitivity to changes is, perhaps, of the utmost importance. The fact that relational databases are not designed to handle changes is crucial, having minded the nature of today's data, which is more dynamic, more intense, more voluminous, more unpredictable, more variable, and more diverse than ever before. Moreover, relational databases are neither designed for heterogeneous data, nor for scale and for mixed workloads, which make them a mismatch for modern app development.

Nowadays it becomes quite clear that relational databases cannot cope successfully with the challenges posed by today's data, and justifies what Vivek Kundra, the former CIO of the U.S. Federal Government, said in 2009: "This notion of thinking about data in a structured, relational database is dead."

Graph databases, on the other hand, belong to the family of NoSQL databases; they address one of the great macroscopic business trends of today: "leveraging complex and dynamic relationships in highly connected data to generate insight and competitive advantage" [1]. Unlike the relational databases, which store data to efficiently store facts, graph databases store both facts and the relationships among the facts, making certain types of analyses more efficient and intuitive. Emerging as a major driver of innovation during recent years, graph databases have already exhibited many advantages over relational databases, like storing large volumes of data that might have little to no structure, sharing data across multiple servers in the cloud, speeding the development, boosting the horizontal scalability, demonstrating superior performances, and supporting iterative algorithms and other data mining and machine learning algorithms.

The paper is organized as follows. Section 2 provides an insight into some of the most relevant research in this area. A generic B2C e-Commerce relational schema, which is used as a basis for this study, is presented in Section 3. In Section 4, the mapping of the B2C e-Commerce relational schema into a corresponding graph data model is described. Section 5 concludes.

II. RELATED RESEARCH

There is intense ongoing research in the area of mapping / converting / transforming / migrating relational databases into graph databases during recent years.

De Virgilio et al. propose a methodology for converting a relational to a graph database by exploiting their schema and the constraints of the source database. They also provide experimental results that demonstrate the feasibility of their approach and the efficiency of query execution against the target database [2]. Bordoloi & Kaita propose a method for transforming a relational database to a graph database model by transforming dependency graphs for the entities in the system into star graphs, which are then transformed into a hypergraph model [3]. Wardani & Küng propose a methodology for mapping and converting relational models to graph models without any semantic loss [4]. De Virgilio et al. present R2G, a tool for automatic migration of relational databases to a Graph Database Management System (GDBMS) [5]. Recognizing the fact that identifying correlations and relationships between entities within and across different data sets (or databases) is of a great importance in many domains, Lee et al. develop a reconfigurable and reusable graph construction tool, named *Table2Graph*, based on a Map-Reduce framework over Hadoop, with an aim of constructing a graph-based model from relational source databases [6]. Recently, Filho et al. proposed some heuristics, aiming at mapping systematization from relational model data to graph representation, in order to provide support for an adequate choice of the graph model, according to the type of analysis to be performed [7].

[1]Ilija Hristoski is with the Faculty of Economics at the "St. Kliment Ohridski" University in Bitola, 133 Gjorche Petrov St, 7500 Prilep, North Macedonia, E-mail: ilija.hristoski@uklo.edu.mk.

[2]Tome Dimovski and Violeta Manevska are with the Faculty of Information and Communication Technologies at the "St. Kliment Ohridski" University in Bitola, Partizanska St, 7000 Bitola, North Macedonia.

III. B2C e-Commerce Relational Schema

In order to achieve the goal of this study, as a starting point, we use the generic B2C e-Commerce relational data model published by Williams in 2009 [8]. The database model consists of 14 tables, drawn from the following relational schema:

CUSTOMER
      (**customer_id**, …)
CUSTOMER_PAYMENT_METHOD
      (**customer_payment_id**, …, *customer_id**, *payment_method_code**)
REF_PAYMENT_METHOD
      (**payment_method_code**, …)
PRODUCT
      (**product_id**, …, *product_type_code**)
REF_PRODUCT_TYPE
      (**product_type_code**, …, *parent_product_type_code**)
ORDER
      (**order_id**, …, *customer_id**, *order_status_code**)
REF_ORDER_STATUS_CODE
      (**order_status_code**, …)
ORDER_ITEM
      (**order_item_id**, …, *product_id**, *order_id**, *order_item_status_code**)
REF_ORDER_ITEM_STATUS_CODE
      (**order_item_status_code**, …)
INVOICE
      (**invoice_number**, …, *order_id**, *invoice_status_code**)
REF_INVOICE_STATUS_CODE
      (**invoice_status_code**, …)
PAYMENT
      (**payment_id**, …, *invoice_number**)
SHIPMENT
      (**shipment_id**, …, *order_id**, *invoice_number**)
SHIPMENT_ITEM
      (**shipment_id**, **order_item_id**)

In the above relational schema, non-key attributes are being omitted, i.e. only table names, along with corresponding primary and foreign keys are given, due to simplicity reasons. Primary keys are bolded and underlined, whilst foreign keys are italicized.

IV. Mapping the Relational Schema into a Graph Data Model

Before transforming the relational database schema into a graph database model, it is convenient to point out the process of transformation, which can be carried out through a number of steps [9-11]:

- Each entity table is represented by a label on nodes;
- Each row in an entity table becomes a particular node;
- Columns on those tables become node properties;
- Technical primary keys should be removed, whilst keeping business ones;
- Unique constraints should be added for business primary keys;
- Indexes should be added for frequent lookup attributes;

- Each foreign key should be replaced with a relationship to the other table, and removed afterward from the original table;
- Data with default values should be removed, there is no need to store those;
- Data in tables that is denormalized and duplicated might have to be pulled out into separate nodes to get a cleaner model;
- Indexed column names might indicate an array property;
- Simple JOIN tables become relationships;
- Attributed JOIN tables become relationships with properties.

Five tables in the original relational schema: CUSTOMER, REF_ PAYMENT_METHOD, CUSTOMER, REF_ORDER_ STATUS_CODE, REF_ORDER_ITEM_STATUS_CODE, and REF_INVOICE_STATUS_CODE, which, besides other non-key attributes, does not include any foreign keys, but solely a primary key, can be directly converted into nodes. The names of these tables become node labels, whilst columns on those tables become node properties (Fig. 1)
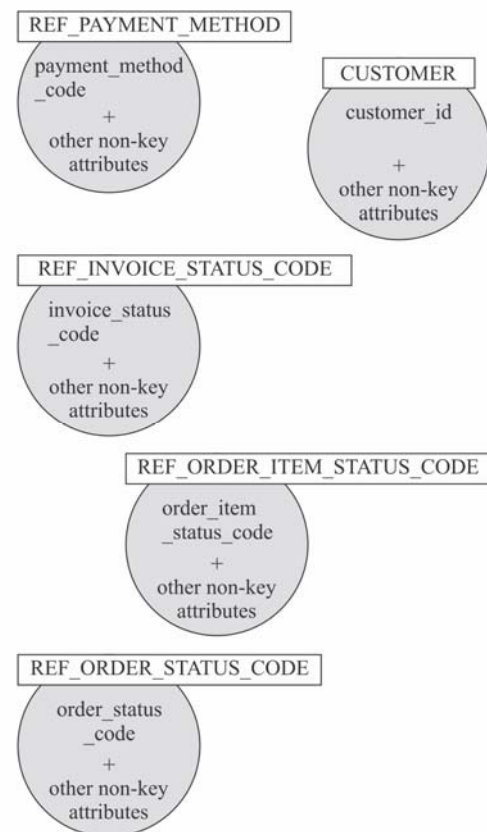
Fig. 1. Nodes coming out from the tables that do not include any foreign keys

Semantic relations between tables in the relational schema can be both identified and drawn from foreign keys, starting from the table with a foreign key(s) and following the relationship(s) to the table(s) where those foreign keys are primary keys (Table 1).

TABLE I
FINDING THE SEMANTIC RELATIONS BETWEEN TABLES

Semantic relation	Logical link (foreign key)
CUSTOMER_PAYMENT_METHOD – *BelongsTo* → REF_PAYMENT_METHOD	payment _method_code
CUSTOMER_PAYMENT_METHOD – *IsUsedBy* → CUSTOMER	customer_id
ORDER – *IsPlacedBy* → CUSTOMER	customer_id
ORDER – *HasOrderStatus* → REF_ORDER_STATUS_CODE	order_status _code
INVOICE – *IsIssuedFor* → ORDER	order_id
INVOICE – *HasInvoiceStatus* → REF_INVOICE_STATUS_CODE	invoice_status _code
PAYMENT – *IsMadeFor* → INVOICE	invoice_number
SHIPMENT – *CorrespondsTo* → ORDER	order_id
SHIPMENT – *IsCoveredBy* → INVOICE	invoice_number
PRODUCT – *IsOf* → REF_PRODUCT_TYPE	product_type _code
ORDER_ITEM – *RefersTo* → PRODUCT	product_id
ORDER_ITEM – *IsPartOf* → ORDER	order_id
ORDER_ITEM – *HasOrderItemStatus* → REF_ORDER_ITEM_STATUS_CODE	order_item _status_code
REF_PRODUCT_TYPE – *IsSubtypeOf* → REF_PRODUCT_TYPE	product_type _code

The logical links like *BelongsTo*, *IsUsedBy*, *IsPlacedBy*, etc., are implemented through the mechanism of foreign keys, but they are not obvious, because neither the existence nor the meaning of such semantic relations is not explicitly coded in the relational database schema. After identifying the semantic relations originating from the foreign keys, the latter ones are being removed from the tables. Those tables also become nodes, and the semantic relations originating from these become relations pointing towards corresponding nodes in a graph data model (Fig. 2).
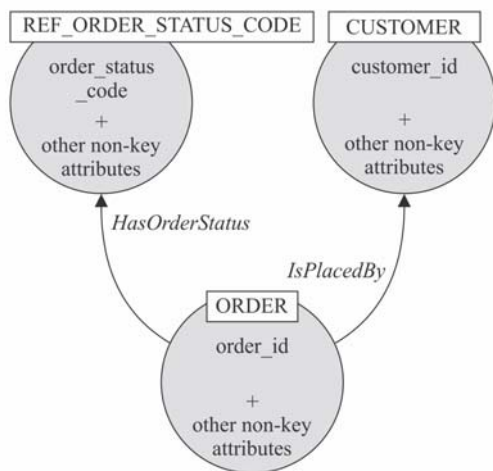
Fig. 2. Transforming foreign keys of the table ORDER into relations to nodes CUSTOMER and REF_ORDER_STATUS_CODE

The table REF_PRODUCT_TYPE is a specific one since it contains a foreign key that references its own primary key. This is a recursive relationship of cardinality 1:M. In this particular case, the foreign key transforms into a semantic relation *IsSubtypeOf* between two nodes with the same label, REF_PRODUCT_TYPE, which represent two distinct product types (Fig. 3).
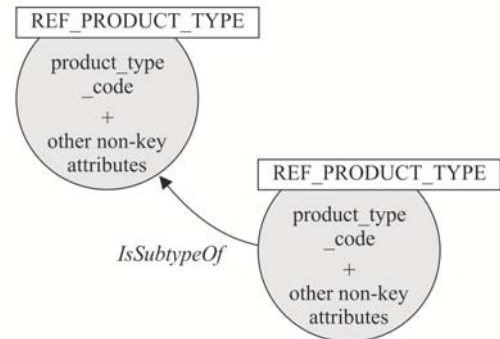
Fig. 3. Transforming the foreign key *parent_product_type_code** of the table REF_PRODUCT_TYPE into a semantic relation *IsSubtypeOf*

Finally, the table SHIPMENT_ITEM has been derived from an M:N relationship between the entity types SHIPMENT and ORDER_ITEM in the original E-R diagram. Since it represents a simple JOIN table without any additional attributes, it should not be represented as a node, but rather as a relationship, starting either from the node SHIPMENT to the node ORDER_ITEM, or vice-versa. Fig. 4 displays this relationship, named *Includes* (Fig. 4).
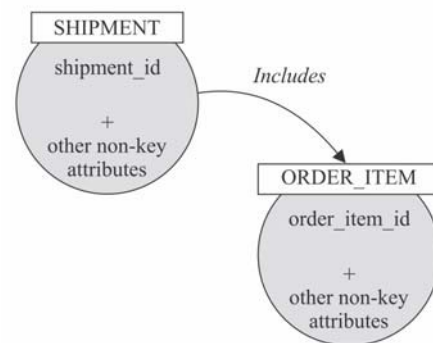
Fig. 4. Transforming the table SHIPMENT_ITEM into a semantic relation *Includes* between the nodes labeled SHIPMENT and ORDER_ITEM

The complete B2C e-Commerce graph data model has been implemented in Neo4j using the Cypher query language (CQL) and is given in Fig. 5. When it comes to Neo4j, it is worth pointing out that the Neo4j ETL (Extract-Transform-Load) tool allows one an automated translation of both relational data structures and actual data stored in an existing relational database into a graph data model. It includes a 3-step process that allows the user to specify a relational database via a JDBC setup, then to edit the data model

mapping that the tool creates for the graph, and finally, to import all of the data itself into a Neo4j database.

Fig. 5. The resulting B2C e-Commerce graph data model

V. CONCLUSION

The resulting graph data model differs from the original relational data model in several aspects: (1) There are no NULLs, i.e. non-existing value entries/properties are not present; (2) The graph model has no JOIN tables; (3) The graph model has no artificial primary keys or foreign keys; (4) Relationships are more detailed; (5) There is no need for intermediary tables, because they are all translated directly into relationships. The resulting graph data model captures the relationships among entity types in a highly intuitive and simplified manner because it is more expressive than any other corresponding relational data model. It provides a substantial basis for conducting advanced data analytics in a time-efficient manner, like e-Commerce recommendations, personalizations, security, real-time fraud detection, etc.

REFERENCES

[1] I. Robinson, J. Webber and E. Eifrem, *Graph Databases: New Opportunities for Connected Data*, Second Edition, O'Reilly Sebastopol, CA, USA, 2015.

[2] R. De Virgilio, A. Maccioni and R. Torlone, "Converting Relational to Graph Databases", First International Workshop on Graph Data Management, Experience and Systems (GRADES 2013), Conference Proceedings, New York, NY, USA, pp. 1–6, 2013.

[3] S. Bordoloi, B. Kalita, "Designing Graph Database Models from Existing Relational Databases", International Journal of Computer Applications, vol. 74, no. 1, pp. 25–31, 2013.

[4] D. W. Wardani, J. Küng, "Semantic Mapping Relational to Graph Model", 2014 International Conference on Computer, Control, Informatics and Its Applications (IC3INA 2014), Conference Proceedings, Bandung, Indonesia, pp. 160–165, 2014.

[5] R. De Virgilio, A. Maccioni and R. Torlone, "R2G: a Tool for Migrating Relations to Graphs", 17th International Conference on Extending Database Technology (EDBT 2014), Conference Proceedings, Athens, Greece, pp. 640–643, 2014.

[6] S. Lee, B. H. Park, S-H. Lim and M. Shankar, "Table2Graph: A Scalable Graph Construction from Relational Tables Using Map-Reduce", 2015 IEEE First International Conference on Big Data Computing Service and Applications, Conference Proceedings, pp. 294–301, Redwood City, CA, USA, 2015.

[7] S. P. L. Filho, M. C. Cavalcanti and C. M. Justel, "Graph Modeling from Relational Databases", 2017 XLIII Latin American Computer Conference (CLEI 2017), Conference Proceedings, Cordoba, Argentina, pp. 843–842, 2017.

[8] B. Williams, "A Data Model for e-Commerce", 2009. URL: http://www.databaseanswers.org/data_models/e_commerce/index.htm (Accessed March 30, 2019)

[9] M. Hunger, R. Boyd and W. Lyon, *The Definitive Guide to Graph Databases for the RDBMS Developer*, e-Book, Neo Technology, 2016. URL: https://neo4j.com/whitepapers/rdbms-developers-graph-databases-ebook/ (Accessed April 02, 2019)

[10] M. Hunger, *From Relational to Graph: A Developer's Guide*, e-Book, DZone, Inc., 2016. URL: https://dzone.com/storage/assets/2054302-dzone-refcardz-231-neo4j.pdf (Accessed April 2, 2019)

[11] S. Yang, *How to Map Relational Data to a Graph DB in Four Steps*, e-Book, TIBCO Software Inc., Palo Alto, CA, USA, 2018. URL: https://www.tibco.com/sites/tibco/files/resources/sb-graph-database-final.pdf (Accessed April 4, 2019)
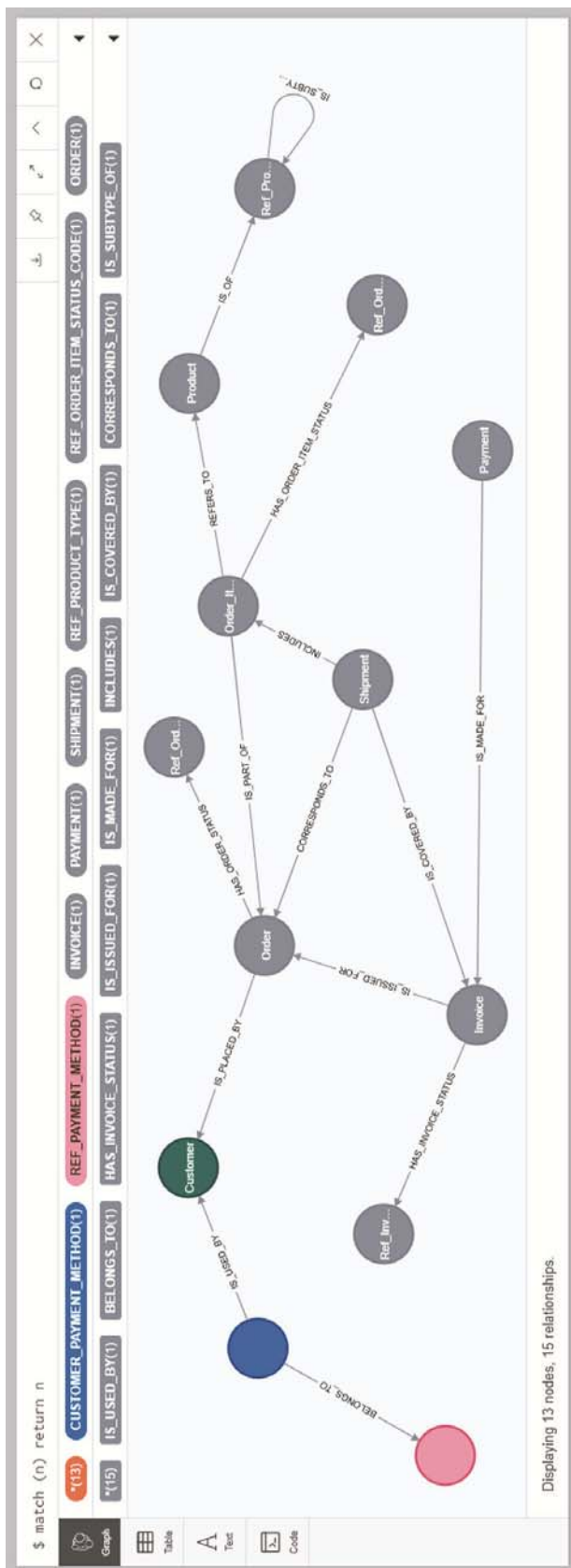
FINANCIAL PORTFOLIO OPTIMIZATION USING CLUSTERING ALGORITHMS

Ivana P. Marković[1], Jovica M. Stanković[2], Jelena Z. Stanković[3] and Miloš B. Stojanović[4]

*Abstract –***The decision to invest in capital market instruments is based on the analysis of return and risk, which are characteristics of investment in these instruments, as well as the possibility of their successful prediction in the future. The development of the stock market enhances the possibilities for investment, what maintains the interest of investors in this field of study. The presented empirical study focuses precisely on developing markets, on the Belgrade Stock Exchange in particular. The goal of the research is to, by applying cluster algorithms, identify various groups of shares, which, if included in the portfolio, would enable the diversification of risk.**

Keywords – dynamic time warping, K-means, hierarchical clustering, portfolio optimization.

I. INTRODUCTION

Portfolio optimization represents one of the most important aspects of making investment decisions. Despite the fact that it was presented to the academic and professional audiences back in the 1950s, modern portfolio theory is the first quantitative framework for the diversification of investment portfolios, which is widely being used today. The simplified initial assumptions in terms of investor preferences and the probability distribution of the financial assets' returns have made it possible to form a model of optimization, which, according to Markowitz's approach, is based only on the values of the expected return and risk. A rational investor will prefer efficient portfolios, which provide maximal return for an acceptable level of risk. However, in cases when a large number of financial instruments are being considered, high dimensionality can prevent the precise evaluation of a complex correlation structure and risk. Thus, the application of Markowitz's model in the optimization of large portfolios can result in the allocation of financial means to suboptimal investment alternatives.

The problem of diversification when creating effective portfolios in comparison to more conventional methods could also be solved by the application of numerous alternative methods. In numerous studies, the algorithms of machine learning have proved to be quite effective in the creation of effective portfolios. At the same time, the most frequently used approaches include artificial neural networks [1], [2], genetic algorithms [3], and cluster analysis [4] and [5].

Cluster analysis represents a technique of supervised machine learning which attempts to establish previously unknown and useful connections among the data, whereby different insight into the structure of the data is obtained. A special type of clustering is the clustering of time series, which are classified as dynamic data since their characteristics can change over observed period of time. Additionally, clustering of financial time series is especially challenging task, taking into consideration the fact that the financial market is a complex, evolving and dynamic system whose behavior is pronouncedly non-linear.

Most of the research that focuses on cluster analyses with the aim of portfolio optimization refers to the developed financial markets, as shown in [5] and [6]. On the other hand, the presented empirical study focuses on developing markets, and the Belgrade Stock Exchange in particular. This aim of this paper is to apply a cluster algorithm in the identification process of various groups of shares, whose inclusion in the portfolio would provide risk diversification. The subject matter of the analysis is portfolio optimization of the shares which are constituents of the BELEX15 and BELEX*line* index.

The paper is organized as follows: the second part presents the theoretical basis of the applied clustering methods. The third describes the used data and the experimental framework, while the fourth part presents the results of the cluster analysis and the results of the portfolio optimization. In the final part, some of the conclusions and directions for further research are presented.

II. CLUSTERING ALGORITHMS

This paper analyzes the influence of two most frequently used groups of clustering algorithms, partitional and hierarchical clustering algorithms, on portfolio optimization. In partitional clustering, the dataset object is divided into non-overlapping clusters, while in the case of hierarchical clustering, nested clusters are organized as subclusters.

A. Partitional clustering algorithms

K-means is one of the most widely used algorithms in this group thanks to its easy implementation, simplicity, efficiency, and empirical success [7].

Let $X = \{x_i\}$ $1;...;$ n be the set of n d-dimensional points to be clustered into a set of K clusters, $C = \{C_k; k = 1; ... ; K\}$. Let μ_k be the mean of cluster C_k while x_i is a data point belonging

[1]Ivana P. Marković is with the Faculty of Economics at the University of Niš, Trg kralja Aleksandra Ujedinitelja 11, 18000 Niš, Serbia, e-mail: ivana.markovic@eknfak.ni.ac.rs
[2]Jovica M. Stanković is with the Faculty of Economics at the University of Niš, Trg kralja Aleksandra Ujedinitelja 11, 18000 Niš, Serbia, e-mail: jovica.stankovic@eknfak.ni.ac.rs
[3]Jelena Z. Stanković is with the Faculty of Economics at the University of Niš, Trg kralja Aleksandra Ujedinitelja 11, 18000 Niš, Serbia, e-mail: jelenas@eknfak.ni.ac.rs
[4]Miloš B. Stojanović is with the College of Applied Technical Sciences Niš, Aleksandra Medvedeva 20, 18000 Niš, Serbia, e-mail: milos.stojanovic@vtsnis.edu.rs

to the cluster C_k. The squared error between μ_k and the points in cluster C_k are defined as:

$$W(C_k) = \sum_{x_i \in C_k}(x_i - \mu_k)^2 \qquad (1)$$

Each observation (x_i) is assigned to a given cluster such that the sum of squares (SS) distance of the observation to their assigned cluster centers μ_k is minimal. The total within-cluster variation is defined as [8]:

$$tot.withinss = \sum_{k=1}^{K} W(C_k) = \sum_{k=1}^{K}\sum_{x_i \in C_k}(x_i - \mu_k)^2 \qquad (2)$$

The *total within-cluster sum of square* measures the compactness of the clustering, and it should be minimized.

A K-means algorithm starts with an initial partition with K clusters, by randomly selecting k objects from the data set as the initial cluster centers or means, and assigns observations to the closest cluster centers. After all the observations are assigned to the clusters, new cluster centers are computed, and the distance between each observation and new obtained cluster centers is recalculated. These steps are repeated until the cluster assignments stop changing or the maximum number of iterations is reached so that the total within-cluster variation between points in the cluster and the corresponding centroid is minimized.

B. Hierarchical clustering

In hierarchical clustering, the clusters are determined by applying an agglomerative or divisive algorithm. Agglomerative clustering uses a bottom-up approach to perform hierarchical clustering, so that each observation is primarily found in its special cluster, and then the clusters are combined into larger clusters until finally all the observations belong to the same cluster or until a stopping criterion is satisfied. A divisive algorithm generates clusters following the opposite procedure, that is in a top-down manner, beginning with one cluster which contains all the observations and then later determining the subclusters.

In this paper we used two categories of hierarchical clustering, the principal agglomerative clustering algorithms single linkage and complete linkage, which use linkage criteria to determine the metric which is used when grouping clusters. The single linkage algorithm calculated the minimum distances between all observations of pairs of clusters, while the complete linkage algorithm calculated the maximum distance between observations of pairs of clusters.

A similarity measure is of essential importance for each clustering algorithm, but similarity measures on time series data are much harder to determine than on constant data, because of their own continuous nature. This is why the similarity measure is, as a rule, carried out in an approximate manner [9]. When clustering time series, various types similarity measures are used, considering that each measure reflects a different type of similarity among time series data.

The shape-based measures as the Euclidean distance and Dynamic Time Warping (DTW) [10], which find similarities in the time series in the domain of time and shape, are widely used for clustering of times series data [9]. This is why in the experimental part of this study we used these two specific measures.

The Euclidean metric is defined as the root of the square difference between co-ordinates of pairs of objects [8] and the Euclidean distance between two points X_i and X_j, with the m dimension calculated as:

$$\mathrm{Dist}X_iX_j = \sqrt{\sum_{k=1}^{m}(X_{ik} - X_{jk})^2} \qquad (3)$$

On the other hand, DTW aligns two time series., $Q = q_1, q_2, \ldots, q_n$ and $P = p_1, p_2, \ldots, p_m$, using matrix M_{nxm}, in order to minimize their difference. The element *(i,j)* of the matrix M represents the distance $d = (q_i, p_j)$ between two points q_i and p_j where in order to calculate the distance various methods can be used. In this paper we used a standard Euclidean distance. The warping path, denoted by W, $W = w_1, w_2, \ldots w_k, \ldots w_K$ while $max(m,n) \leq K \leq m+n-1$, actually represents a group of matrix elements which define the mapping between Q and P, with the k-th element $w_k = (i_k, j_k)$. The wrapping path needs to satisfy three conditions: the continuity condition, the boundary condition and the monotonicity condition [10]. The optimal path is determined by applying dynamic programming in order to find the wrapping path W which minimizes the wrapping cost, that is, minimize the distance between the two time series:

$$DTW(Q,P) = \min_{w}[\sum_{k=1}^{K} d(w_k)] \qquad (4)$$

where $d(w_k) = (q_{i_k}, p_{i_k}) = (q_{i_k} - p_{i_k})^2$. The complete derivations can be found in [11].

III. EXPERIMENTAL DATA

The value of BELEX*line* and BELEX15 indexes are determined by the prices of the most liquid shares which are continuously being traded on the regulated market of the Belgrade Stock Exchange. An adequate evaluation of the basic characteristics of financial time series requires a certain duration of the financial time series, which in the opinion of analysts reduces the number of available stocks to 29. Taking into consideration the required conditions which the companies issuing the shares need to meet in order for them to be included in these lists, the starting point in this study was the assumption that they are the most liquid securities on the Belgrade Stock Exchange.

The data used in this paper were taken from the website of the Belgrade Stock Exchange (www.belex.rs). The series of values of the index, as well as of individual shares, include records from January 1, 2008 to December 31, 2017, which in total includes 2524 days of training. Clustering algorithms are executed using 29 assets observation each of which with the 2524 trading data.

Afterwards, we distributed assets inside each cluster relying on the Omega ratio as the optimization strategy for asset allocation. The Omega ratio allows comparisons between risk and return at different threshold levels for various asset choices [12]. Assuming that $F(x)$ is a cumulative distribution function of the rate of return on investment and r_m is the minimum acceptable return, the Omega ratio can be calculated using the following equation:

$$\Omega_t = \frac{\int_{r_m}^{\infty}(1-F(x))dx}{\int_{-\infty}^{r}F(x)dx} \qquad (5)$$

As a general measure for the evaluation of the quality of the obtained cluster analysis, in this paper we used three different economic performance measures: the return r_t, risk σ_t and the

Sharpe ratio SR_t, which are calculated according to the following equations:

$$r_t = logCP_t - logCP_{t-1} \qquad (6)$$

$$\sigma_t = \sqrt{\frac{\sum_{t=1}^{n}(r_t - \bar{r})^2}{n-1}} \qquad (7)$$

$$SR_t = \frac{r_t}{\sigma_t} \qquad (8)$$

where CP represents closing price at t, $t = 1, 2, ..., n$

IV. RESULTS AND DISCUSSION

In this paper we tested the potential of clustering in portfolio optimization using two different clustering algorithms: K-means with an Euclidean distance, as a representative of partitional clustering, and a single linkage (SL) and a complete linkage (CL) agglomerative clustering algorithm by applying the DTW metric.

Fig.1 shows the dendrogram of hierarchical clustering by using a single-linkage method on the data set, while Fig. 2 shows the dendrogram for the same time period but with the application of a complete linkage method.
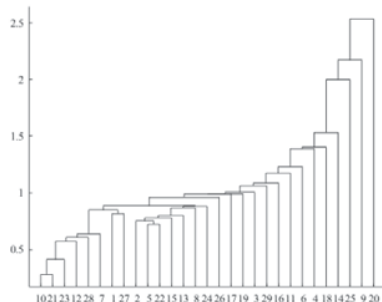
Fig. 1. Single linkage
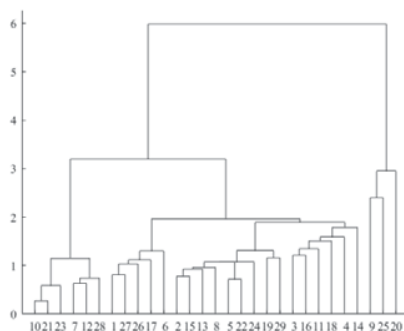
Fig. 2. Complete linkage

According to the figures, we can note a difference in the content and distribution of the clusters, depending on the applied methods.

Fig. 3. shows the results of the application of the DTW metric on the studied group for two shares AGBC and VBAV over a three-month period.
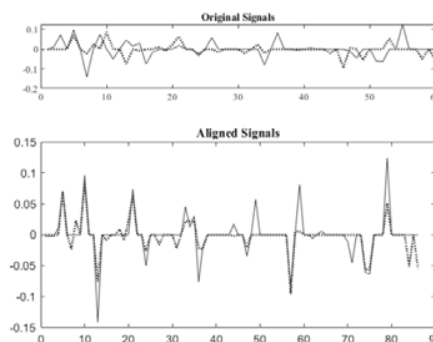
Fig.3. Example of DTW alignment between AGBC (1) and VBAV(27) shares in a three-month period

It is known that the performance of each clustering algorithm is affected by the number of created clusters. In this paper we selected the optimal number of clusters using the Calinski-Harabaz index [13]. This score is given as ratio of means between clusters and the within-cluster dispersion. Typically, for K-means the number of clusters is determined using the Elbow method by computing K-means clustering using different values of K. Next, the wss (within sum of square) is plotted according to the number of clusters. The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters.

TABLE I
THE SELECTION OF SHARES BASED ON THE ALGORITHM AND MODEL

Clustering algorithms	Shares
K-means (K=4)	JMBN, GLOS, TIGR, PRGS
CL (K=2)	PRGS, JMBN
SL (K=2)	PRGS, GLOS
BELEX15	NIIS, KMBN, AERO, FITO, MTLC, TGAS, ALFA, ENHL, JESV, SJPT, IMPL
BELEX*line*	AERO, NIIS, MTLC, KMBN, FITO, JMBN, TGAS, ALFA, ENHL, FINT, EPEN, JESV, LSTA, SJPT, IMPL, DNOS, KOPB, EPIN, TIGR, VDAV, GLOS, VBSE

Based on the data shown in Table I, we can note that the K-means algorithm a greater number of clusters, 4 compared to 2 or 3 which is the number of clusters obtained by applying the method of hierarchical clustering. In Table I we can also note a difference in the selected shares based on the applied algorithms and models of clustering. The table presents the shares of both market indexes, BELEX15 and BELEX*line*. In relation to the market indexes, we can note that the clustering algorithms recommend the use of a smaller number of shares. The results obtained by the portfolio optimization based on applying the clustering method, compared to the values of the market indexes BELEX15 and BELEX*line* during the monitored test period are represented in tables II and III.

TABLE II
TEST PERIOD 01.01.-01.03.2018

	Return	Risk	Sharpe	Omega
K-means	-0.214	0.214	-0.998	0.696
CL	-0.191	0.273	-0.699	0.767
SL	-0.233	0.244	-0.955	0.705
BELEX15	-0.299	0.177	-1.689	0.525
BELEX*line*	-0.222	0.100	-2.226	0.428

Based on the first test period, which includes the first quartal of 2018, we can note that the K-means algorithm allow the creation of a portfolio which is characterized by a significantly smaller risk, but a less favorable relationship between return and risk, measured by the Sharpe and Omega ratios, compared to the portfolios obtained by using algorithms of hierarchical clustering. The complete linkage method is used to create a portfolio which realizes the best results based on all the suggested algorithms, but also market indexes, considering that the Sharpe ratio of such a portfolio is 2.5 times smaller than the BELEX15 index, that is, 3 times smaller than the BELEX*line* index.

TABLE III
TEST PERIOD 01.01.-31.12.2018

	Return	Risk	Sharpe	Omega
K-means	-1.309	0.401	-3.265	0.551
CL	-1.042	0.504	-2.066	0.676
SL	-1.417	0.567	-2.500	0.609
BELEX15	-1.460	0.396	-3.687	0.479
BELEX*line*	-1.193	0.257	-4.648	0.387

During the second test period, which includes all of 2018, we can note that all the clustering models provide better results for the studied evaluation parameters compared to the values of the market indexes BELEX15 and BELEX*line*. The application of the K-means algorithm enables the creation of portfolios whose performances are improved compared to those of the studied market indexes in terms of the return and risk ratio, while the complete linkage method provides a portfolio which offers the most favorable ratio between return and risk and significantly improves the features of the optimized portfolio.

Experiments have shown that the traditional Euclidean distance metric cannot always be a suitable distance function in financial time series analysis, while DTW is able to find optimal global alignment between sequences and capture specific similarities.

V. CONCLUSION

Considering that learning from data and adapting to financial market changes are a specific and favorable advantage of machine learning methods, we have seen an increased use of these methods in financial analyses. As a special type of machine learning, clustering algorithms have also recently become the focus of practitioners, as these algorithms can both identify and incorporate the fast changes of dynamic financial markets.

The results obtained in this study indicate the possibility of creation a portfolio by using hierarchical clustering algorithms in the portfolio selection task, which would result in improved performances compared to the portfolios of the market indexes in the sense of a stable diversified portfolio that shows a lower risk measured in terms of return volatility.

In the future, we plan to investigate the effect of expanding the group of shares used to create a portfolio on the performance of a clustering algorithm, testing the adequate timing for rebalancing a portfolio, analyzing other strategies for the selection of shares from the cluster which are included in the realization of the portfolio and analyzing the use of more advanced algorithms for clustering which might additionally improve the clustering of financial time series with the aim of portfolio optimization.

REFERENCES

[1] Fernández and S. Gómez, "Portfolio selection using neural networks", Computers & Operations Research, vol. 34, no. 4, pp. 1177-1191, 2007. Available: 10.1016/j.cor.2005.06.017

[2] P. KO and P. LIN, "Resource allocation neural network in portfolio selection", Expert Systems with Applications, vol. 35, no. 1-2, pp. 330-337, 2008.

[3] T. Chang, S. Yang and K. Chang, "Portfolio optimization problems in different risk measures using genetic algorithm", Expert Systems with Applications, vol. 36, no. 7, pp. 10529-10537, 2009. Available: 10.1016/j.eswa.2009.02.062.

[4] S. R Nanda, B., Mahanty and M. K. Tiwari, "Clustering Indian stock market data for portfolio management", Expert Systems with Applications, vol. 37, no.12, pp.8793-8798,2010.

[5] V. Tola, F. Lillo, M. Gallegati and R. Mantegna, "Cluster analysis for portfolio optimization", Journal of Economic Dynamics and Control, vol. 32, no. 1, pp. 235-258, 2008.

[6] D. León, A. Aragón, J. Sandoval, G. Hernández, A. Arévalo and J. Niño, "Clustering algorithms for Risk-Adjusted Portfolio Construction", Procedia Computer Science, vol. 108, pp. 1334-1343, 2017. Available: 10.1016/j.procs.2017.05.185.

[7] A. Jain, "Data clustering: 50 years beyond K-means", Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, 2010. Available: 10.1016/j.patrec.2009.09.011.

[8] A. Singh, A. Yadav and A. Rana, "K-means with Three different Distance Metrics", International Journal of Computer Applications, vol. 67, no. 10, pp. 13-17, 2013. Available: 10.5120/11430-6785.

[9] T. Fu, "A review on time series data mining", Engineering Applications of Artificial Intelligence, vol. 24, no. 1, pp. 164-181, 2011. Available: 10.1016/j.engappai.2010.09.007.

[10] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series". In KDD workshop, vol. 10, no. 16, pp. 359-370, 1994.

[11] J.B Kruskal. and M Liberman, "The symmetric time-warping problem: from continuous to discrete" In: Kruskal J.B., Sankoff D., editors. Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison. Stanford: CSLI Publications; pp. 125–161, 1999.

[12] C. Keating and W. F. Shadwick, "A universal performance measure", Journal of performance measurement, vol.6, no.3, pp. 59-84, 2002.

[13] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis", Communications in Statistics - Theory and Methods, vol. 3, no. 1, pp. 1-27, 1974. Available: 10.1080/03610927408827101.

Efficient Bentness Testing for Cryptographic Function using Statistical Analysis of Binary Functions

Miloš Radmanović[1]

Abstract – **Bent functions are special set of cryptographic binary functions. These functions make very small subset of the total number of bent functions, especially for large number of variables. There is no a formal method for construction of bent functions. Therefore, methods for construction of bent functions are obtained by using the discovery of bent functions within a group of functions. Testing of bentness across a group of functions, even for small numbers of variables, requires a lot of processing time. Thus, this paper proposes an efficient method for bentness testing of binary functions using statistical analysis of binary functions. The standard method for bentness testing is based on the usage of the fast Walsh transform calculations. This method uses conditional testing of the initial set of particular Walsh coefficients. In this paper, the selection of coefficients for conditional testing is determined using statistical analysis of binary functions. Experimental results showed that the proposed method can be efficiently used for bentness testing for the functions of 6 to 10 variables.**

Keywords – **Cryptography, binary function, bent, Walsh transform, statistical method.**

I. INTRODUCTION

Bent functions are binary functions with extreme nonlinearity properties. They are actively studied in cryptography, logic synthesis, switching theory, coding theory, and other areas. They become very important for their intensive applications in cryptography.

Bent functions have specific properties and various characterizations. They ensure the cryptographic effectiveness and they can resist to various cryptanalysis attacks. They exist only for the even number of variables. There is no a precise general definition of the structure of bent functions. Also, there is no a formal method for construction of all bent functions. Thus, during recent years, it has been developed a lot of methods for construction subsets of bent functions that have particular properties. The most known methods for construction of bent functions are based on applying combinatorial, algebraic and permutation methods. For example, combinatorial construction methods are Maiorana-McFarland, partial spreads, Dobbertin, iterative constructions, and etc [1]. The most widely known algebraic constructions are monomial bent functions in the Kasami, Gold, Dillon and Canteaut-Leander case, hyper bent functions, Niho bent functions, and etc [2]. The permutations construction methods describes how new bent functions can be obtained from a known bent function, for example in [3].

However, bent functions obtained in this way constitute a small subset of all bent functions, especially for large number of variables [4]. Further, the subsets of bent function generated by proposed deterministic methods do not provide any, for example, cryptographic quality to constructed bent functions.

In cryptographic applications, bent functions need to be non-deterministic. Therefore, bent functions are determined by using a discovery of random bent functions, but the searching time may become prohibitively large when the number of variables is greater than approximately 12. The existing methods are mainly focused on the reduction of the searching time [5].

The most common characterization of bent functions is the equal absolute values of all coefficients of their Walsh spectra. All coefficients has the absolute value $2^{n/2}$ [1].

Testing of all Walsh coefficients requires computation of all 2^n coefficients and related comparisons. This computation, even for small numbers of variables, requires a lot of processing time. Consequently, the number of n-variable Boolean bent functions is known only for $n \leq 8$ [4]. The general number of bent functions is an open problem. Note that, the number of Bent functions increases rapidly with increasing n. Thus, before testing of all Walsh coefficients, the method for bentness testing uses conditional testing of the initial set of particular Walsh coefficients. This initial set of Walsh coefficients usually includes: the first, second, and middle element of the Walsh spectrum.

The efficiency of using conditional testing for particular Walsh coefficients depends on the structure of binary function for which the Walsh spectrum will be calculated. Therefore, in this paper it is proposed an efficient method for bentness testing of binary functions using statistical analysis of binary functions. The statistical analysis of binary functions shows the most efficient selection of the Walsh coefficients for conditional bentness testing. The statistical method is based on the counting of how many times a particular Walsh coefficient of binary n-variable non-bent functions has the absolute value $2^{n/2}$. For the binary functions of many variables, statistical analysis can be applied on the set of 1 million of functions that satisfied bent criteria. The most common bent criteria is restriction of binary function in the spectral Reed-Muller domain, Since the algebraic degree of an n-variable bent function is less or equal to $n/2$, the number of non-zero elements of Reed-Muller spectrum vector is limited and their positions also in the spectrum vector are restricted. Thus, proposed method uses computation of the fast Reed-Muller transform and the computation of the fast Walsh transform.

These statistical results indicate that there is statistically significant relationship between the selection of the Walsh coefficients for conditional bentness testing and efficiency of bentness testing. Experimental results showed that the proposed method can be efficiently used for bentness testing for the functions of 6 to 10 variables.

[1]Miloš Radmanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.radmanovic@elfak.ni.ac.rs

II. Preliminaries

The Reed-Muller transform [6] represents an important operator for obtaining AND-EXOR expressions of binary functions. The Reed-Muller transform matrix of order n, denoted by $R(n)$, is defined recursively as:

$$R(n) = \bigotimes_{i=1}^{n} R(1), \ R(1) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \quad (1)$$

The Walsh transform [6] is based on a set of orthogonal functions defined by J. L. Walsh which are an extension of a set of functions defined by H. Rademacher. Analogously to previous transforms, the Walsh transform matrix of order n in Hadamard ordering, denoted by $W(n)$, is defined as:

$$W(n) = \bigotimes_{i=1}^{n} W(1), \ W(1) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (2)$$

The spectrum of a binary function f given by truth vector $F = \begin{bmatrix} f(0), f(1)_{....,} f(2^n - 1) \end{bmatrix}^T$ is computed as:

$$S_f = T(n)F, \quad (3)$$

where $T(n)$ is any of the three matrices $R(n)$, or $W(n)$, with computations performed in $GF(2)$ for the Reed-Muller transform, and in the set of rational numbers for the Walsh transforms.

A binary function $f(x_1, x_2, ..., x_n)$ in $(1,-1)$ encoding is called bent if all Walsh coefficients in vector $S_{f,W}$ have the same absolute value $2^{n/2}$ [1].

Algebraic degree of bent functions $f(x_1, x_2, ..., x_n)$ in Reed-Muller spectral domain is at most $\dfrac{n}{2}$ for $n > 2$ [2].

Walsh transform [6] is an important mathematical tool for the analysis of Boolean functions. It can be shown that with $(1,-1)$ encoding of Boolean function values, the Walsh coefficients are even integers in the range 2^{-n} to 2^n. Since the first row of $W(n)$ is equal to constant 1, conditional bentness testing requres calculation of the first Walsh coefficient $S_{f,w}(0)$ expressed as:

$$S_{f,w}(0) = \sum_{i=0}^{2^n - 1} F(i). \quad (4)$$

Analogously to previous, second row $W(n)$ takes the successively values +1 and -1. The conditional bentness testing requires calculation of the second Walsh coefficient $S_{f,w}(1)$ expressed as:

$$S_{f,w}(1) = \sum_{i=0}^{2^{n-1} - 1} (F(2i) - F(2i+1)). \quad (5)$$

Since the middle row of $W(n)$ takes the value +1 in the first half and -1 in the second half, the conditional bentness testing requires calculation of the middle Walsh coefficient $S_{f,w}(2^{n-1})$ expressed as:

$$S_{f,w}(2^{n-1}) = \sum_{i=0}^{2^{n-1} - 1} F(i) - \sum_{i=2^{n-1}}^{2^n - 1} F(i). \quad (6)$$

The calculations of first, second and middle Walsh coefficient are exploited in the standard method for the bentness testing of the binary function in the Reed-Muller spectral domain in order to avoid computation of all Walsh coefficients for bent testing.

The recursive definition of the Reed-Muller and the Walsh transform matrices, expressed in Eq. (1), and Eq. (2) respectively, is the fundamental for the definition of fast Reed-Muller and the fast Walsh transform algorithm similar to a fast Fourier transform (FFT) algorithm [6].

The computation of the fast transform algorithm consists of the repeated application of the same "butterfly" operations determined by the basic transform matrices. Figure 1 shows the "butterfly" operations for the Reed-Muller and the Walsh transform matrices [6].
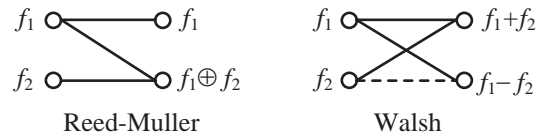
Fig 1. The "butterfly" operations for the Reed-Muller and the Walsh transform matrices

The "butterfly" operations [6] are performed in each step over a different subset of data. Fig. 2 shows the flow graphs of the fast Walsh transform algorithm for computation of the Walsh spectrum of a three-variable logic function f given by the truth-vector $F = \begin{bmatrix} f(0), f(1)_{....,} f(7) \end{bmatrix}^T$.
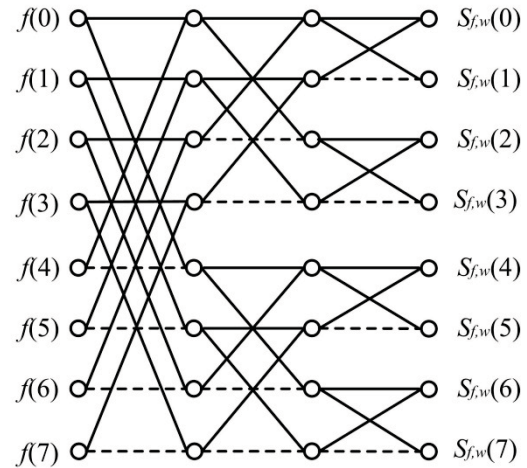
Fig. 1. The flow graph of the fast Walsh transform algorithm of the Cooley-Tukey type for the computation of the Walsh spectrum of a three-variable Boolean function.

This algorithm is highly exploited for benthess testing in the algortihm for discovering bent functions in the Reed-Muller domain.

III. STATISTICAL ANALYSIS OF THE VALUES OF WALSH COEFFICIENTS

The statistical analysis of the Walsh coefficients values is based on the counting of how many times a particular Walsh coefficient of binary n-variable non-bent functions has the absolute value $2^{n/2}$.

For example, for a binary function of 6 variables, there exist 64 Walsh coefficients. Table 1 shows example of statistical analysis of how many times a particular Walsh coefficient of 6-variable non-bent functions has the absolute value $2^{6/2} = 8$.

TABLE I
EXAMPLE OF THE STATISTICAL ANALYSIS OF HOW MANY TIMES A PARTICULAR WALSH COEFFICIENT OF 6-VARIABLE NON-BENT FUNCTIONS HAS THE ABSOLUTE VALUE 8

Index of Walsh coefficient	Equal to $\left\lvert 2^{6/2} \right\rvert$ for non-bent functions	Index of Walsh coefficient	Equal to $\left\lvert 2^{6/2} \right\rvert$ for non-bent functions
1	57108751	33	61302754
2	62280216	34	62684614
3	62280758	35	62685119
4	63193292	36	63201318
5	62156341	37	62658738
6	62818400	38	63216388
7	62818979	39	63216660
8	63104008	40	62467837
9	61287800	41	62604235
10	62656734	42	63299127
11	62657004	43	63299512
12	63192238	44	62534014
13	62736518	45	63376543
14	63201946	46	62613453
15	63201937	47	62613447
16	62440487	48	61723928
17	61302747	48	62616310
18	62684605	50	63316595
19	62685140	51	63316740
20	63201319	52	62545379
21	62658714	53	63397936
22	63216367	54	62624877
23	63216671	55	62625138
24	62467828	56	61741281
25	62604222	57	63475839
26	63299110	58	62690483
27	63299519	59	62690478
28	62534001	60	61805616
29	63376520	61	62714230
30	62613426	62	61882980
31	62613444	63	61882810
32	61723904	64	60913426

The statistical analysis is applied on the testing of set of 1 million binary functions that satisfied bent criteria in Reed-Muller domain. For a binary function of 6 variables, it is evident that statistical analysis shows that first Walsh coefficient for non-bent functions in the least number of cases have absolute value 8. It means that first Walsh coefficient is the most optimal solution for conditional bentness testing. The next Walsh coefficient, that in the least number of cases has absolute value 8, is the last Walsh coefficient. Note that for binary function of 6 variables the difference between statistical values of the first and the last Walsh coefficient is not so high, but it will increases rapidly with increasing number of variables. The next Walsh coefficient, that in the least number of cases has absolute value 8, is the middle Walsh coefficient. In the case of functions of 6 variables it is Walsh coefficient that has index 33.

The statistical analysis of the Walsh coefficients values showed that, initial set of Walsh coefficients need to include: the first, the last, and the middle Walsh coefficient for conditional testing, respectively. It means that before testing of all Walsh coefficients, the method for bentness testing need to perform conditional testing of the first, then testing of the last and finally of the middle Walsh coefficient. The efficiency of the proposed method for bentness testing will be experimentally tested by using a discovery of random bent functions.

IV. METHOD FOR BENTNESS TESTING OF BINARY FUNCTIONS

An outline of the proposed method for the bentness testing of binary functions is given as Algorithm 1.

Algorithm 1 Bentness testing

1: (1,−1) encoding of a binary function function
2: Bent testing of the first Walsh coefficient, if failed go to the step 8.
3: Bent testing of the last Walsh coefficient, if failed go to the step 8.
4: Bent testing of the middle Walsh coefficient, if failed go to the step 8.
5: Transition from truth vector to Walsh spectrum using the fast Walsh transform.
6: Bent testing of all Walsh coefficients, if failed go to the step 8.
7: Return bent function found.
8: Return bent function not found.

Note that in previous version of bentness testing methods [7], [8], the conditional testing was done using the first, second, and last coefficient of the Walsh spectrum. The reason for this is that the first, middle and last Walsh coefficient have simple formula for calculation and also the implementation is easy.

The proposed method, using results from statistical analysis of the values of Walsh coefficients, made correction in previous approach in step 2. Instead of calculation and conditional testing of the second Walsh coefficient, now the method uses calculation and conditional testing of the last Walsh coefficient.

V. Experimental Results

In this section, the proposed method for bentness testing will be experimentally tested within algorithm for discovering of bent functions with predefined number of non-zero coefficients in Reed-Muller domain. A short outline of the algorithm for discovering bent functions in the RM domain is given as Algorithm 2. The "Bentness testing" algorithm is given as Algorithm 1 within previous section.

Algorithm 2 Discovery of bent function in RM domain

1: Set the number of function variables n and the number of non-zero RM coefficients

2: Random generation of the possible non-zero coeficients in the RM spectrum.

3: Conversion of the random RM spectrum to the binary vector by using the fast RM transform algorithm.

4: If the "Bentness testing" of binary vector is not successful, then go to step 1.

5: Obtain a bent function.

In this paper, the same algorithm for discovering of bent functions in Reed-Muller domain is implemented with standard bentness testing algorithm and also it is implemented with the proposed bentness testing algorithm which is derived form statistical analysis of Walsh coefficients. Experimental results compare the number of discovered functions for the predefined number of non-zero coefficients in RM spectrum. Comparison of these numbers is motivated by increasing the number of discovered bent function when the proposed method is used. Note that discovering of bent function is faster when the number of non-zero coefficients is small.

For experimental purposes, it is developed C++ implementation of Algorithms 1 and 2. The computations are performed on an Intel i7 CPU at 3.66 GHz with 12 GBs of RAM.

Table 2 shows the total number of discovered functions per 1 second for the predefined number of non-zero coefficients in RM domain for standard and proposed method. The data in tables are sorted in the increasing order of the number of non-zero coefficients and the number of variables.

It should be noticed that the number of discovered bent functions of 6, 8, and 10 variables for proposed method is about 3% larger than using the standard method. Note that for the largest binary functions, experiments were not performed, due to very long CPU computation time.

VI. Conclusion

This paper proposes an efficient method for bentness testing of binary functions using statistical analysis of the values of Walsh coefficients of binary functions. The standard method for bentness testing before performing the fast Walsh transform calculations uses conditional testing of the initial set of particular Walsh coefficients. The selection of coefficients for conditional testing within standard methods is determined by complexity of calculations.

TABLE II
NUMBER OF DISCOVER FUNCTIONS PER 1 SECOND FOR STANDARD AND PROPOSED METHOD

Num. of function variables	Num of non-zero RM coefficients	Number of discovered bent finctions per 1 s	
		standard	proposed
6	10	650	671
6	20	635	653
6	30	627	644
8	10	0.691	0.712
8	20	0.553	0.564
10	10	0.031	0.032
10	20	0.035	0.036

The conditional testing of the first, the second and the middle Walsh coefficients are exploited in the standard method for bentness testing. The statistical analysis shows that for conditional testing it is more efficient to use the last Walsh coefficient instead of the second. Experimental results confirm that exploiting proposed conditional testing can help to improve the computation performances by 3%.

We can conclude that, when processing time is a critical parameter, the proposed method for bentness testing should be performed. Future work will be on improving other aspects of the methods for bentness testing.

Acknowledgements

References

[1] N. Tokareva, *Bent Functions, Results and Applications to Cryptography*, Academic Press, 2015.

[2] S. Mesnager, *Bent Functions, Fundamentals and Results*. Springer International Publishing, 2016.

[3] J. Seberry, X. Zhang, "Constructions of bent functions from two known bent functions", Australasian Journal of Combinatorics. vol. 9, pp. 21-34, 1994.

[4] P. Langevin, G. Leander, "Counting all bent functions in dimension eight 99270589265934370305785861242880", Designs, Codes and Cryptography vol. 59, pp. 193-201, 2011.

[5] A. Grocholewska-Czurylo, J., Stoklosa, "Generating bent functions", Proc. Advanced Computer Systems Eighth International Conference, ACS2001, pp. 361-370. Springer US, Mielno, Poland, 2002.

[6] M. A. Thornton, R. Drechsler, D. M. Miller, *Spectral Techniques in VLSI CAD*, Springer, 2001.

[7] M. Radmanović, R. Stanković, "Random generation of bent functions on multicore CPU platform". Proc. ICEST 2016, pp. 239-242. Ohrid, Macedonia,2016.

[8] M. Radmanović, "Efficient random generation of bent functions using GPU platform", Proc. 12th Int. Workshop on Boolean Problems, IWSBP2016, pp. 167-173. Freiberg, Germany 2016.

Brain Rhythms, Pascal Triangle, and Brain-Computer Interface

Stevo Bozinovski[1] and Adrijan Bozinovski[2]

Abstract – **The ranges of brain rhythms are empirically defined and a theoretical process is needed to offer definition of those ranges. The paper makes an observation that the central frequencies of brain rhythms can be defined with a integer sequence from the Pascal triangle. That way the paper defines a theoretical process underlying the ranges of brain rhythms.**

Keywords – **brain rhythms, central frequencies Pascal Triangle, brain-robot interface**

I. INTRODUCTION

At the beginning of 20th century two works pointed toward electrical activity of the animal and human brain. In 1913 Pravdich-Nemensky [1] showed electrical brain activity an animal and Berger showed a human brain activity [2]. Berger named it Elektrenkephalogramm which today in English literature is translated as electroencephalogram (EEG). He also mentioned that in an EEG a certain frequency rhythm is dominant under certain conditions, and named it alpha rhythm. Since then various other rhythms were observed, and named delta theta, beta, and gamma, with additional variants.

Brain rhythms are often used in Brain-Computer Interface (BCI), i.e. in controlling external objects using EEG signals. The Brain-Computer Interface challenge was stated by Vidal in 1973 [3]. Vidal proposed the use of EEG rhythms as well as evoked brain potentials as features of EEG that will be used to control external devices. He suggested that the Contingent Negative Variation (CNV) potential be used in a BCI. After Vidal's challenge in 1973, the first five applications of BCI were: (1) control of a cursor-like object on a screen using the Visual Evoked Potential (VEP) [4]; (2) control of a buzzer using the CNV potential [5-9]; (3) control of a physical object (a robot) using the alpha rhythm [10-15]; and (4) writing text on a screen using the P300 potential [16].

In this paper we will consider brain rhythms and focus on their ranges and central frequencies.

II. BRAIN RHYTHMS

Brain rhythms are signals obtained from a human electroencephalogram (EEG) after filtering certain frequencies with band pass filters.

Figure 1 shows filtering an EEG with band pass filters and obtaining brain rhythms. Figure 1 shows 6 signals.

[1]Stevo Bozinovski is with the Department of Mathematics and Computer Science, South Carolina State University, Orangeburg, SC, USA. Email: sbozinovski@scsu.edu

[2]Adrijan Bozinovski is with the School of Computer Science and Information Technology, University American College, Skopje, Macedonia. Email: bozinovski@uacs.edu.mk

Fig. 1. EEG rhythms obtained in our lab

The signal on the top of the Figure 1 is the EEG signal obtained from a subject. It is the unprocessed (i.e., "raw") EEG, whereas the signals below it are the signals obtained when the "raw" EEG is band pass filtered in the alpha, beta, delta, and theta frequencies, respectively. The definitions of the bands may vary, and in Figure 1 the definitions are: delta: 1-5Hz, theta: 4-8 Hz, alpha: 8-13 Hz, beta: 13-30Hz [17].

Various interpretations are given for mental states at which a brain rhythm is dominant. The most well known interpretation is that the alpha rhythm appears in a wake brain but relaxed from doing a particular task. For example, when the eyes are closed and the visual cortex is not active, this rhythm is dominant at the back of the head. Placing electrodes in the occipital and parietal area might record an EEG where the alpha rhythm is visually seen from the raw EEG, without any signal processing. A variant of the alpha rhythm is the mu rhythm, which has the same frequency band as the alpha rhythm but is recorded from the brain sensorimotor area. It is dominant when there is no movement (imagined or actual) of the limbs such as hands and legs. The beta rhythm is dominant when the brain is awake and engaged. The border between the alpha and beta can be used as a switch for application of a BCI. For example, the first control of a physical object (a robot) using BCI was based on the alpha/beta switch [9-11, 18].

In this paper we will not discuss interpretations of other brain rhythms. However, we will make an observation that the delta rhythm is simply a signal obtained by a low pass filter from an EEG. The obtained signal might be used as an average activity of an EEG. It gives the slow brain potential around which the EEG signal stochastically oscillates.

III. DEFINING BRAIN RHYTHMS. OUR CONTRIBUTION

The brain rhythms are defined by their frequency bands. The frequency ranges are defined differently depending on the author and application. Here is an example of definition of the frequency bands of the brain rhythms [19]:

delta:	0-4Hz
theta	4-8 Hz
alpha	8-13 Hz
beta	13-30 Hz
gamma	30-100 Hz

So a theoretical process defining frequency ranges is needed. In this paper we propose a theoretical integer sequence which defines the central frequencies and ranges of brain rhythms. Some previous works[20][21] on integer sequences by the second author are registered in Online Encyclopedia of Integer Sequences (OEIS). Here we point out toward a integer sequence derived from the Pascal triangle.

The Pascal triangle is a pattern of generating integers. It starts with number 1 at level 0. Level 1 contains the sequence (1,1). The process is shown in Figure 2.

```
n = 0                        1
n = 1                     1     1
n = 2                  1     2     1
n = 3               1     3     3     1
n = 4            1     4     6     4     1
n = 5         1     5    10    10     5     1
n = 6      1     6    15    20    15     6     1
n = 7    1     7    21    35    35    21     7     1
n = 8  1    8    28    56    70    56    28     8     1
n = 9 1   9   36   84   126   126   84    36    9     1
etc.
```

Fig. 2. The Pascal triangle

The Pascal triangle can be constructed using the rule that each number in a level is a sum of the two number above it , and each level starts and ends with the number 1. The Pascal triangle generates various well known integer sequences, for example positive integers $\{n\}_1^\infty$, triangular numbers $\{n(n+1)/2\}_1^\infty$ (e.g. [22]) etc.

Well known interpretations of the rows of a Pascal triangle is that they represent coefficients of the expansion of $(a+b)^n$ for n=0,1, etc. .

Considering the Pascal triangle, we made an observation that the central frequencies of the brain rhythms can be defined by the middle values of the Pascal triangle. Figure 3 shows our definition of central frequencies of brain rhythms.

```
n = 0                        1
n = 1                     1     1
n = 2                  1     2     1
n = 3               1     3     3     1
n = 4            1     4           4     1
n = 5         1     5    10    10     5     1
n = 6      1     6    15    20    15     6     1
n = 7    1     7    21    35    35    21     7     1
n = 8  1    8    28    56    70    56    28     8     1
n = 9 1   9   36   84   126   126   84    36    9     1
```
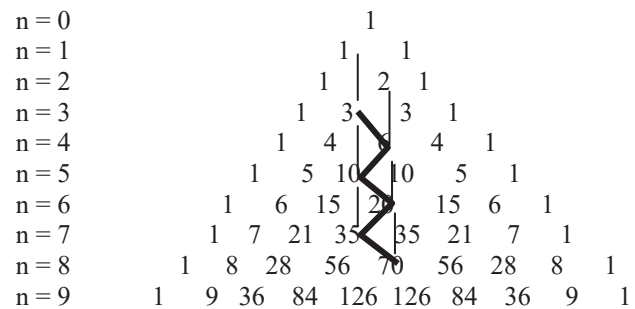
Fig. 3. Our definition of central frequencies and ranges of brain rhythms

So we define brain rhythms by their central frequencies and deviations. Each central frequency with its deviation can be considered as a fuzzy number. The following is our definition of brain rhythms obtained from Figure 3.

rhythm	central frequency $\pm$ variation	frequency range
delta	3 ± 1 Hz	2-4 Hz
theta	6 ± 2 Hz	4-8 Hz
alpha:	10 ± 3 Hz	7-13 Hz
beta	20 ± 6 Hz	14-26 Hz
gamma1	35 ± 10 Hz	25-45 Hz
gamma2	70 ± 20 Hz	50-90 Hz

The central frequency for each band is taken from the middle values of the Pascal triangle. That defines the sequence 3,6,10,20,35,70. This proposes a mathematical phenomenon which defines the central frequencies of brain rhythms.

Table I shows comparison between our theoretically obtained ranges and other authors' empirical ranges of brain rhythms.

Table I. Comparison of our theoretical ranges with empirical ranges used by other authors.

author	year	delta	theta	alpha	beta	gamma	Lgamma	Hgamma
Kelly et al.[23]	2005			8-14	18-26			
Jia et al.[24]	2011	0-4	4-8	8-12	12-30		30-80	>80
Kumar [19].	2012	0-4	4-8	8-13	13-30	30-100		
Biopac[17]	2013	1-5	4-8	8-13	13-30			
Wiki.alpha [25]	2019	0.5-3	4-7	8-12	16-31	32-100		
Wiki.eeg [26]	2019	0-4	4 - 8	8 -14	> 14			
Wiki..eeg [27]	2019	0-4	4-7	7-13	14-30	30-100		
Wiki.neur.osc[28]	2019	1-4	4-8	8-12	13-30		30-70	70-150
Perotti [29]	2019	4-12					25-45	
This paper.	2019	2-4	4-8	7-13	14-26		25-45	50-90

As can be seen from Table I, our theory about a sequence from Pascal triangle defining central frequencies of brain rhythm, fits well into the empirical definitions. Also our definition states clearly that there are two gamma bands, as some authors defined empirically.

According to our theory, the deviations around the central frequency of a band are also defined with integers from Pascal triangle. The definition does not insist that deviations are fixed. There is a possibility that deviations are defined with

different values from Pascal triangle. For example the delta band can be defined as 3+2 Hz (range 1-5Hz) and Upper Gamma band can be defined as 70+35Hz (range 35-105Hz). However the central frequencies are defined from middle values of rows of the Pascal triangle.

In addition to having mathematical background, this definition has a cognitive background. It is related to a very important *cognitive operation of choice*. Indeed the numbers inside the Pascal triangle represent number of combinations (choices) when one chooses k elements from a set of n elements, or subsets from a set. Most often used notation is "n over k" which is read "n choose k". We can use a function named *choices(n, k)*. For example if the set has 4 elements and we need to chose 2 elements, the Pascal triangle gives the value *choices(4,2)* = 4!/(2!2!) = 6. The n-th row of the Pascal triangle gives number of choices for each k from 0 to n.

IV. Conclusion

This paper discovers a relation between central frequencies of brain rhythms and a sequence obtained from the middle values of the Pascal triangle. The sequence is 3,6,10,20,35,70. The deviations from central frequencies are also defined from the numbers of the Pascal triangle and they define ranges of brain rhythms. This work makes a contribution toward finding a theoretical process underlying brain rhythms. The knowledge can be used in brain-computer interface and in other applications of EEG.

References

[1] V. Pravdich-Nemensky "Ein Versuch der Registrerung der elektrischer Gehirnerscheinungen" Zentralblatt fuer Physiologie, vol. 27, pp. 951-960, 1913.

[2] H. Berger, "Ueber das Elektrenkephalogramm des Menschen.". Archiv fuer Psychiatrie, vol. 87, pp. 35-, 1929.

[3] J. Vidal, "Toward direct brain-computer communication". Annual Review of Biophysics and Bioengineering, pp. 157-180, 1973.

[4] J. Vidal, "Real-time detection of brain events in EEG". Proceedings of the IEEE, vol. 65, pp. 633– 641, 1977.

[5] L. Bozinovska, M. Sestakov, G. Stojanov, S. Bozinovski, "Intensity variation of the CNV potential during the biofeedback training guided by a personal computer" [abstract]. Neurologija; suppl. 2, vol. 37, p. 76, 1988. (In Serbian)

[6] L. Bozinovska, S. Bozinovski, G. Stojanov, M. Sestakov, "Introduction of biofeedback in the CNV paradigm", Proc Conference ETAN, Novi Sad, pp. XII.93-98, 1989. (In Serbian)

[7] L. Bozinovska, G. Stojanov, M. Sestakov, S. Bozinovski, "CNV pattern recognition - A step toward cognitive wave observation". In: Torres L, Masgray E, Lagunas M, editors. Signal Processing: Theories and Applications. Proc Fifth European Signal Processing Conference (EUSIPCO 90); Barcelona. Amsterdam: Elsevier Science Publishers, pp. 1659-1662, 1990.

[8] L. Bozinovska, S. Bozinovski, G. Stojanov, "Electroexpectogram: Experimental design and algorithms", Proc. IEEE International. Biomedical. Engineering Days, Istanbul, pp. 58–60, 1992.

[9] S. Bozinovski, L. Bozinovska, "Brain-Computer Interface in Europe: The thirtieth anniversary" Automatika, vol. 60, no. 1, pp. 36-49, 2019.

[10] S. Bozinovski, M. Sestakov, L. Bozinovska, "Control of a mobile robot using alpha rhythm from a human brain", Proc. Symposium JUREMA, Zagreb, pp. 247-249, 1988 (In Croatian)

[11] S. Bozinovski, M. Sestakov, L. Bozinovska, "Using EEG alpha rhythm to control a mobile robot", In: Harris G., Walker C. editors. Proc. 10th Annual Conference of the IEEE Engineering in Medicine and Biology Society, track 17, Biorobotics, New Orleans. LA. vol. 10, pp. 1515–1516, 1988.

[12] S. Bozinovski, M. Sestakov, G. Stojanov, L. Bozinovska, "Bioelectric control of mobile robots", Proc. 6th Yugoslav symposium for applied robotics and flexible automation, pp., 237-242, Novi Sad, 1989 (In Macedonian)

[13] S. Bozinovski, "Mobile robot trajectory control: From fixed rails to direct bioelectric control", In: Kaynak O, editor. Proc. IEEE International Workshop on Intelligent Motion Control, Istanbul, vol 2, pp. 463-467, 1990.

[14] Lebedev M. Augmentation of sensorimotor functions with neural prostheses. Opera Medica and Physiologica . 2 (3):211–227, 2016.

[15] M. Lebedev, M. Nicolelis, "Brain-machine interfaces: from basic science to neuroprostheses and neurorehabilitation", Physiological Review vol. 97, pp. 737-867, 2017.

[16] L. Farwell, E. Donchin, "Talking off the top of your head: a mental prosthesis utilizing event-related brain potentials", Electroencephalography and Clinical Neurophysiology, vol. 70, pp. 510–523, 1988.

[17] Biopac. *Biopac Student Lab*. Biopac Systems, Inc, 2013.

[18] S. Bozinovski, A. Bozinovski, "Mental States, EEG Manifestations, and Mentally Emulated Digital Circuits for Brain-Robot Interaction", IEEE Transactions on Autonomous Mental Development, vol. 7, no. 1, pp. 39-51, 2015.

[19] S. Kumar, P. Bhuvaneswari, "Analysis of Electroencephalography (EEG) signals and its categorization - A study". Procedia Engineering, vol. 38, pp. 2525-2536, 2012.

[20] A. Božinovski, G. Tanev, Sequence A277267 in the *On-Line Encyclopedia of Integer Sequences*, 2016, (accessed July 7, 2019) Available from: https://oeis.org/A277267.

[21] A. Božinovski, Sequence A279521 in the *On-Line Encyclopedia of Integer Sequences*, 2016, (accessed July 7, 2019) Available from: https://oeis.org/A279521.

[22] S. Bozinovski, "An overlooked identity in number theory" Contributions of Macedonian Academy of Science, Section of Natural, Mathematical and Biotechnical Sciences, vol. 35, no. 1, pp. 53–56, 2014.

[23] S. Kelly, E. Lalor, R. Reilly, J. Foxe. Independent brain-computer interface control using visual spatial attention-dependent modulations of pareto-occipital alpha. Proc. 2nd International IEEE EMBS Conference on Neural Engineering. Arlington, VA, 2005

[24] X. Jia, A.Kohn, Gamma Rhythms in the Brain. PLoS Biology, 9 (4): 1-4, 2011|

[25] https://en.wikipedia.org/wiki/Alpha_wave, June 2019

[26] https://en.wikipedia.org/wiki/Electroencephalography, p.16, June 2019

[27] https://en.wikipedia.org/wiki/Electroencephalography, p.16-17, June 2019

[28] https://en.wikipedia.org/wiki/Neural_oscillation June 2019

[29] L. Perotti, J. DeVito, D. Bessis, Y. Dabaghian, Discrete Structure of the Brain Rhythms. Nature 9:1105, 2019

Security and Standardization at E-learning platforms

Dragan Zlatkovic[1], Nebojsa Denic[2], Milos Ilic[3] and Milena Petrovic[4]

Abstract – **The process of developing e-learning is based on standards in the field of information security (ISO) and learning technologies (IEEE, IMS), and criteria for quality assurance have been taken as input parameters. The overall development was carried out through the PDCA cycle and with upgraded LTSA architecture.**

Keywords – **E-learning, standards, security, PDCA cycle, LTSA.**

I. INTRODUCTION

Quality assurance in e-learning is a natural continuation of education quality in general. Thanks to the specificities of this type of teaching/learning, special procedures and schemes for quality assurance have been developed at different levels: at the level of educational software, at the level of the individual course institution, at the level of the study program or institution. Because quality of e-learning is associated with security, these schemes also address security issues through their criteria.

Regardless of the purposes of the information system used, there are general safety recommendations and standards that are defined by specific documents. ISO International Standardization Organization has provided best practices for information security management through its set of ISO 27K standards [1].

When creating any model, it is necessary to take into account the standards that deal with the given area in general (in this case, the information security), as well as standards dealing with security in a specific field (e-learning), as part of a specific form of regulation. In this paper is a brief overview of standardization in security and relevant standardization in e-learning.

II. STANDARDISATION OF SECURITY

Official standardization is presented through a set of ISO/IEC 27K family documents [2]. Some of the essential standards related to "information technology - security techniques" are:

[1]Dragan Zlatkovic is with the Faculty of Mathematics' and Computer Sciences at Alfa BK University in Belgrade, 3 Palimira Toljatija, Belgrade 11070, Serbia, E-mail: dragan.zlatkovic@sbb.rs

[2]Nebojsa Denic is with the Faculty of Natural Sciences at University of Pristina, 29 Ivo Lola Ribara, Kosovska Mitrovica 38220, Serbia, E-mail: nebojsa.denic@pr.ac.rs

[3]Milos Ilic is with the Faculty of Information Technologies at Alfa BK University in Belgrade, 3 Palimira Toljatija, Belgrade 11070, Serbia, E-mail: milos.ilic@alfa.edu.rs

4 Milena Petrovic is with the Faculty of Natural Sciences at University of Pristina, 29 Ivo Lola Ribara, Kosovska Mitrovica 38220, Serbia, E-mail: milena.petrovic@pr.ac.rs

- ISO/IEC 27000 – Information security management systems — Overview and vocabulary;
- ISO/IEC 27001 - Information technology - Security Techniques - Information security management systems-Requirements;
- ISO/IEC 27002 – Code of practice for information security controls - virtually a detailed catalog of information security controls that might be managed through the ISMS, etc.

The security management philosophy in the 2005 standard version relied on the PDCA methodology, Fig. 1, while in the new version (2013), this approach is placed in the other plan, with an emphasis on the quality of organization's security and compatibility with different standards, especially with ISO 9000.

Entry to this circle is the requirements and expectations of interested parties-stakeholders, and the exit is security management.
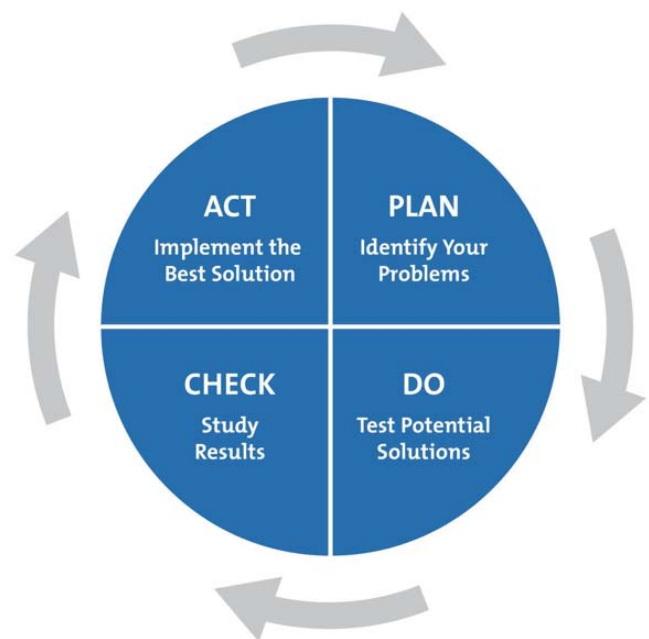
Fig. 1. The Plan–Do–Check–Act (PDCA) Quality Cycle

The standard ISO/IEC 27002 (2015) presents a particular set of good practice recommendations, which can be applied in a wide range of information systems, including an e-learning system [1].

The first step is the risk assessment, which establishes, quantifies and prioritizes the risks to the given criteria for accepting risks and goals that are important for the organization itself. The results should serve as a guide to further management actions and prioritization in managing security risks and implementing security measures [3].

Fig. 2, presents the elements that figure in the 27K standards.
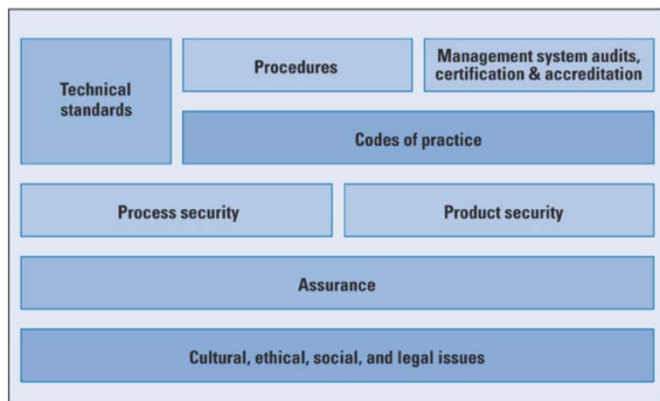
Fig. 2 – Information Security Management Elements (according to ISO Standard)

The risk can be assessed at the level of the whole organization, its parts, specific system components or even at the level of the service. After carrying out a risk assessment, it is necessary to define the appropriate risk management mode [4]:

- apply the appropriate protection measure;
- accept risk;
- avoid risk by not carrying out activities that would lead to it;
- transfer risk to another (suppliers, insurance company).

Through eleven classes of protection measures, many safety-relevant categories have been defined, such as, for example, "Information security awareness, education, and training," and given is the goal of the protection measure and one or more protection measures applied to achieve the goal. ISO "protection measure" is called a control [5].

Guidelines for auditors on information security controls provide instructions to auditors to review the implementation and the correctness of the protection measures, including verifying the technical compatibility of these measures and the established standards of the organization. If the protection process covers the e-learning system, these recommendations can be further exploited precisely to verify that system or its protection measures [4].

III. INFORMATION SECURITY WITHIN E-LEARNING STANDARDS

Some organizations deal with the creation of standards and recommendations in the field of e-learning. Among the most important are the Learning Technology Standardization Committee (LTSC), the IMS Global Learning Consortium, the Aviation Industry CBT Committee (AICC) and the U.S. Department of Defense Advanced Distributed Learning (ADL).

The e-learning standards generally consist of several parts [6]. The first describes the data model and provides norms and content abstracts, the other is a formal description (mainly

through XML), and the third represents the API, an interface for collaboration with other systems.

A very detailed overview of e-learning standards is given in [7].

A. IEEE 1484

The most famous and most detailed standard is IEEE 1484 [8]. Within the segment 1484.1-2003 of the IEEE Standard for Learning Technology - Learning Technology Systems Architecture (LTSA), a high-level e-learning model is also given. The task of this architecture is to provide a high-level framework for the development of various e-learning systems and facilitate their evaluation and comparison.

In addition to the IEEE 1484 working group, LTSA was developed through the activities of various other organizations [8]:

- DoD Advanced Distributed Learning (ADL)
- Aviation Industry Computer-Based Training (CBT) Committee (AICC)
- American National Standards Institute, Information Infrastructure Standards Panel (ANSI IISP)
- Architecture Abstraction Hierarchy Reference Model, by Frank Belz, Dan Suthers, Tom Wheeler.
- Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE)
- Tool/Agent Communication, by Steven Ritter - Carnegie Mellon University (CMU)
- Common Object Request Broker Architecture of Object Management Group (OMG), Medical Informatics (CORBAMED)
- Apple Computer's Educational Object Economy (EOE)
- Educom's Instructional Management Systems Project (IMS)
- International Standards Organization - International Electrotechnical Committee, Joint Technical Committee 1 - Information Technology, Business Team on Electronic Commerce (ISO-IEC JTC1 BT-EC)
- Global Information Infrastructure - Standards Roadmap: a catalog and analysis of GII standards (ISO-IEC JTC1 GII)
- Cultural Adaptability Workshop (ISO-IEC JTC1 CAW)
- Standards Operations Roundtable (ISO-IEC JTC1 SORT).

LTSA components are [8]:
- Processes: Learner, Evaluation, System Coach, Delivery process;
- Flows: Behavior, Assessment, Performance, Query Index, Content Index, Locator Index, Learning Content, Multimedia, Learning Style;
- Data storage: Records Database, Knowledge Library.

The "Learner Process" is an abstraction of students and can represent an individual student, a group of learners who learn individually, a group that determines collaboratively, and so on.

The student receives a multimedia setting, and his behavior is watched. At this level of abstraction, multimedia and observed behavior are shown separately. However, real implementations usually combine these elements into one or more interface modules, such as window systems, presentation in a web browser, specialized applications, etc.

The learning style is established in cooperation with the System Instructor.

The process of behavior represents the student's coded behavior, from the student to the evaluation process. In this process, practice is embedded in an appropriate context by matching learning content with a specific range of behavioral responses.

Encoding behavior is how behavior information is organized, for example, key press, mouse click, voice command, etc. Coding represents the student's behavior independent of the content of learning [3].

The evaluation process results in the evaluation information and sends the evaluation details to the System Instructor. The evaluation process creates information about achievement, which is stored in the Records Database.

The evaluation process uses the learning content object to provide a context student's behavior and determine the appropriate evaluation.

The evaluation process sends out information about achievement and keeps the database (for example, "question 12, answered correctly – student spent 57 seconds") [5].

The knowledge library keeps different data (tutorials, tools, experiments ...). Related material is provided based on the Context Index (metadata). Based on these metadata, the delivery process supplies Content [9].

The delivery process transforms content acquired from the knowledge library into the appropriate multimedia form.

LTSA implementation in online learning

As already stated, LTSA is a generic architecture. In the concrete implementation (for example, in e-learning), mapping of individual components into appropriate representations of the electronic learning system is achieved. Fig. 3, shows that the user - the learner is mapped into the LTSA learner component and a database of courses in the knowledge library.
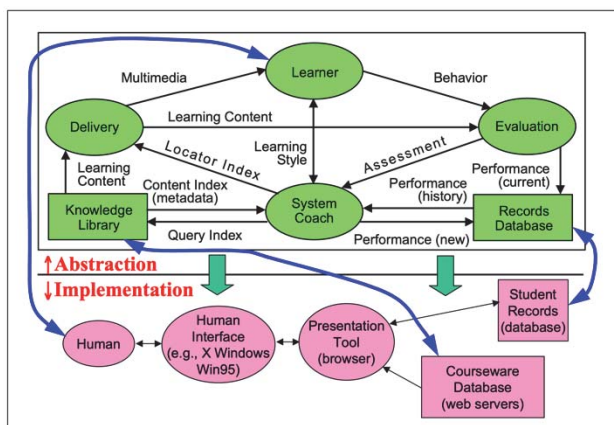

Fig.3. Mapping the LTSA components of the online learning scenario [8]

Security elements in IEEE 1484

Table I briefly shows the security related features of IEEE 1484 standard [10].

TABLE I
SECURITY FEATURES DEFINED IN IEEP P1848

Model	Specification	Model	Specification
Session – View Security	D	Non-Repudiation	I
Security Parameter Negotiation	D	Repudiation	I
Security Extension	D	Privacy	N
Access Control	D	Confidentiality	N
Identification	I	Encryption	N
Authentication	O	Data Integrity	N
De-identification	O	Validation of Certificates	N
Authorization	I	Digital Signature	N
Delegation	I		

D – Defined: the model and/or requirements are defined or provided.
I – Implementation-dependent: the detailed methods are depended on detail implementations.
O – Outside the scope: the methods are outside the standard.
N – Non-specified: the standard doesn't specify the model and requirements.

Within the LTSA standard, a particular part - Public and Private Information (PAPI) deals with the student himself, that is, the syntax and semantics of his information and forms of access. The elements deal with skills, abilities, contact information, learning style, performance, personal portfolio, security parameters. The standard provides different types of information review, according to the appropriate roles: teacher, student, parent, employer [11].

PAPI specifically addresses the issue of access to external repositories and provides for the creation of the so-called the surrogate of the identifier, by which the student registers to the external service, thereby eliminating the possibility of its monitoring [7].

B. IMS Global Learning Consortium – LIP

IMS Global Learning Consortium is an organization that develops open e-learning recommendations, addressing key issues and challenges in distributed learning environments across a range of specifications, including metadata, enterprise specifications, learning package specifications, tests and questions [12]. Among other things, the IMS specification Learner Information Package (IMS LIP) addresses the interoperability of the student's information system with other learning systems [13]. Student data is designed as a collection of student data and typically contains a record of the level of education, a learning diary, a lifelong learning diary, and so on.

IMS LOM deals with privacy issues in Version 1.0, and in this segment attaches great importance. There are two mechanisms proposed [13]:

- Support for the inclusion of data describing the level of privacy, access rights, and data integrity. These data are defined through a special meta-structure.

- Support for user data that would be used to secure data transfer. This information is defined as a student's security key (learner security key).

Security keys constitute a structure that keeps different keys providing communication between learners and systems and e-learning services.

Unlike PAPI, IMS besides data also provides a model or metadata that supports modeling. Also, this model itself is extremely flexible.

C. EDUCAUSE – Internet2 EduPerson

EduPerson's specifications came from the Internet2 Initiative and EduCause [14]. It represents an attempt to model of typical student information, as well as other entities in the institution, as part of the Lightweight Directory Access Protocol Scheme (LDAP), which would facilitate the creation of institutional directories by providing appropriate templates.

EduPerson is not broad as PAPI, and since it covers a more whole class of users than students (employees, alumni), it provides less specific information: name, nickname, organization, contact information, photo, preferred language, etc.

Necessary attributes related to security are userCertificate and userSMIMECertificate. They define the X.509 student certificate or the certificate for Multipurpose Internet Mail Extensions (MIME) e-mail applications.

Other standards related to e-learning are mainly dealing with content and do not treat security issues. For example [15]:

- The AICC focuses on practical aspects, providing recommendations for platforms for e-learning, peripherals, audio devices, and similar implementation details.
- The ARIADNE deals mainly with metadata to share and reuse materials.
- The ADL-SCORM is oriented towards methods of aggregation, description, and sequencing of learning objects.

IV. CONCLUSION

Despite the existence of different standards and their detailed specification, the issue of their implementation remains open. In many modern e-learning environments, support for standards was not included at the start, either because they were not developed at that time, either because of the lack of attachment of crucial importance to standards (assessing that implementation is unprofitable). Therefore, for systems already in operation, it is not easy to add support for standards, because in some segments it would require complete reengineering.

Specific general international standards are dealing with information security and which can be used to design a security information system or an appropriate module in any environment. These options relate to the design methodology, but also specific practices (according to ISO 27002). PDCA is recognized as a methodology suitable for designing modules.

There are a large number of non-compliant e-learning standards, and the security of information considers a number

of these standards. In practice, these elements are not implemented on a broader scale. The reasons for this can be sought in the complexity of the proposed solutions. On the other hand, the standards are relatively flexible so that they can be expanded and adapted to a particular need.

REFERENCES

[1] "Standard SRPS ISO/IEC 27001" Institute for Standardization of Serbia, Belgrade, 2015.

[2] "ISO 27000," 2013. [Online]. Available: http://www.27000.org. [Accessed: 02 Apr-2019].

[3] D. Zlatkovic, N. Denic, M.Petrovic and M. Ilic, „An Qualitative Analysis of the Effects of E-learning on Students Intristic Motivation", ITOP'19, Conference Proceedings, pp.185-193, Cacak, Serbia, 2019.

[4] V. Nikolic, D. Petkovic, N. Denic, M. Milovancevic, "Appraisal and review of e-learning and ICT systems in teaching process", *Physica A: Statistical Mechanics and its Applications*, Vol. 513, pp.456-464, 2019. https://doi.org/10.1016/j.physa.2018.09.003.

[5] M. Milosevic, "Development and Implementation of the Security Moduls and its Impact on the Quality of E-learning", Cacak, FTN, 2016.

[6] L. E. Anido-Rifon, M. J. Fernandez-Iglesias, M. Caeiro-Rodríguez, J. M. Santos-Gago, M. Llamas-Nistal, L. Alvarez Sabucedo, and R. Míguez Perez, "Standardization in computer-based education," *Computer Standards & Interfaces*, vol. 36, no. 3, pp. 604–625, 2014.

[7] B. Jerman-Blazic and T. Klobucar, "Privacy provision in e-learning standardized systems: status and improvements," *Computer Standards & Interfaces*, vol. 27, no. 6, pp. 561–578, 2005.

[8] F. Farance, and J. Tonkel, "IEEE P1484.1/D9, 2001-1130 Draft Standard for Learning Technology - Learning Technology Systems Architecture (LTSA)," pp. 1–120, 2001.

[9] N. Denic, and D. Zlatkovic, "A Study of the Potentials of the Distance Learning System", IConSE'17, Conference Proceedings, vol. 8, pp. 30-39, 2017.

[10] K. El-Khatib, L. Korba, Y. Xu, and G. Yee, "Privacy and Security in E-learning", *International Journal of Distance Education Technologies*. vol.1, no. 4, pp. 1-19, 2003.

[11] E. R. Weippl, *Security in e-learning*, New York, Springer, 2005.

[12] C. Smythe, *IMS Learner Information Package Summary of Changes*, IMS Global Learning Consortium, 2005.

[13] C. Smythe, F. Tansey, and R. Robson, *IMS learner information package information model specification*, IMS Global Learning Consortium, 2001.

[14] "eduPerson Object Class Specification (201602)", Internet2 Middleware Architecture Committee for Education, Directory Working Group, 2016.

[15] T. K. Shih and J. C. Hung, *Future directions in distance learning and communications technologies*. Hershey, PA: Idea Group Pub, 2007.

With Artificial Intelligence towards Intelligent Logistics and Supply Chains: the state of the

Evelin Krmac[1]

Abstract – **This paper focuses on the review of the importance of Artificial Intelligence (AI) solutions adoptions from scientific and professional points of view, as well as of the key challenges associated with the adoption of AI solutions in logistics and supply chain industry. Namely, AI is not only one of the biggest and most promising commercial opportunities for todays business, but it is becoming also an important field of study for academic researchers. It seems that at the moment, the interest and awareness of the relevance of AI for logistics and supply chains are higher among professionals than among scholars.**

Keywords – **Artificial Intelligence, AI, Logistics, Supply Chain, digitalization.**

I. Introduction

As in any other industry, logistics and supply chains strive to perform their activities and business processes intelligently as much as possible. And intelligently means digital. Nowadays we are increasingly confronted with concepts such as Intelligent logistics or Supply Chain, Industry 4.0 (or even 5.0), Digital Logistics or Supply Chain, Smart Logistics or Supply Chain and similar, both in professional and scientific environments. To achieve such a level of intelligence, the major and the most important role plays the technology.

Technology has always been and will always be very important for the optimization of the operational performance and efficiency of each company. It offers many key opportunities, but it offers also some key challenges. Among the latter are high costs and requirements for the technology implementation, ethical and legal concerns, the resistance of workforces, safety issues, integration issues, and other. Therefore, it is extremely important to approach the selection and implementation of technology thoughtfully and in a mature way.

Logistics and supply chains more than ever strive to be smarter, cost-effective, more flexible, accurate, efficient, precise, faster, sustainable, transparent, customer-centric, and to be available 'as a Service'. With current advancements in technology, this is more accessible than ever before.

II. Technologies in logistics and SC

[1]Evelin Krmac is with the Faculty of Maritime Studies and Transport of University of Ljubljana, Pot pomorščakov 4, Portorož 6320, Slovenia, E-mail: evelin.krmac@fpp.uni-lj.si.

To help logistics and supply chain players select or invest in the 'right', 'promising enough' technology, various predictions and classifications of promising technologies were prepared. In Table 1 three of them are presented. While Gartner [1] made a list of Top 10 strategically important technologies independent of the type of the industry, McKinsey [2] and DHL [3] offer classifications of technology trends dedicated for supply chains and logistics, respectively.

TABLE I
THE MOST IMPORTANT TECHNOLOGY TRENDS IN LOGISTICS.

*Gartner**	*McKinsey***	*DHL****
Autonomous Things	Big Data and Advanced Analytics	Cloud logistics (H, <5)
Augmented Analytics	Automation	Big Data Analytics (H, <5)
AI-driven Development	Artificial Intelligence	Internet of Things (H, <5)
Digital Twins	Autonomous and smart vehicles	Robotics & Automation (H, <5)
Empowered Edge*1	Supply Chain Cloud	Artificial Intelligence (H, >5)
Immersive Technologies*2	3D printing	3D printing (H, >5)
Blockchain		Self-driving Vehicles (H, >5)
Smart Spaces (i.e. advanced digital workplace or connected factory)		Augmented Reality (M, <5)
Digital Ethics and Privacy		Low-cost Sensor Solutions (M, <5)
Quantum Computing		Blockchain (M, >5)
		Next-generation Wireless (M, >5)
		Unmanned Aerial Vehicles (M, >5)
		Virtual Reality & Digital Twins (L, >5)
		Bionic Enhancement (L, >5)

*: Top 10 Strategic Technology Trends for 2019 [1]
*1: Cloud, devices (AI chips, greater compute capabilities, more storage), 5G [1]
*2: Conversational Platforms, Augmented Reality, Mixed Reality, and Virtual Reality [1]
**: the classification was made according to their Digital Supply Chain Compass [2]

***: the classification was made according to their Logistics Trend Radar [3] – first the impact (H-high, M-medium, L-low) was taken into account, then the relevance in years (< 5, >5)

Inspecting these technologies it is clear that AI solutions are embedded in, or are enabling many of them. On the other hand, the rise of the power of AI will be possible just because of advances of other technologies, like radio frequency identification (RFID), big data and advanced analytics, sensor solutions, new generation communication networks, and other as well.

A. Artificial Intelligence

Artificial Intelligence (AI) is one of the 'hottest' technologies (already) today, and according to Costello [4], "AI adoption in organizations has tripled in the past year and is a top priority for CIOs". Following DHL [5], AI is a "trend having much greater impact on logistics than expected in the past". The fact is that many tools, products, and applications used for logistics and supply chain support today already rely on AI.

The goal of AI is to create such a technology that gives the 'human' intelligence to the machines, - i.e., computers, computer-controlled robots, or software -, so that they can imitate humans and behave in an intelligent manner. Because the machines demonstrate intelligence, they are sometimes called also the machine intelligence.

AI solutions most often used or implemented for logistics and supply chain are machine learning approaches (through machine translation, speech recognition, image classification, and information retrieval), robotics, computer vision, natural language processing, and expert systems as the synonym of decision support system.

One of the greatest benefits of the AI solutions is the ability to learn and extract insights and knowledge from unstructured data (unstructured text, documents of various types, videos, online browsing data, conversations and posts on social media, emails, letters, photos, and other sources).

B. Key applications of AI in logistics and supply chain

As the supply chains and logistics itself are complex networks of product or service, information and financial flows between suppliers and customers, - i.e., physical and digital networks, which must function optimally and harmoniously, AI solutions, which are mostly network-based too, are very suitable for the problem solving in logistics and supply chain management (LSCM).

According to [2], [3], and [5], the key applications of AI in LSCM are:

Intelligent decision making support
Knowledge-based decision support systems are the result of AI used for decision-making. Such systems behave like an expert, consultant that is capable to gather and analyze data, identify problems from these data, and finally find and evaluate the solutions. Decision-making with the support of knowledge-based decision support systems is much more efficient and fast.

Back-office automation or automation of knowledge work
A large amount of data-repetitive tasks could be faster, cheaper, and more accurately executed using the cognitive automation, - i.e., a combination of AI and software robots (bots) that are integrated into existing IT applications. Bots are programmed for the execution of tasks according to predefined rules and (only) structured inputs. They can provide AI tools with a waste amount of business data while using AI technology like natural language processing, the extracted data from unstructured document or source of information (invoices or contract) can be classified and given to bots as input data.

Automated planning and scheduling
Ad hoc and real-time planning (demand forecasting), that dynamically changes over time as a result of changed requirements or constraints, is giving to the SC much more flexibility and creates a new business models such as Supply Chain-as-a-Service. It is possible when big data and advanced analytics together with AI techniques (Bayesian networks and machine learning) are used.

Predictive logistics
Machine learning-based tools can be used for predictions of delays, transit times, interest for a specific product/service (which and when the interest will be maximal), prospects for the global trade, (supplier) risks.

AI-powered customer experience with logistics provider - Customer personalization
Voice agent (a speaker or a software tool able to communicate via Facebook Messenger or SMS) offers a voice-based service in tracking ordered parcels and providing with the information about delivery times, locations, and other information.

Anticipatory logistics uses AI for delivering goods to customers before their effective order and maybe before they realize to need them. Anticipations are designed based on predictions made from browsing behavior, purchase history and other data.

Automation of customer interactions by voice with chatbots, playing the role of virtual assistants, allows complex dialogs with customers and therefore real-time customer assistance.

Collaborative human-machine environments (i.e. use of collaborative robots)
Collaborative robots are assisting workers while carrying out repetitive tasks, working with workers, who perform more intelligent, technically demanding tasks (so-called 'digital work'). Virtual Reality tools support workers in training and remote collaboration.

"Uberization" of transport and new transport concepts
The cooperative intelligent transport systems, intelligent route optimization offering dynamic routing facilities, last-mile delivery using autonomous unmanned aerial vehicles or autonomous ground vehicles, use of autonomous trucks/fleets, truck platooning or caravanning of groups of semi-trucks, etc. are some of the most important results of AI implementation in transport.

Logistics network orchestration

Harmonization of physical and digital networks on which logistics companies depend, using AI technologies.

Smart logistics assets or "Seeing, speaking and thinking logistics assets" [5]

Smart logistics is created by use of robotics (sorting robots, collaborative robots), computer vision systems (visual inspection, inventory management and execution, vision-based sorting), conversational interfaces (voice-based picking, conversation of logistics operators with IT system – i.e., system can interpret the meaning of speech/voice information and after "understanding" which product was mentioned, connects the data in ERP, WMS, or TMS about this product and allow the input, store or retrieval of data), autonomous vehicles (self-learning and self-navigating AGVs – Autonomous Guided Vehicles) as added value to the human workforce.

III. THE SCIENTIFIC LITERATURE REVIEW OF AI APPLICATIONS IN LOGISTICS AND SC

A. The methodology of the literature review

The review of open-access scientific original or review papers published until now was conducted in May 2019 by browsing the two of the most important and largest scientific databases, ScienceDirect and Scopus. As the purpose of the review was to found out how frequently the researchers have studied the application of Artificial Intelligence tools and approaches in the field of logistics and/or supply chain and which AI tools or approaches were studied, the search of databases was performed using keywords such as Logistics AND ("Artificial Intelligence" OR "Intelligence" OR "AI") and "Supply Chain" AND ("Artificial Intelligence" OR "intelligence" OR "AI"). The search was focused on titles, abstracts, and keywords for English-written full-text free-available scientific journal papers resulting in 95 papers found. After the initial reading of titles, abstracts, and keywords, 23 papers were excluded because the techniques used were not from the AI field, and 5 because they were duplicates. Remaining 67 papers were considered relevant after full-text reading and they entered the classification process. The overall search process is shown in Fig. 1.

B. Results of the literature review and classification

The data about the authors, title, aspect(s) studied, AI approaches used, the scope and the results of the article, as well as the publication year and journal title, were collected during the full-text reading.
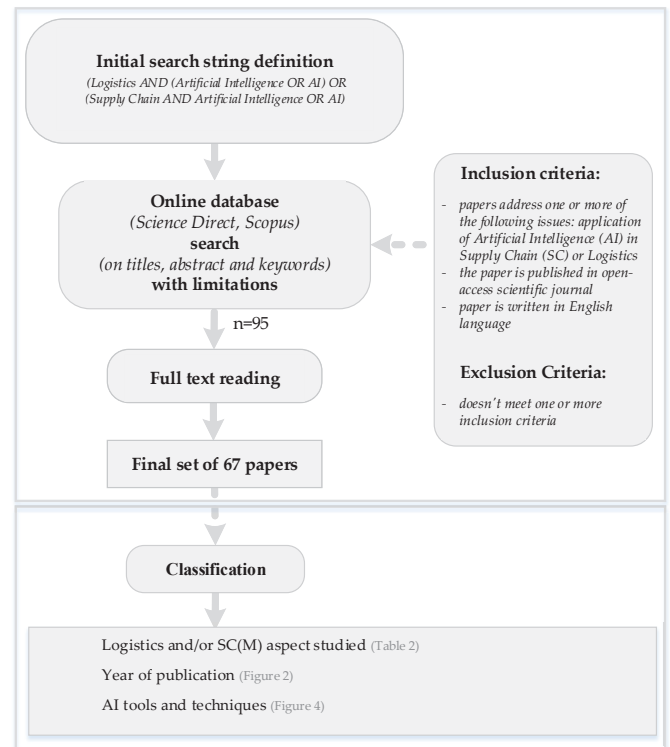

Fig. 1. The overall search process

As expected, the most often AI approaches were used for the optimization purposes (11), planning and scheduling problem solving (11), and forecasting and predicting (8). The review of other aspects of logistics and supply chain, as well as some short details about the reviewed articles are given in Table II. The empty cell in the column AI approach means that the literature review of was performed.

TABLE II.
LOGISTICS AND/OR SC(M) ASPECT STUDIED.

Aspect studied	Ref.	AI approach	Year	Summary
allocation	[6]	tabu search, path relinking	2012	a hybrid metaheuristic for dynamic berth allocation problem (port logistics) with the goal to minimize the total time the vessels stay at the port
automation and control	[7]	a rule-based system, machine learning	2018	approach for proactive management of raw milk quality with a high level of accuracy
	[8]	machine learning	2014	new detection and avoidance mechanisms of all counterfeit parts and forged documentation for electronic component SC

Aspect studied	Ref.	AI approach	Year	Summary
	[9]	distributed AI, multi-agent system, real-time decision making	2009	forecast of the use of radio frequency identification (RFID) technologies integrated into ICT framework in a cooperative intelligent logistics systems
	[10]	fuzzy logic, genetic algorithms	2016	a proposal for a combined application to control the procurement process in the enterprise (inventory control under uncertain conditions)
	[11]	deep learning	2018	3D object recognition and autonomization of logistics
big data	[12]		2018	review of Big Data Applications (BDA) in supply chain
	[13]	support vector machine and hierarchical clustering with multiscale bootstrap resampling	2018	proposes a big-data analytics-based approach that considers social media (Twitter) data for the identification of supply chain management issues in food industries
	[14]		2018	review of big data analytics and applications for logistics and SCM by examining novel methods, practices, and opportunities
cooperation	[15]	machine learning, deep learning	2017	the concept of a research center for issues of human interaction in logistics, intralogistics, and human and machine cooperation
decision making	[16]	bees algorithm, swarm intelligence	2010	a container loading support system (CLSS)
	[17]	sequential decision making, binary decision tree, heuristic algorithms	2011	formulation of a port of entry inspection sequencing task as a problem of finding an optimal binary decision tree for an appropriate Boolean decision function
	[18]		2014	review and analysis of past Simulation and Modeling efforts to support decision making in healthcare SCM
	[19]	decision-making	2018	design and implementation of decision support
	[20]	approximate dynamic programming, real-time dynamic programming	2009	a real-time method for solving multistage capacity decision problems in a manufacturing environment
forecast	[21]	Bayesian network	2019	airline network delay propagation model development
	[22]	machine learning, supervised models, predictive analytics, penalized regression	2018	forecast of taxi-timeout
	[23]	LSSVR model - vector regression	2017	selection for container throughput forecasting
	[24]	Croston's method, SBA and TSB, exponential smoothing, nearest neighbor	2016	forecast of sporadic demand (SCM) in the automotive industry
	[25]	machine learning	2014	a framework to solve the dynamic bike sharing repositioning problem
	[26]	machine learning, neural networks, support vector machine	2009	
	[27]		2017	review of insight into Big Data applications and Smart Farming
	[28]	logistic regression, support vector machine and back-propagation neural networks	2013	proposed novel ensemble learning approach for corporate financial distress forecasting in fashion and textile SC
optimization	[29]	multi-agent learning, deep reinforcement learning networks combined with link-state protocol and preliminary supervised learning, deep neural networks	2019	optimization of the distributed packet routing system
	[30]	least square-support vector machine, continuous general variable neighborhood search	2017	a novel model for solving non-linear regression problems (a case study of a supplier selection and evaluation problem in the cosmetics industry)
	[31]	intelligent agents, virtual reality	2017	a conceptual framework for the development of real-time intelligent observational platform supported by advanced intelligent agents
	[32]	self-optimization	2016	an introduction to the concept of self-optimizing production systems
	[33]	bees algorithm, swarm-based optimization	2013	optimization method for optimum configuration of a given supply chain problem which minimizes the total cost and the total lead-time
	[34]	decision support systems, heuristics	2013	metaheuristics for solving a complex optimization problem from logistics
	[35]	stochastic programming, heuristic methods, simulation-based methods	2012	modeling and optimization of SCM systems

Aspect studied	Ref.	AI approach	Year	Summary
optimization	[36]	deterministic and stochastic shrinking ball (DSB and SSB) approaches	2010	presents, analyzes, and compares three random search methods for solving stochastic optimization problems with uncountable feasible regions
	[37]	case-based reasoning, multi-agent, fuzzy logic, neural networks	2009	a multi-artificial intelligence system aimed to provide quality logistics solutions to achieve high levels of service performance in the logistics industry proposed
	[38]	machine learning, data mining, genetic algorithms, neural networks, knowledge discovery, classification, prediction, goal programming, rule induction	2006	evolutionary/genetic algorithm (GA)-based neural approach that incorporates asymmetric Type I and Type II error costs on financial and medical data
	[39]	proactive uncertainty management techniques, simulated annealing	2004	techniques for proactive uncertainty management determine the release dates of different jobs based (just-in-time job shop environment)
personalized product SC	[40]	multi-agent, negotiating mechanisms, models, and tactics	2018	model for personalized product SC
planning and scheduling	[41]	machine learning	2016	an approach for predictive inbound logistics
	[42]	review of classical and probabilistic planning algorithms	2013	planning
	[43]	agent-based system, decision support systems, genetic algorithm, tabu search, simulated annealing	2008	2 scenarios for solving Production-Distribution Planning Problem in a DSS framework
	[44]	machine learning, decision theory, and distributed AI	2016	a novel real-time path planning system for SC of road construction
	[45]	deep reinforcement learning, neural networks, a continuous-variable feedback control algorithm	2019	a learning-based logistics planning and scheduling (LLPS) algorithm that controls the admission of order requests and schedules the routes of multiple vehicles altogether
	[46]	multiobjective swarm intelligence algorithm, multiobjective gravitational search algorithm	2013	model for strategic planning and optimizing cost and CO_2 emissions in an environmentally friendly automotive supply chain
	[47]	stroke graphs	2013	an algorithm that solves the supply network configuration and operations scheduling problem in a mass automization company that faces alternative operations for one specific tool machine order in a multiplant context
	[48]	genetic algorithm, neighborhood search	2011	algorithm for the resource-constrained project scheduling problem
	[49]	constraint logic programming	2000	an algorithm which will allow the creation of partial schedules for reducing the search space and their combining to obtain the global schedule
	[50]	genetic algorithm (random keys, Bernoulli crossover, immigration type mutation)	1999	an algorithm that considers the scheduling problem to minimize total tardiness given multiple machines, ready times, sequence dependent setups, machine downtime, and scarce tools
	[51]	genetic algorithms	1998	an algorithm for the resource-constrained project scheduling problem
	[52]	ant colony optimization	2001	an application of ant colony optimization to address a production-sequencing problem when two objectives are present
production	[53]	multi-objective optimization, decision support systems	2017	an efficient multi-objective archived simulated annealing approach and a visualization technique for the multi-objective inter-terminal truck routing problem by specifically considering truck emissions
routing	[54]	genetic algorithm, beam search algorithm	2002	a method of routing yard-side equipment during loading operations in container terminals
	[55]	genetic algorithm	2014	the integrated design and planning of bioenergy supply chains problem
	[56]	neural networks, backpropagation algorithm	2011	a combination of advanced technologies to form an integrated system that helps achieve lean and agile logistics workflow
SCM improvement	[57]	expert systems - knowledge engineering	2017	system for measuring the necessities of professional contracts regarding insurance coverage and improve the supply chain management using IT

| Aspect studied	Ref.	AI approach	Year	Summary
	[58]	data-driven modeling, constraint logic, and mathematical programming	2017	a novel approach that would allow the flexible modeling and solving of food supply chain management (FSCM) problems
	[59]	case-based and rule-based reasoning, inference method, neural networks, Delphi method, data mining	2014	third-party (3PL) selection problem - review on criteria and methods
selection	[60]	neural network, locally linear neuro-fuzzy model, locally linear model tree learning algorithm. multi-layer perceptron, radial basis function neural network, least square-support vector machine	2011	an effective AI approach to improve the supplier selection (cosmetics industry)
	[61]		2017	a general overview of the present and future trends in sustainable SC
technology	[62]		2019	review of freight transport modeling techniques
transportation	[63]	deep belief, vector regression network method	2019	method for flight delay prediction
	[64]	machine learning, Bayesian classification	2019	approach for shipment size choice in strategic interregional freight transport models
	[65]	fuzzy logic, genetic algorithms	2018	a new model for in-plant transportation control with the AGV
	[66]	naive Bayesian, Bayesian network, logistic regression, multilayer perceptron, support vector machine, decision table, and C4.5 algorithms	2016	comparison of the relative performance of different algorithms for the detection of transportation modes and activity episodes
	[67]	multi-agent	2013	a structure model for military container transportation in campaigns logistics
	[68]	combination of operational research techniques with AI search methods	2013	a new hybrid approach for intermodal transportation problem solving
	[69]		2019	review of Artificial Intelligence (AI) adoption within warehouses
warehouse	[70]	tabu search	2014	algorithm for resource-constrained project scheduling problem optimization of truck-dock assignment in the cross-dock management system
	[71]	least squares temporal difference learning method	2012	algorithm for a vehicle-dispatching problem-solving in warehouse management
	[72]	object recognition methods, convolutional neural network	2016	a novel system to support order pickers in warehouses (using smartwatch and low-cost camera)
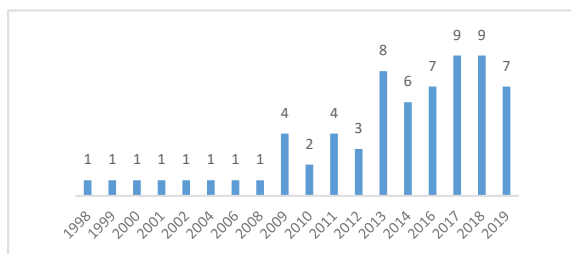
Fig. 2. Publication years
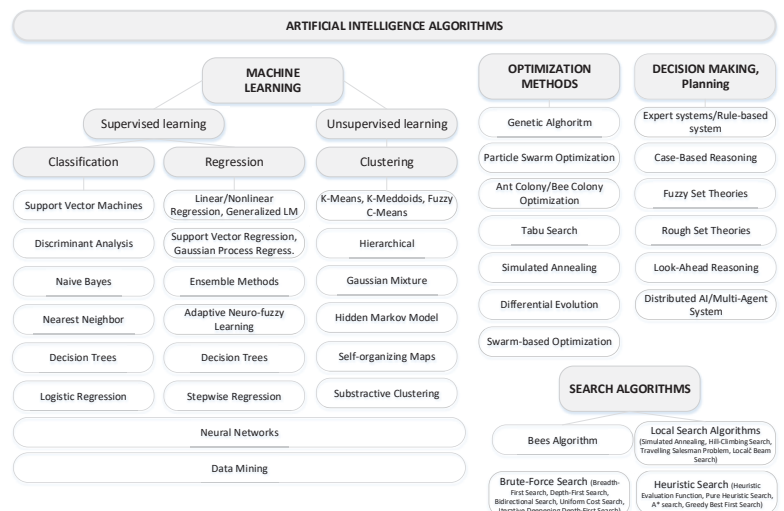
Fig. 3. Most 'popular' AI algorithms [73], [74] and [75]

Fig. 2 represents the number of publications by year. It could be seen that the rise of a number of papers is significant in the period of last ten years, while more than half of these papers were published in last 5 years, and that the number of publications is rising.

In theory and in scientific research as well there are really many AI algorithms, techniques, and approaches in use. Fig. 3 presents an attempt to summarize and classify some of them from the most often used within studies, - i.e. machine learning, optimization, decision making, planning, and searching.

Many of them were used also in the reviewed studies. The most frequently used are shown in Fig. 4.
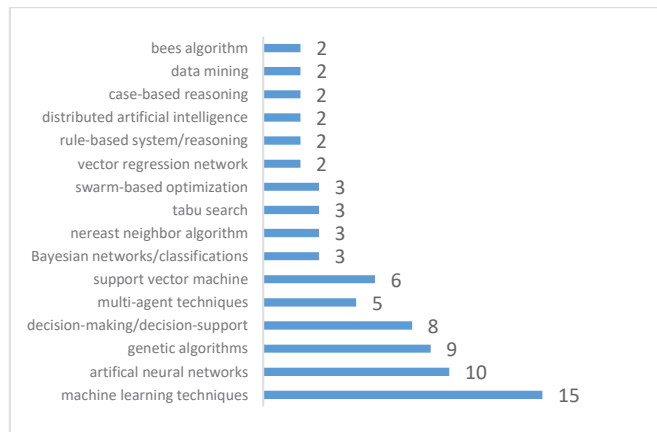
Fig. 4. AI algorithms used in the reviewed studies

Exploring the last three years' studies, - i.e. from 2017 to 2019, we can conclude, that the most often studied/used AI techniques were the machine learning techniques. Machine learning indeed is becoming one of the most important AI approaches. Combined with other AI approaches, like natural language processing, object recognition, robots, and others, it's a very promising tool for many applications in logistics and supply management industry.

IV. CONCLUSION AND DISCUSSION

There are many possibilities for the application of AI solutions across the supply chain and logistics operations. In all cases, there are some strong prerequisites that should (or better must) be fulfilled. Firstly, the 'basic' IT systems and technologies used for information (and partially) decision support must provide a large amount of data, that represent an adequate input for AI systems and tools. Secondly, the AI implementation should not be made only because it is popular and it promises 'a lot', but because there is a strong belief and awareness, that it will represent added value and substantial benefit for the business. Therefore, the business problem should be clearly defined and AI technologies appropriately selected. [76, 77] Finally, because of complexity and sometimes also the immaturity of AI technologies, for successful introduction all possible options should be carefully considered. Namely, an AI solution can be built in-house, can be bought, or outsourced. Sometimes the hybrid approach is most appropriate. [76]

Regarding the maturity of field of Logistics and Supply Chain Management (LSCM) on one side and the technology and AI maturity on the other, in next years it is therefore normal to expect higher level and stronger presence of AI in everyday logistics and supply chain operations, with the role of "accelerating the path towards a proactive, predictive, automated, and personalized future" [5] for logistics. Considering the increasing interest for AI solution from the business, the number of scientific papers offering various solutions and an insights in AI science is expecting to be much higher in the next years.

REFERENCES

[1] G. Inc., "Top 10 Strategic Technology Trends for 2019," ed, 2019.
[2] McKinsey&Company, "Supply Chain 4.0 - the next generation digital supply chain," ed, 2016.
[3] D. T. Research, "Logistics Trend Radar," ed, 2019.
[4] K. Costello, "Gartner Predicts the Future of AI Technologies," ed, 2019.
[5] D. T. Research, "Artificial Intelligence in Logistics," ed, 2018.
[6] E. Lalla-Ruiz, B. Melián-Batista, and J. Marcos Moreno-Vega, "Artificial intelligence hybrid heuristic based on tabu search for the dynamic berth allocation problem," Engineering Applications of Artificial Intelligence, vol. 25, no. 6, pp. 1132-1141, 2012, doi: https://doi.org/10.1016/j.engappai.2012.06.001.
[7] A. Zakeri, M. Saberi, O. K. Hussain, and E. Chang, "An early detection system for proactive management of raw milk quality: an australian case study," IEEE Access, vol. 6, pp. 64333-64349, 2018, doi: 10.1109/ACCESS.2018.2877970.
[8] U. Guin, K. Huang, D. Dimase, J. M. Carulli, M. Tehranipoor, and Y. Makris, "Counterfeit integrated circuits: A rising threat in the global semiconductor supply chain," Proceedings of the IEEE, vol. 102, no. 8, pp. 1207-1228, 2014, doi: 10.1109/JPROC.2014.2332291.
[9] J. C. Q. Dias, J. M. F. Calado, A. L. Osório, and L. F. Morgado, "RFID together with multi-agent systems to control global value chains," Annual Reviews in Control, vol. 33, no. 2, pp. 185-195, 2009, doi: https://doi.org/10.1016/j.arcontrol.2009.03.005.
[10] P. Więcek, "Intelligent Approach to Inventory Control in Logistics under Uncertainty Conditions," Transportation Research Procedia, vol. 18, pp. 164-171, 2016, doi: https://doi.org/10.1016/j.trpro.2016.12.023.
[11] M. Thiel, J. Hinckeldeyn, and J. Kreutzfeldt, "Deep learning for 3D object recognition in logistics," Logistics Journal, vol. 2018, no. 1, 2018, doi: 10.2195/lj_Proc_thiel_de_201811_01.
[12] D. Arunachalam, N. Kumar, and J. P. Kawalek, "Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice," Transportation Research Part E: Logistics and Transportation Review, vol. 114, pp. 416-436, 2018, doi: https://doi.org/10.1016/j.tre.2017.04.001.
[13] A. Singh, N. Shukla, and N. Mishra, "Social media data analytics to improve supply chain management in food industries," Transportation Research Part E: Logistics and Transportation Review, vol. 114, pp. 398-415, 2018, doi: https://doi.org/10.1016/j.tre.2017.05.008.
[14] K. Govindan, T. C. E. Cheng, N. Mishra, and N. Shukla, "Big data analytics and application for logistics and supply chain management," Transportation Research Part E: Logistics and Transportation Review, vol. 114, pp. 343-349, 2018, doi: https://doi.org/10.1016/j.tre.2018.03.011.

[15] F. Zeidler, H. Bayhan, A. K. R. Venkatapathy, and M. T. Hompel, "Reference field for research and development of novel hybrid forms of human machine interaction in logistics," Logistics Journal, vol. 2017, 2017, doi: 10.2195/lj_Proc_zeidler_de_201710_01.

[16] T. Dereli and G. S. Das, "Development of a decision support system for solving container loading problems," Transport, vol. 25, no. 2, pp. 138-147+Ib+IIb, 2010, doi: 10.3846/transport.2010.17.

[17] D. Madigan, S. Mittal, and F. Roberts, "Efficient sequential decision-making algorithms for container inspection operations," Naval Research Logistics, vol. 58, no. 7, pp. 637-654, 2011, doi: 10.1002/nav.20472.

[18] E. Abukhousa, J. Al-Jaroodi, S. Lazarova-Molnar, and N. Mohamed, "Simulation and modeling efforts to support decision making in healthcare supply chain management," The Scientific World Journal, vol. 2014, 2014, doi: 10.1155/2014/354246.

[19] M. Zhao and X. Liu, "Development of decision support tool for optimizing urban emergency rescue facility locations to improve humanitarian logistics management," Safety Science, vol. 102, pp. 110-117, 2018, doi: https://doi.org/10.1016/j.ssci.2017.10.007.

[20] N. E. Pratikakis, M. J. Realff, and J. H. Lee, "Strategic capacity decision-making in a stochastic manufacturing environment using real time approximate dynamic programming," Naval Research Logistics, vol. 57, no. 3, pp. 211-224, 2010, doi: 10.1002/nav.20384.

[21] C.-L. Wu and K. Law, "Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model," Transportation Research Part E: Logistics and Transportation Review, vol. 122, pp. 62-77, 2019, doi: https://doi.org/10.1016/j.tre.2018.11.004.

[22] T. Diana, "Can machines learn how to forecast taxi-out time? A comparison of predictive models applied to the case of Seattle/Tacoma International Airport," Transportation Research Part E: Logistics and Transportation Review, vol. 119, pp. 149-164, 2018, doi: https://doi.org/10.1016/j.tre.2018.10.003.

[23] G. Xie, N. Zhang, and S. Wang, "Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology," Transportation Research Part E: Logistics and Transportation Review, vol. 108, pp. 160-178, 2017, doi: https://doi.org/10.1016/j.tre.2017.08.015.

[24] K. I. Nikolopoulos, M. Z. Babai, and K. Bozos, "Forecasting supply chain sporadic demand with nearest neighbor approaches," International Journal of Production Economics, vol. 177, pp. 139-148, 2016, doi: https://doi.org/10.1016/j.ijpe.2016.04.013.

[25] R. Regue and W. Recker, "Proactive vehicle routing with inferred demand to solve the bikesharing rebalancing problem," Transportation Research Part E: Logistics and Transportation Review, vol. 72, pp. 192-209, 2014, doi: https://doi.org/10.1016/j.tre.2014.10.005.

[26] J. Shahrabi, S. S. Mousavi, and M. Heydar, "Supply chain demand forecasting: A comparison of machine learning techniques and traditional methods," Journal of Applied Sciences, vol. 9, no. 3, pp. 521-527, 2009, doi: 10.3923/jas.2009.521.527.

[27] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big Data in Smart Farming – A review," Agricultural Systems, vol. 153, pp. 69-80, 2017, doi: 10.1016/j.agsy.2017.01.023.

[28] G. Xie, Y. Zhao, M. Jiang, and N. Zhang, "A novel ensemble learning approach for corporate financial distress forecasting in fashion and textiles supply chains," Mathematical Problems in Engineering, vol. 2013, 2013, doi: 10.1155/2013/493931.

[29] D. Mukhutdinov, A. Filchenkov, A. Shalyto, and V. Vyatkin, "Multi-agent deep learning for simultaneous optimization for time and energy in distributed routing system," Future Generation Computer Systems, vol. 94, pp. 587-600, 2019, doi: https://doi.org/10.1016/j.future.2018.12.037.

[30] B. Vahdani, S. M. Mousavi, R. Tavakkoli-Moghaddam, and H. Hashemi, "A new enhanced support vector model based on general variable neighborhood search algorithm for supplier performance evaluation: A case study," International Journal of Computational Intelligence Systems, vol. 10, no. 1, pp. 293-311, 2017, doi: 10.2991/ijcis.2017.10.1.20.

[31] Z. Asgari and F. P. Rahimian, "Advanced Virtual Reality Applications and Intelligent Agents for Construction Process Optimisation and Defect Prevention," Procedia Engineering, vol. 196, pp. 1130-1137, 2017, doi: https://doi.org/10.1016/j.proeng.2017.08.070.

[32] E. Permin et al., "Self-optimizing Production Systems," Procedia CIRP, vol. 41, pp. 417-422, 2016, doi: https://doi.org/10.1016/j.procir.2015.12.114.

[33] E. Mastrocinque, B. Yuce, A. Lambiase, and M. S. Packianather, "A multi-objective optimization for supply chain network using the bees algorithm," International Journal of Engineering Business Management, vol. 5, no. 1, pp. 1-11, 2013, doi: 10.5772/56754.

[34] J. Schönberger and H. Kopfer, "A Procedural Framework for Hybrid Decentralized Decision Problem Solving," IFAC Proceedings Volumes, vol. 46, no. 9, pp. 724-729, 2013, doi: https://doi.org/10.3182/20130619-3-RU-3018.00517.

[35] S. Chen, Y. Zheng, C. Cattani, and W. Wang, "Modeling of biological intelligence for SCM system optimization," Computational and Mathematical Methods in Medicine, vol. 2012, 2012, doi: 10.1155/2012/769702.

[36] S. Andradóttir and A. A. Prudius, "Adaptive random search for continuous simulation optimization," Naval Research Logistics, vol. 57, no. 6, pp. 583-604, 2010, doi: 10.1002/nav.20422.

[37] Y. K. Tse, T. M. Chan, and R. H. Lie, "Solving complex logistics problems with multi-artificial intelligent system," International Journal of Engineering Business Management, vol. 1, no. 1, pp. 37-48, 2009, doi: 10.5772/6781.

[38] P. Pendharkar and S. Nanda, "A misclassification cost-minimizing evolutionary-neural classification approach," Naval Research Logistics, vol. 53, no. 5, pp. 432-447, 2006, doi: 10.1002/nav.20154.

[39] R. Bollapragada and N. M. Sadeh, "Proactive release procedures for just-in-time job shop environments, subject to machine failures," Naval Research Logistics, vol. 51, no. 7, pp. 1018-1044, 2004, doi: 10.1002/nav.20036.

[40] C. Yang, T. Xu, R. Yang, and Y. Li, "Multi-agent single-objective negotiation mechanism of personalized product supply chain based on personalized index," Advances in Mechanical Engineering, vol. 10, no. 10, 2018, doi: 10.1177/1687814018795785.

[41] D. Knoll, M. Prüglmeier, and G. Reinhart, "Predicting Future Inbound Logistics Processes Using Machine Learning," Procedia CIRP, vol. 52, pp. 145-150, 2016, doi: https://doi.org/10.1016/j.procir.2016.07.078.

[42] A. R. Sousa and J. J. P. Z. S. Tavares, "Toward Automated Planning Algorithms Applied to Production and Logistics," IFAC Proceedings Volumes, vol. 46, no. 24, pp. 165-170, 2013, doi: https://doi.org/10.3182/20130911-3-BR-3021.00081.

[43] A. Kazemi and M. H. Fazel Zarandi, "An agent-based framework for building decision support system in supply chain management," Journal of Applied Sciences, vol. 8, no. 7, pp. 1125-1137, 2008, doi: 10.3923/jas.2008.1125.1137.

[44] [44] R. Kuenzel, J. Teizer, M. Mueller, and A. Blickle, "SmartSite: Intelligent and autonomous environments,

machinery, and processes to realize smart road construction projects," Automation in Construction, vol. 71, pp. 21-33, 2016, doi: https://doi.org/10.1016/j.autcon.2016.03.012.

[45] Y. Kang, S. Lee, and B. D. Chung, "Learning-based logistics planning and scheduling for crowdsourced parcel delivery," Computers & Industrial Engineering, vol. 132, pp. 271-279, 2019, doi: https://doi.org/10.1016/j.cie.2019.04.044.

[46] A. Sadrnia, N. Ismail, N. Zulkifli, M. K. A. Ariffin, H. Nezamabadi-Pour, and H. Mirabi, "A multiobjective optimization model in automotive supply Chain networks," Mathematical Problems in Engineering, vol. 2013, 2013, doi: 10.1155/2013/823876.

[47] J. Maheut and J. P. Garcia-Sabater, "Algorithm for complete enumeration based on a stroke graph to solve the supply network configuration and operations scheduling problem," Journal of Industrial Engineering and Management, vol. 6, no. 3 SPL.ISS, pp. 779-795, 2013, doi: 10.3926/jiem.550.

[48] S. Proon and M. Jin, "A genetic algorithm with neighborhood search for the resource-constrained project scheduling problem," Naval Research Logistics, vol. 58, no. 2, pp. 73-82, 2011, doi: 10.1002/nav.20439.

[49] s. Buthod-Garçon, M. Dumoulin, and L. Trilling, "Experiments with Constraint Based Long Term Scheduling of Jobs," IFAC Proceedings Volumes, vol. 33, no. 17, pp. 941-943, 2000, doi: https://doi.org/10.1016/S1474-6670(17)39530-7.

[50] B. A. Norman and J. C. Bean, "A Genetic Algorithm Methodology for Complex Scheduling Problems," Naval Research Logistics, vol. 46, no. 2, pp. 199-211, 1999, doi: 10.1002/(SICI)1520-6750(199903)46:2<199::AID-NAV5>3.0.CO;2-L.

[51] S. Hartmann, "A competitive genetic algorithm for resource-constrained project scheduling," Naval Research Logistics, vol. 45, no. 7, pp. 733-750, 1998, doi: 10.1002/(SICI)1520-6750(199810)45:7<733::AID-NAV5>3.0.CO;2-C.

[52] P. R. McMullen, "An ant colony optimization approach to addressing a JIT sequencing problem with multiple objectives," Artificial Intelligence in Engineering, vol. 15, no. 3, pp. 309-317, 2001, doi: https://doi.org/10.1016/S0954-1810(01)00004-8.

[53] L. Heilig, E. Lalla-Ruiz, and S. Voß, "Multi-objective inter-terminal truck routing," Transportation Research Part E: Logistics and Transportation Review, vol. 106, pp. 178-202, 2017, doi: https://doi.org/10.1016/j.tre.2017.07.008.

[54] K. Y. Kim and K. H. Kim, "Heuristic algorithms for routing yard-side equipment for minimizing loading times in container terminals," Naval Research Logistics, vol. 50, no. 5, pp. 498-514, 2003, doi: 10.1002/nav.10076.

[55] K. K. Castillo-Villar, "Metaheuristic algorithms applied to bioenergy supply chain problems: Theory, review, challenges, and future," Energies, vol. 7, no. 11, pp. 7640-7672, 2014, doi: 10.3390/en7117640.

[56] C. K. M. Lee, W. Ho, G. T. S. Ho, and H. C. W. Lau, "Design and development of logistics workflow systems for demand management with RFID," Expert Systems with Applications, vol. 38, no. 5, pp. 5428-5437, 2011, doi: https://doi.org/10.1016/j.eswa.2010.10.012.

[57] S. Shokouhyar, S. Seifhashemi, H. Siadat, and M. M. Ahmadi, "Implementing a fuzzy expert system for ensuring information technology supply chain," Expert Systems, vol. 36, no. 1, 2019, doi: 10.1111/exsy.12339.

[58] P. Sitek, J. Wikarek, and P. Nielsen, "A constraint-driven approach to food supply chain management," Industrial Management and Data Systems, vol. 117, no. 9, pp. 2115-2138, 2017, doi: 10.1108/IMDS-10-2016-0465.

[59] A. Aguezzoul, "Third-party logistics selection problem: A literature review on criteria and methods," Omega, vol. 49, pp. 69-78, 2014, doi: https://doi.org/10.1016/j.omega.2014.05.009.

[60] B. Vahdani, S. H. Iranmanesh, S. M. Mousavi, and M. Abdollahzade, "A locally linear neuro-fuzzy model for supplier selection in cosmetics industry," Applied Mathematical Modelling, vol. 36, no. 10, pp. 4714-4727, 2012, doi: https://doi.org/10.1016/j.apm.2011.12.006.

[61] M. Merlino and I. Sproġe, "The Augmented Supply Chain," Procedia Engineering, vol. 178, pp. 308-318, 2017, doi: https://doi.org/10.1016/j.proeng.2017.01.053.

[62] L. Tavasszy and H. Friedrich, "Supply chain elements in freight transport modelling," Transportation Research Part E: Logistics and Transportation Review, vol. 121, pp. 1-3, 2019, doi: https://doi.org/10.1016/j.tre.2018.11.009.

[63] B. Yu, Z. Guo, S. Asian, H. Wang, and G. Chen, "Flight delay prediction for commercial air transport: A deep learning approach," Transportation Research Part E: Logistics and Transportation Review, vol. 125, pp. 203-221, 2019, doi: https://doi.org/10.1016/j.tre.2019.03.013.

[64] R. Piendl, T. Matteis, and G. Liedtke, "A machine learning approach for the operationalization of latent classes in a discrete shipment size choice model," Transportation Research Part E: Logistics and Transportation Review, vol. 121, pp. 149-161, 2019, doi: https://doi.org/10.1016/j.tre.2018.03.005.

[65] G. Kłosowski, A. Gola, and T. Amila, "Computational Intelligence in Control of AGV Multimodal Systems," IFAC-PapersOnLine, vol. 51, no. 11, pp. 1421-1427, 2018, doi: https://doi.org/10.1016/j.ifacol.2018.08.315.

[66] T. Feng and H. J. P. Timmermans, "Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data," Transportation Planning and Technology, vol. 39, no. 2, pp. 180-194, 2016, doi: 10.1080/03081060.2015.1127540.

[67] L. Qinzhen, L. Rongsheng, L. Yanxia, and L. Shillips, "Study on campaigns logistics military container transportation management information system based on multi-agent," Journal of Applied Sciences, vol. 13, no. 21, pp. 4495-4500, 2013, doi: 10.3923/jas.2013.4495.4500.

[68] J. García et al., "Combining linear programming and automated planning to solve intermodal transportation problems," European Journal of Operational Research, vol. 227, no. 1, pp. 216-226, 2013, doi: https://doi.org/10.1016/j.ejor.2012.12.018.

[69] K. Mahroof, "A human-centric perspective exploring the readiness towards smart warehousing: The case of a large retail distribution warehouse," International Journal of Information Management, vol. 45, pp. 176-190, 2019, doi: https://doi.org/10.1016/j.ijinfomgt.2018.11.008.

[70] Z. Miao, S. Cai, and D. Xu, "Applying an adaptive tabu search algorithm to optimize truck-dock assignment in the crossdock management system," Expert Systems with Applications, vol. 41, no. 1, pp. 16-22, 2014, doi: https://doi.org/10.1016/j.eswa.2013.07.007.

[71] R. M. Estanjini, K. Li, and I. C. Paschalidis, "A least squares temporal difference actor-critic algorithm with applications to warehouse management," Naval Research Logistics, vol. 59, no. 3-4, pp. 197-211, 2012, doi: 10.1002/nav.21481.

[72] R. Grzeszick, S. Feldhorst, C. Mosblech, M. ten Hompel, and G. A. Fink, "Camera-assisted pick-by-feel," Logistics Journal, vol. 2016, 2016, doi: 10.2195/lj_Proc_grzeszick_en_201610_01.

[73] MathWorks, "What Is machine Learning? 3 things you need to know," ed, 2019.

[74] A. Badar, U. B.S., and A. S. Junghare, "Study of Artificial Intelligence Optimization Techniques applied to Active Power Loss Minimization," ed. International Conference on Advances in Engineering & Technology – 2014 (ICAET-2014), IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), 2014, pp. 39-45.

[75] J.-C. Pomerol, "Artificial intelligence and human decision making," vol. 99, ed: European Journal of Operational Research, 1997, pp. 3-25.

[76] G. Inc., "Five Questions for a Successful AI Project," ed, 2019.

[77] R. Burkhardt, N. Hohn, and C. Wigley, "Leading your organization to responsible AI," ed: McKinsey Analytics, 2019.B. Milovanovic, Z. Stankovic, S. Ivkovic and V. Stankovic, "Loaded Cylindrical Metallic Cavities Modeling using Neural Networks", TELSIKS'99, Conference Proceedings, pp.214-217, Nis, Yugoslavia, 1999.

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

Interconnecting Wireless Sensor Networks Into IoT

Mirko Kosanovic[1] and Milos Kosanovic[2]

Abstract – In recent years, we are witnessing an increasing presence of a large number of different sensors that supply us with different data. The Wireless Sensor Network (WSN) has become a basic technology that has enable us to collect these data and that almost all complex phenomena in the physical world convert a simple set of units and zeros. Further evolution of this technology into Internet of Things (IoT) expands the possibilities of much application. This paper presents some of the basic approaches and technologies that are used today. At the end of the work, a typical solution for IoT application is proposed.

Keywords – **wireless sensor node, Internet of Things, cloud computing, mobile application**

I. INTRODUCTION

WSN increasingly have an impact in our everyday lives. They reveal a wide range of applications across different domains, including health, assistance and improved living conditions, industrial and production monitoring, traffic and transport controls and many other areas. The dream of intelligent devices that independently perform many complex tasks became a java because it was realized by the implementation of the IoT concept. The IoT makes it possible to connect WSN into entire Internet infrastructure through various communication and information technologies. Already in 2008, the number of Internet connected objects on our planet has exceeded the total number of people living on it. At the moment, there are over 10 billion interconnected facilities, and according to Cisco Systems, Dave Evans futurists, it is expected that at the end of 2020 this figure will reach about 50 billion interconnected computer objects [1]. His vision is that in the future, the world will be full of networked, smart devices that will have unique identification (IP number or radio frequency identification - RF number). All of these devices will be equipped with a large number of different sensors and will be able to completely independently collect a large number of different data that will be exchanged or forwarded to the parent device (sink), which will enable them to be available on the Internet by wireless connection. It has emerged as the ultimate product of the three core technologies that in recent years have undergone the greatest progress and development: wireless communications, micro electro mechanical systems (MESN) and Internet technologies. One of the main advantages of IoT is that it allows a large number of people to create very powerful and cheap applications with very little knowledge and experience. All this is made possible by the development of a large

number of micro-platforms, such as Arduino, Propeller and Microchip PIC families, and their integration through wireless communication with the Internet. In other words, this technique requires engineers to equally share knowledge of hardware and software. This is necessary in order to apply this knowledge in a coherent and integrated way in order to obtain an applicable IoT application [2, 3]. On the other hand, there has been a major development of Internet technologies that have contributed to realizing a large amount of data from a large number of computers, and this is almost always presented to clients. First of all it refers to new generation of distributed computing technologies: Cluster computing, Grid computing, Big Data and Cloud computing. They were now a precondition for developing and using new applications, WEB service oriented software. These technologies also have the ability to use messaging mechanisms such as Email, SMS or to exchange messages via social networks and blogs [4]. All of these features were ideal for their benefits to be utilized to connect to a single distributed source of a large number of data, which are WSN.

The rest of this paper is organized as follows: Section 2 provides the actual open issues of development of IoT. The advantages of 6LoWPAN technology are explained in Section 3. Section 4 discusses about the CoAP application protocols and its applicability into IoT applications. The types of web services that are applicable in IoT applications are discussed in Section 5. In Section 6 we present one proposal for an construction of standard IoT application. At the end, Section 7 concludes this paper by looking at the future research and recommendations which are required to make the IoT application more effective.

II. RESEARCH QUESTIONS

Following the IoT vision, WSN has made significant progress in recent years and has become closer to the structure and organization of standard protocols from the TCP/IP suite. Traditional WSN has emerged from simple, isolated systems for the monitoring natural phenomena into interoperable, powerful systems that are connected to the Internet. Cloud computing and IoT, which are two distinct technologies, are expected to become an integral part of the Future Internet considering the rate at which they are being adopted and used. This development has taken place in several different fields, of which they are particularly distinguished:

1. **Protocol customization**: new customizable protocols have been developed that could be implemented on modest resources available to sensors that were compatible with standard protocols from a TCP/IP suite. 6LoWPAN and CoAP are typical protocols that have enabled WSN to become more powerful in terms of interoperability and scalability.

[1]Mirko _R. Kosanovic is with the College of Applied Technical Science Nis, Aleksandra Medvedeva 20, 18000 Nis, Serbia, E-mail: milos.kosanovic@vtsnis.edu.rs.

[2]Milos M. Kosanovic is with the College of Applied Technical Science Nis, Aleksandra Medvedeva 20, 18000 Nis, Serbia, E-mail: mirko.kosanovic@vtsnis.edu.rs

2. **Powerful WSN application** – today sensor nodes(SN) have many multimedia sensors such as camera, microphone, radio frequency identification reader (RFID reader), radio frequency identification TAG (RFID -TAG)). Such progress has increased the interoperability and strength of WSNs and opens one new array: Big Data. It's a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not important the amount of data, than what someone do with these data. Big data can be analyzed for insights that lead to prediction of events, better decisions or strategic moves.

3. **Advancement in hardware technologies** – today SNs have significantly more powerful resources and computing capabilities: from 8b to 32b microcontroller, from 8MHz to 80MHz, from 10KB to 128KB RAM, from 48KB to 512KB flash memory. On other side we have cheaper hardware (Arduino) and an increasing number of different Wi-Fi modules(ESP2866) that connected SN with Internet.

4. **Development of new operating systems** - Another requirement to consider is the ability to remotely change the services (program code) provided by the SN during its lifecycle. This possibility enabled the use of many advanced techniques like Web services or possibility of use Web Socket protocols, which has enabled many real-time applications.

III. 6LoWPAN TECHNOLOGY

6LoWPAN is an open standard defined in RFC 6282 by the Internet Engineering Task Force (IETF). Initially, 6LoWPAN was designed to support the IEEE 802.15.4 standard for low-power wireless networks in the 2.4-GHz area. Now it become a standard that has been adapted and can be used in a variety of other networking technologies, such as the Sub-1GHz low - power RF, Bluetooth Smart, power line control (PLC) and low-power Wi-Fi. It is based on standard IEEE 802.15.4 (LoWPAN) that is applied to devices that have very limited resources such as wireless SNs. One of the basic prerequisites for each SN is the minimum power consumption in order to achieve the maximum life expectancy [5].

A lot of different technologies have been developed for wireless communication between SNs that allow access to the Internet, but it is quite clear that technologies based on the TCP/IP protocol suite have an advantage [6]. As most protocols from the TCP/IP suite represent very demanding protocols, with only headers being large enough (IPv4-24 bytes, IPv6-40 bytes, UDP-8 bytes, TCP-24 bytes), it is clear that the full implementation of these protocols in the WSN was a complete failure, from the point of energy efficiency and efficiency. For only 2-3 payload data, which is usually the size of the data transmitted through WSN, we would have to transmit 30 bytes, and therefore we spend a large amount of electricity, which indirectly means that life of these SNs was reduced to only a few days [7]. Technology 6LoWPAN put this requirement as the primary goal that must be fulfilled.

6LoWPAN network connects to other IP networks through one or more edge routers whose task is to transfer IP packets between different networks (see Figure 1). Connecting to other IP networks can be provided through various links-links such as Ethernet, Wi-Fi or 3G/4G. As the 6LoWPAN standard defines IPv6 protocol operations over the IEEE 802.15.4 standard, routers must also support an IPv6 transition mechanism that allows the connection of the 6LoWPAN network to IPv4 networks, as defined in the RFC 6146. This frees SN in the 6LoWPAN network to have this mechanism implemented, while at the same time allowing them to interact with IPv4 networks without interruption. Generally speaking, the basic goal of each 6LoWPAN network is to enable the use of standard Web services, such as REST, XML, JSON, and other devices with very limited resources, such as SNs, in environments that have a very high percentage badly received messages, LLNs (Low-Power and Lossy Networks) networks.
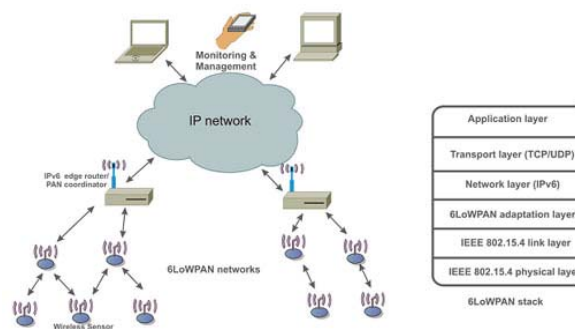

Fig 1. Typically connect devices to a 6LoWPAN network

IV. CoAP APPLICATION PROTOCOL

The integration of the REST architecture within WSN is not at all a simple task from the simple fact that this is a very limited resource available to the SN within that network. Typical SNs are powered by a battery, a strictly limited source of power, and have several kB memory and CPU that have limited computing capabilities. Consequently, any direct application of the original protocols from the TCP/IP suite is completely inapplicable because in this way the lifetime of all SNs in the network would be very short [8]. That is a main reason to develop a new special protocol, by the IETF working group, called the CoAP (Constrained Application Protocol), with the goal of being the basic Web transport protocol that would be applied to WSN and which would replace the HTTP protocol. CoAP seeks to apply the same data transfer system as the HTTP, but with significantly less resource requirements. It supports one part of the HTTP function, but also extends this set with its own functions to simplify communication between two SN, i.e. enabled M2M (Machine to Machine) communication. In this way, the services offered by Web services are gaining importance because now each SN can use them and participate in their expansion [9].

The core intention of the CoAP protocol is to provide a generic Web protocol that the SNs can communicate with. It is very similar to HTTP protocol (Fig.2), but its goal is not simply to compress HTTP packets, but to implement a reduced set of system messages that will enable M2M communication. To this end, CoAP uses the following four

messages: CON (Confirmable), NON (Non-Confirmable), ACK (Acknowledgment) and RST (Reset). The operating mechanism of this protocol is similar to the client-server communication model because the client always requires some service from the server. After the requested request, the server responds including the unique message ID received with the request (see Figure 3). By looking at Figures 2 and 3, it can be seen how much the CoAP protocol is used up because it only uses two packets in normal communication, as opposed to an HTTP protocol that requires 7 packets. A particular advantage of the CoAP protocol is that minimum resources are needed that each SN needs to be able to be implemented: 10 kB RAM and 100 kB of memory space for program code (RFC 7228) [10].
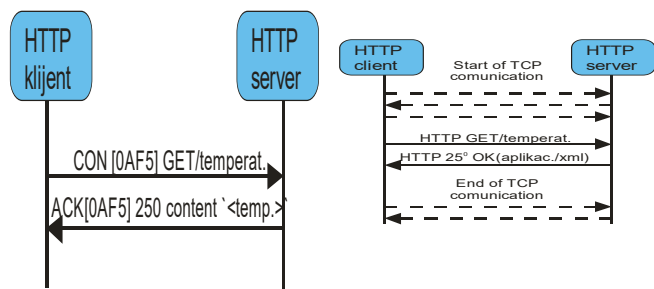
Fig.2. HTTP and CoAP comunication

V. WEB SERVICES

Web services represent a new look towards writing applications that are applied on the Internet. Their goal is to enable the connection of various information systems installed in distributed environments. They represent one type of distributed application that consists of several application functions that can be programmed from anywhere that has the ability to connect to the Internet. In doing so, the consumer and the provider of these applications use the messaging system to exchange their requests and answers, most often through the HTTP, whilst that communication does not depend of the resources which a provider and a consumer of these services have. In other words, the Web service is not, at all, tied to a specific hardware and software platform used by both parties in communication. It's just enough to support some of the protocols for sharing information such as HTTP or SMTP protocols. A typical architecture in which Web services are used (see Fig. 4) is a client-server architecture that allows one network component to play the role of a service provider, service user or service broker. With their simplicity and great capabilities, they have today become an integral part of almost all modern information systems.

The very rapid development of the Internet as well as related technologies led to the development of many technologies that were applied to Web services that were changing rapidly. From today's point of view, two Web development platforms have emerged as the leading: Microsoft.NET and Sun J2EE. Regardless of the chosen platform, all Web services can easily call each other. In general, we can split all Web services into two large groups: Big Web Services and RESTful Web Services (see Figure 4).

"Big" web services are based on the SOAP protocol for easy access to the sites and often contain a WSDL language for describing Web services. WSDL enables us to call each Web service when it is called, and which contains this function, it can describe itself, the operations it supports how we can use them. The details that WDSL provides can be messages, operations, links, and Web service locations. A SOAP message is usually an XML document defined by an XML schema. In the SOAP Web service, SOAP messages are not the subject of a programmer's interest in creating a Web service or writing a client code for using the Web service, but they only enable the communication of a client application that calls the Web service and the Web service itself. This type of Web service is suitable when we need asynchronous processing, reliability and pre-formulated State-full operations. If the application needs further information, SOAP provides an additional specification in the Web services structure to support the query (security, transactions, coordination, etc.) [11].
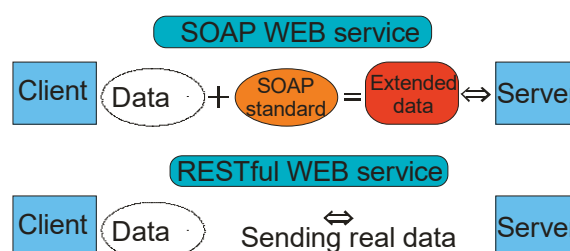

Fig.3. SOAP i RESTful Web services

RESTful web services are based on architecture called REST - Representational State Transfer. REST is not dependent on any protocol, but almost every RESTful service uses HTTP as the basic protocol. These services are much better integrated with HTTP than SOAP services, and as such do not require XML, SOAP messages, or WSDL definitions. The introduction of REST Web services into an attempt is to overcome the complexity of SOAP Web services. Basically, there is a big difference between these two types of Web services and does not exist so that the SOAP Web service using the HTTP protocol for messaging is one specific case of the REST Web service[11].

VI. PROPOSED SYSTEM AND ARCHITECTURE

In order to address the all mentioned issues of flexibility and functionality of IoT in this paper, we present one suggestion for standard IoT application through a layered structure of IoT. Each layer in this structure has its own defined task to be performed. The basic architecture of a IoT based application can be considered through four different layers (see Fig.4):

a. **Perception Layer** - consists of different connected physical sensors which are supposed to be used for monitoring events in our application.

b. **Interface Layer** – represent control units which are collecting data from the perception layer. This unit mainly consists of different tiny, inexpensive wireless sensor nodes which are network organized and are deployed over a wide

geographical area, capable to integrate continuous and unobtrusive measurement, computing and wireless communication, completely autonomously.

c. **Network Layer** - The processing unit leads the accumulated data into the Network Layer, which provides two way communication and connectivity pathway via Cloud or WEB infrastructure, after which the data finally reaches the Application Layer and end user.

 d. **Application layer** - The Application Layer via which control and monitoring of the appliances, installed at mobile, lap-top or tablet, takes place from anywhere in the world.
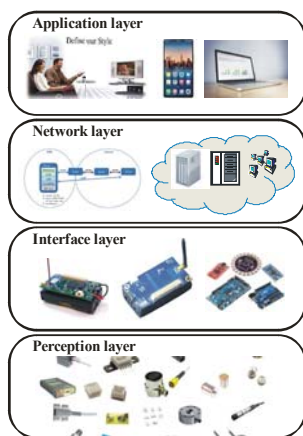
Fig. 4. Layer architecture of IoT application

The architecture presented in this paper can be customized in different ways in order to accommodate different application scenarios with minimum recoding and design. This system allows authorized home owners to remotely control and monitor events via smart phone application with graphical user interface (GUI). The proposed system represents a classical client-server application based on Arduino microcontroller system, Wi-Fi (ESP2866 shield), ThingSpeak platform and the Android compatible mobile application. The ESP8266 WiFi Module is a self-contained SOC with integrated TCP/IP protocol stack that can give any microcontroller access to your home WiFi network. It offers a complete and self-contained Wi-Fi networking solution, allowing it to either host the application or to offload all Wi-Fi networking functions from another application processor. This shield has the capability to be used both, as a client or a server. Since Arduino already supports a TCP/IP stack, we have focused on implementing software to connect it to the remote user. Software of the proposed IoT application is divided into three main parts: client application on smart mobile, server software application on some Web or Cloud devices and microcontroller firmware. The client application for Android system can be done using Google Android Programming IDE - App Inventor and Java programming language. The client application communicates with the microcontroller through ThingSpeak cloud. ThingSpeak is an open-source IoT platform that lets you collect and store sensor data in the cloud and develop IoT applications. This platform has its own API to store and retrieve data from sensor nodes using the HTTP protocol and let you to analyze and visualize

your data in MATLAB, and then act on the data. Sensor data can be sent to ThingSpeak from different microcontroller system such as: Arduino, Raspberry, BeagleBone Black or other hardware[12]. Using the REST API request methods such as GET, POST, PUT, and DELETE, we can create a channel and update its feed, update an existing channel, clear a channel feed, and delete a channel. We sent a Java Script Object Notation (JSON) GET Request to ThingSpeak by using REST API Web Service and channel ID and field number within its parameters. We received the response from ThingSpeak in JSON format and populated the tables in the Android application by using JSON Parser.

VII. Conclusion

In this paper we discuss current technologies that enable us to take advantage of all IoT applications. The fact that we can present a huge amount of collected data from different sources to a large number of clients, provides unprecedented opportunities for the development of a huge number of applications. On the other hand, the simplicity of building an IoT application is simply fascinating. We plan further research to be in the direction of implementing a CoAP protocol into IoT application as well as the benefits that can be gained from considering large amounts of data – Big Data analytics.

References

[1] D. Evans, The Internet of Things [INFOGRAPHIC], http://blogs.cisco.com/diversity/the-internet-of-things-infographic/ pos. 25.02.2019

[2] Ian G. Smith, *The Internet of Things 2012 New Horizons*, Halifax, UK, 2012

[3] J. Gubbia, R.Buyyab, S.Marusic, M.Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions", Future Generation Computer SysteSN 29, 2013, pp.1645–1660

[4] Adrian McEwen, Hakim Cassimally, "Designing the Internet of Things", John Wiley and Sons, 2014

[5] Raghunathan V., Ganerival S., Srivastava M., "Emerging techniques for long lived wireless sensor networks", IEEE Communication Magazine, Vol.44, No.4, 2006

[6] M. Beraka, H. Mathkour, S. Gannouni, H. Hashimi, "Applications of Different Web Service Composition Standards," *IEEE International Conference on Cloud and Service Computing (CSC)*, pp.56-63, Nov. 2012

[7] M.Kosanovic, M.Stojcev, "Connecting Wireless Sensor Networks to Internet"*, Facta Universitatis, Series: Mechanical Engineering Vol. 9, No 2, 2011, pp.169-182

[8] J. Rellermeyer at all, "The Software Fabric for the Internet of Things", In Proceedings of the First International Conference on the Internet of Things, Zurich, 2008

[9] D. Guinard, V. Trifa, T. Pham, O. Liechti, "Towards Physical Mashups in the Web of Things", Proceedings of INSS 2009, Pittsburgh, USA, 2009.

[10] Shelby, Z.; Hartke, K.; Bormann, C., "The Constrained Application Protocol (CoAP) "; RFC 7252; The Internet Engineering Task Force: Reston, VA, USA, June 2014

[11] Leonard Richardson and Sam Ruby, "RESTful Web Services", O'Reilly Media, Inc., May 2007

[12] Marco Schwartz, "Internet of Things with Arduino Cookbook",Packt Publishing, 2016

5G System Support for Mission Critical Communications

Evelina Pencheva[1], Aleksander Nametkov[1], Denitsa Velkova[1], Ventsislav Trifonov[1]

Abstract – **Fifth generation (5G) mobile networks will provide flexibility to customize quality of service for diverse use cases. This enables a wide range of mission critical services, including autonomous vehicles, industrial control, robotics, telesurgery, augmented/virtual reality, etc. These services require very high reliability and availability, and low latency. In this paper, we present the 5G developments to support mission critical communications. Topics discussed in brief include advances in radio access, network slicing, network programmability, and Device-to-Device communications. Utilization of Multi-access Edge Computing in 5G is considered in the context of mission critical services.**

Keywords – **Latency, Reliability, Radio access, Network slicing, Quality of Service, Multi-access Edge Computing.**

I. INTRODUCTION

As to 3GPP definition, mission critical communications require high reliability and availability, low setup and transfer latency, ability to handle large number of devices, priority handling and strong security. The mission critical communications serve public safety, utility sectors, railways and other vertical segments [1]. Historically, this type of communications have been provided by dedicated private mobile radio technologies (e.g. TETRA, GSM-R, iDEN, etc.) which are low capacity, narrowband, fragmented and are de – facto standards. The demand for broadband data applications and location-based services, and the aim to achieve high efficiency make four generation Long Term Evolution (LTE) and LTE-Advanced (LTE-A) reference technology for mission critical applications. LTE is widely supported by mobile network operators which may provide mission critical service by their commercial LTE networks [2], [3]. In comparison with private mobile radio technology, LTE provides higher data rates and better security mechanisms. However, LTE is not designed to comply with reliability, confidentiality and security requirements of mission critical services.

The upcoming 5G systems promise to support the growing needs for enhanced Broadband Communications (eBBC), massive numbers of connected devices (mMTC), and ultra-reliable, low latency communications (URLLC). URLLC services expose strict requirements on low latency and reliability for mission critical communications. Ultra-reliability and low latency are vital for applications such as intelligent transportation, telesurgery, industry automation etc. In this paper, we outline the vision and requirements of mission critical communications in 5G, and discuss 5G technologies that aim to support mission critical applications.

[1]The authors are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: enp@tu-sofia.bg.

The rest of the paper is organized as follows. Next section describes the different types of mission critical services as defined by 3GPP and their requirements. Section III discusses 5G developments aimed to achieve ultra-reliability and low latency. Section IV describes the 5G quality of service mechanisms for provisioning of high availability of mission critical services. Section V presents the capabilities of Multi-access Edge Computing to support mission critical applications. The conclusion summarizes the contribution.

II. MISSION CRITICAL SERVICE TYPES AND REQUIREMENTS

3GPP defines three types of mission critical services [4]. The Mission Critical Push-to-Talk (MCPTT) service supports communication between several users (i.e. group call), where each user has the ability to gain access to the permission to talk in an arbitrated manner. The MCPTT service also supports private calls between two users. The Mission Critical Video (MCVideo) services and Mission Critical Data (MCData) can be used for public safety applications and maritime safety applications and also for general commercial applications (e.g. utility companies, railways and maritime usage). MCVideo defines a service for mission critical video communication using 3GPP transport networks. MCData defines a service for mission critical data services which needs to provide a means to manage all data connections of mission critical users in the field and provide relevant resources to the ones who need it.

Mission critical (MC) refers to meeting the needs of agencies providing Public Safety services such as, but not limited to, Police, Fire department, and Paramedic services. Those needs include high reachability, availability and reliability of the service, low latency, real-time operating capabilities, highly secured operations, inter-operability with other services and systems, private and group communications, handling of emergencies and ability to provide prioritization, pre-emption, queuing and quality of service (QoS). Other examples of MC use cases include autonomous vehicles, industrial machineries in smart factories, virtual/augmented reality, remote early warning sensors, remote radar system and many others.

End-to-end (E2E) latency requirements need to be less than 1 ms and can be achieved by faster, more flexible frame structure and new non-orthogonal uplink access. Ultra-high reliable transmissions can be time multiplexed with nominal traffic through puncturing, and additional base stations for public safety coverage. Ultra-high reliability requires packet error rate (PER) less than 10^{-9} and can be achieved by simultaneous links to both 5G and LTE for extreme mobility and fault tolerance. E2E security is provided by security extensions to air interface, core network and service layer. Fig.1 shows the requirements of some MC use cases.

	Augmented/ Virtual Reality	Industry Control	Autonomic vehicle	Remote robotics/ surgeries
Reliability	Reliable failure detection	Up to 10^{-9} PER	Up to 10^{-6} PER	Up to 10^{-9} PER
E2E latency	< 5 ms	<0.5 ms	<5-10 ms	< 1ms
Specifics	High data rates	Often separate	High mobility	Tactile communi-cations

Fig.1 Requirements of mission critical use cases [5]

Common requirements to different types of MC services are defined in [5]. Efficient group communications are essential for different professional organizations. Geographic groups are based on the location and allow communication in a certain area. Functional groups are created for specific purposes and have to be available anywhere in the network. Fit-to-purpose groups are dedicated to specific tasks within a certain area. Efficient group management is required.

A mission critical system must prioritize different communication groups, e.g. Emergency Group Communications and Imminent Peril Communications provide users with higher priority in resource allocation. Users may have different mission critical applications, e.g. MCPTT, MCVideo, and/or MCData. These applications have different requirements for traffic handling, so users can have different prioritization levels for different applications. Furthermore, network may experience high level of traffic in extreme conditions which are unpredictable, and it must be able to adapt to exceptional MC load.

Simultaneous registrations of multiple devices have to be supported and the service administrator has to be able to limit the number of concurrent logs. Provisioning of location information is essential for MC services. Security requires support the confidentiality and integrity of all user traffic and signalling at the application layer.

III. DEVELOPMENTS FOR MISSION CRITICAL COMMUNICATIONS IN 5G

The 5G introduces new developments to address the requirements of MC communications.

A. Radio Access

The radio access developments for ultra reliability include redundancy by usage of Massive Multiple-Input, Multiple-Output (MIMO) and multi-connectivity. Massive MIMO technology groups the antennas at the transmitter and at the receiver to achieve better spectral efficiency and throughput. 5G Radio Access Network (RAN) will support multi-connectivity for tight interworking among different 5G radio variants (below and above 6GHz carrier frequency), between 4G and 5G, and between different transmission points and carriers of the same radio variant [5].

Low latency is achieved by service-specific optimization of the protocol stack. Multi-connectivity and carrier aggregation are important for ultra-high reliability. In 5G, the functionality of Packet Data Convergence Protocol (PDCP) and Radio Link Control (RLC) are grouped in upper layer, and Media Access Control (MAC) forms lower layer which results in processing and latency gain due to less header processing and function optimization. Multi-connectivity and carrier aggregation is important for URLLC. With the proposed layering upper layer is able to connect to multiple lower layer entities, where the re-transmissions may be done at PDCP level. Further, segmentation is not required due to small and typically fixed packet sizes. Ciphering may be skipped if application level security mechanisms are applied for mission critical service.

B. Network Slicing

Both the 5G RAN and core network should support network slicing. Network slicing is introduced to support the required flexibility in provisioning of QoS required by different services [6]. Network slices are logical networks deployed over the same physical infrastructure. Each network slice provides specific network capabilities and network characteristics. For example, URLLC specific RAN functions include optimal handling of Radio Resource Control states to reduce state change latency, applying acknowledge mode only, prioritization of Random Access Channel, optimized coding for short payloads, and potential omitting of ciphering and header compression.

Network slicing supports slice isolation which prevents from distribution fault or security related events in one slice into other slices.

C. Network Vitualization

Network virtualization means network transition from hardware mode to software mode.

Both 5G RAN and 5G core network designs adopt Software Defined Networking (SDN). SDN splits the control plane and data plane. SDN may be applied to the following types of network functions: network control functions (session management, mobility management and QoS control), connectivity management (packet forwarding) and wireless control functions (such as scheduling and radio link adaptation). These network functions are performed by programmable and logically centralized controller which provides network technology independence.

5G networks will adopt Network Function Virtualization (NFV) also. NFV is a concept for replacing dedicated network devices, such as routers and firewalls, with software running on general purpose servers. It optimizes creation, activation, and provision of services by using the cloud advancements.

D. Device-to-Device Communications

The increasing demand for multimedia communications with high bandwidth and low latency requirements becomes a challenge for cellular networks. A new paradigm that may

face this challenge is Device-to-Device (D2D) communications. D2D communications allow devices in close proximity to communicate directly without involvement of network elements processing of traffic data. D2D communications may address ultra-reliable and low latency requirements through direct communications between devices and thus reducing latency [7]. For MC communications, it is essential to ensure that the communication service is also provided if the network or parts of it are congested. D2D communication may address also reliability. In case of network congestion, a device with good internet connectivity may cash the data and transmit it to other devices. The network coverage may also be expanded by a device acting as a relay between the base station and other devices.

IV. 5G QoS MECHANISMS FOR HIGH SERVICE AVAILABILITY IN 5G

In order to support a large diversity of use case requirements, 5G has to exploit flexible and highly granular end-to-end means to prioritize different traffic types, and to handle appropriately the packets belonging to the same flows. As an example for the latter, it is necessary to assign higher priority to short packet transferring protocol requests and responses than the other protocol packets, in order to avoid delays in connection setup time.

With respect to this context the key limiting factor of LTE QoS architecture is that the finest granularity to distinguish between data is on radio bearer level. Fig.2 shows the QoS architecture of 4G. Evolved packets System (EPS) bearer represents a level of granularity for QoS control which provides a logical transmission path with well defined QoS properties. The EPS bearer is mapped to QoS concepts of the underlying transport. Each EPS bearer is transported over the radio bearer with corresponding QoS characteristics. The EPS bearer QoS is mapped onto Internet protocol (IP) transport layer QoS between Service gateway (SGW) (E-UTRAN radio Access Bearer - ERAB) and Packet Data Network Gateway (PGW) (S5/S8 bearer). The mapping is one-to-one, which means that one radio bearer has to carry data related to different service requirements, but can not treat the packets related to one session differently. Hence, a higher level of QoS granularity is required.
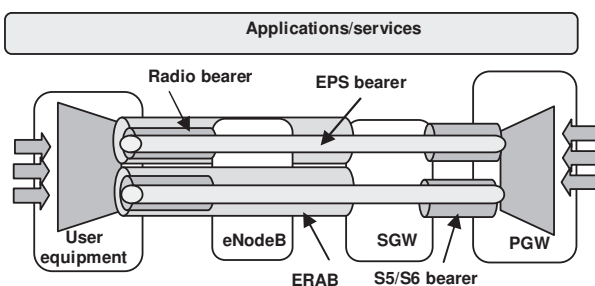
Fig.2 QoS architecture in 4G

The QoS architecture in 5G is based on the concept of QoS flow, which is identified by QoS flow ID (QFI) and is managed by Session Management Function (SMF) in the core

network [8]. In contrast to 4G, in 5G, multiple QoS flows may be assigned to one PDU session. Each QoS flow is characterized by QoS profile, which is provided in a form of QoS rules by SMF to the User equipment (UE). SMF provides to the User Plane Function (UPF) in the core network QoS identifier (5QI) which contains information about guaranteed and non-guaranteed bit rates, and Allocation and Retention Priority in downlink and uplink, respectively. QoS flows can be flexibly assigned to different data radio bearers. The UPF applies service data flow classification in downlink, and the UE evaluates the QoS rules provided by the SMF and assigns the appropriate radio data bearers. Fig.3 shows the QoS architecture in 5G.
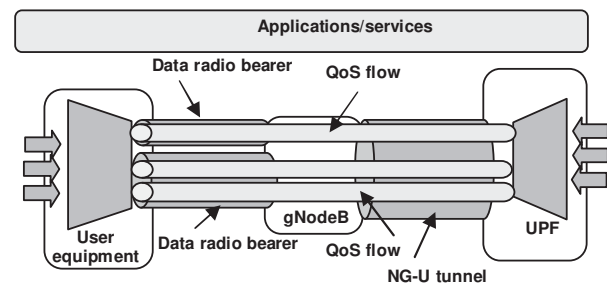
Fig.3 QoS architecture in 5G

V. MULTI-ACCESS EDGE COMPUTING

The E2E latency may be reduced by deployment of Multi-access Edge Computing (MEC). MEC extends the cloud capabilities close to the place where they are used. It introduces data centers at the network edge. MEC can improve user's quality of experience (QoE) and can guarantee maximum utilization of RAN resources. The vicinity to end users enables applications with high bandwidth and low latency requirements. The MEC server contains the mobile edge platform and a virtualized infrastructure which provides cloud intelligence for the mobile edge applications. The mobile edge platform offers an environment for mobile edge applications to discover and to use available mobile edge services, and handles the user plane according traffic rules. Mobile edge applications run in a well-isolated manner on the virtualized infrastructure. Different MEC applications hosted by the MEC platform may belong to different network slices configured in RAN and/or core network.

MEC service platform offers two types of services. Radio network information services provide authorized applications with low level real-time radio and network information related to users and cells. Traffic offload function prioritizes traffic and routes the selected, policy-based, user data stream to and from applications that are authorized to receive the data.

There are many potential scenarios for MEC deployment [9], [10]. Edge may be referred to both RAN nodes and aggregation points in distributed core networks which serve specific requirements. For example, the MEC platform and applications may sit on an entity in the RAN for device offloading scenarios, while the MEC platform and applications may reside close to dedicated distributed core

functional entities for scenarios where mobility and session continuity support is required.

The best way to receive radio network and location information and to manage the bandwidth is the "Bump in the wire" scenario where MEC is deployed much closer to RAN [10]. This scenario exposes the benefit from all MEC advantages such as low latency, efficient bandwidth management and local breakout. It is possible to bundle MEC platform and gNB (5G base station) into a single node where MEC can share the same network function virtualization infrastructure with Cloud Radio Access Network (C-RAN). When MEC is deployed in proximity of gNB or at aggregation point, the MEC platform is located on the N2 reference point between gNB and AMF, while the MEC data plane sits on the N3 reference point between RAN and UPF.

Fig.4 illustrates the MEC and C-RAN co-location deployment scenario. Fig.5 shows standalone MEC deployment at aggregation point.
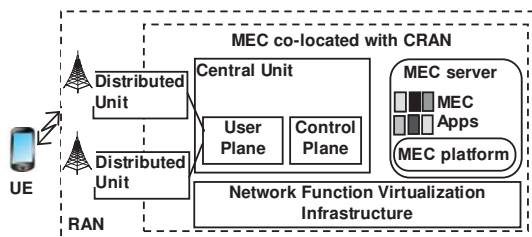
Fig.4 MEC and C-RAN deployment scenario
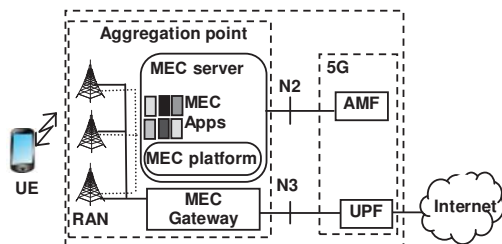
Fig.5 Standalone MEC deployment at aggregation point scenario

The MEC may be co-located with distributed core network functions (i.e. at the same site) [11], [12]. In this case, MEC applications may steer user plane traffic. This type of deployment is appropriate for MCPTT where communication with the core site is optional, as the traffic does not need to pass the backhaul to keep service running. This type of MEC deployment might usually be used by public safety, first responders, and mission critical industrial sites. Fig.6 shows the scenario.
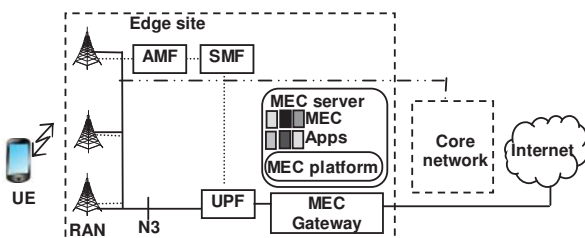
Fig.6 MEC deployment scenario with distributed core functionality

VI. CONCLUSION

In this paper, we consider challenges and developments for provisioning of mission critical communication in 5G networks. The requirements for ultra-reliability and low latency may be addressed by the advanced function exposed by 5G radio access, network slicing, network virtualization, D2D communications, enhanced QoS architecture and deployment of cloud computing at the network edge. 5G radio access supports multi-connectivity and protocol stack optimization. Network slicing, programmable networks and MEC further contribute to latency reduction and provisioning of full mobility and highest reliability.

REFERENCES

[1] Z. Kaleem, et al. "UAV-Empowered Disaster-Resilient Edge Architecture for Delay-Sensitive Communication", Cornell University, Networking and Internet Architecture, arXiv:1809.09617v2 [cs.NI], January, 2019.

[2] K. O. Olasupo, I. Kostanic and T. O. Olasupo, "Performance evaluation of mission critical communications services over LTE networks," IEEE ICPCSI, Chennai, 2017, pp. 273-278.

[3] J. Oueis, V. Conan, D. Lavaux, R. Stanica and F. Valois, "Overview of LTE Isolated E-UTRAN Operation for Public Safety," IEEE Communications Standards Magazine, vol. 1, no. 2, 2017, pp. 98-105.

[4] 3GPP TS 23.179 Functional architecture and information flows to support mission critical communication services, Stage 2, Release 13, v13.5.0, 2017.

[5] 3GPP TS 22.280 Mission Critical Common Requirements (MCCoRe), Stage 1, Release 16, 2018

[6] E., Pencheva, I. Atanasov, D. Kireva, K. Nikolova, "Network Slicing: A Mobility Management Perspective," SGEM 2018, vol.18, Green Design and Sustainable Architecture, pp.641-648

[7] A. Orsino et al., "Effects of Heterogeneous Mobility on D2D- and Drone-Assisted Mission-Critical MTC in 5G," IEEE Communications Magazine, vol. 55, no. 2, pp. 79-87, 2017.

[8] Q. Ye, J. Li, K. Qu, W. Zhuang, X. S. Shen and X. Li, "End-to-End Quality of Service in 5G Networks: Examining the Effectiveness of a Network Slicing Framework," IEEE Vehicular Technology Magazine, vol. 13, no. 2, pp. 65-74, 2018.

[9] S. Kekki, et al. "MEC in 5G networks, " ETSI white paper no 28, June 2018.

[10] F. Guist, et al. "MEC Deploymnets in 4G and Evolution Towards 5G, "ETSI White paper, 2018.

[11] P. Skarin, W. Tärneberg, K. Årzen and M. Kihl, "Towards Mission-Critical Control at the Edge and Over 5G," 2018 IEEE International Conference on Edge Computing (EDGE), San Francisco, CA, 2018, pp. 50-57.

[12] R. Solozabal, A. Sanchoyerto, E. Atxutegi, B. Blanco, J. O. Fajardo and F. Liberal, "Exploitation of Mobile Edge Computing in 5G Distributed Mission-Critical Push-to-Talk Service Deployment," IEEE Access, vol. 6, 2018, pp. 37665-37675.

Warning Message Transmission as a RAN Service

Ivaylo Atanasov[1], Vladislav Vladislavov[1], Ivaylo Asenov[1], Evelina Pencheva[1]

Abstract – **Fifth generation (5G) mobile networks come with the promise to support diverse and demanding performance requirements of very different service types. In order to facilitate service delivery at a reasonable cost, 5G core network is service-oriented. In this paper, we propose an approach to enhance telecom protocol style interface between the Radio Access Network and the core network with web-based Application Programming Interfaces. Following the service-oriented architecture, the functions of Next Generation Application Protocol related to warning message transmission are represented as services, which enable more flexible deployment of new ones. The focus is on warning message transmission, defined as a service. Service description is provided and some implementation aspects are considered.**

Keywords – **5G, Service Oriented Architecture, Radio Access Network, REpresentational State Transfer.**

I. Introduction

Fifth generation (5G) mobile networks will provide three types of services:
- Enhanced Mobile Broadband (eMBB) with requirements of high bandwidth, indoor/hotspot and enhanced wide area coverage;
- Ultra-Reliable and Low Latency Communications (URLLC) with requirements of very low user plane latency, very high reliability and availability, and high mobility;
- Massive Machine Type Communications (mMTC) characterized by high device density and requirements for low power consumption.

In order to support the very wide range of services with diverse requirements, the 5G network architecture has to be flexible, modular and scalable, and has to support high level of programmability and automation in networks [1], [2]. Following this design principles, the 5G core network is centered around services accessible through Application Programming Interfaces (APIs). Service Oriented Architecture enables greater functional and service agility, as the introduction of new service or upgrade of existing one becomes simpler and facilitates the transition to the cloud model [3].

The concept of Every-Thing-as-a-Service is introduced in [4]. The authors argue that every component which used to be essential in the traditional network management including Radio Access Network (RAN) can be viewed as a service. In [5], a service oriented framework for RAN sharing is proposed. The aim is to facilitate radio resource sharing between different mobile operators. Adoption of service-based

[1]The authors are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: iia@tu-sofia.bg.

design facilitates slicing and orchestration of RAN [6].

In this paper, following the service-oriented design of 5G core network, we propose an approach to transition to service-based architecture adopted for the core network of RAN. The control protocol between the Next Generation RAN (NG-RAN) and the core network is Next Generation Application Protocol (NGAP) [7]. In our approach, NGAP functions are defined as services following modular function design. The focus is on warning message transmission function, which enables sending messages from the network to mobile subscribers.

The rest of the paper is organized as follows. Next section describes the functionality of the proposed service illustrated with typical use cases. Section III presents the API definition. Section IV considers some implementation aspects regarding modeling of service logic as seen by the core network and RAN. The conclusion summarizes the contribution.

II. Service Description

The proposed Warning Message Transmission Service (WMTS) exposes functionality of Warning Message Transmission Procedures of NGAP protocol. It may be used to send local content, service context, and real-time information on local access network conditions. The service supports write-replace warning, cancel, reloading of Public Warning System Message (PWS) broadcast, and PWS failure indication. The interaction between WMTS and the core network services follows the REpresentational State Transfer (REST) architectural style. In REST, each entity is represented as a resource. The resource is characterized by its state which can be manipulated by four operations implemented by HTTP requests, namely: CREATE (HTTP POST), READ (HTTP GET), UPDATE (HTTP PUT), and DELETE (HTTP DELETE).

The purpose of write-replace warning procedure is to start or overwrite the broadcasting of warning messages. It is initiated by Access and mobility Management Function (AMF) in the core network. When the AMF wants to start broadcasting of warning messages, it sends a POST request to the resource representing warning messages. The message body contains the content to be broadcasted, the repetition period, and the broadcast area. The NG-RAN node schedules a broadcast of the warning message and starts broadcasting. It acknowledges the request by sending "201 Created" message to the AMF. The message contains the URI (Uniform Resource Identifier) of the resource created. Fig.1 shows the procedure.
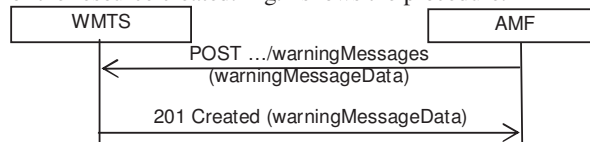

Fig.1 Flow of warning message broadcasting initiation

The purpose of the PWS cancel procedure is to cancel an already ongoing broadcast of a warning message. When the AMF wants to cancel the procedure it sends DELETE request to the resource representing the respective warning message. The NG-RAN node, in turn, stops broadcasting the warning message within that area and discards the warning message for that area, if the broadcasting area is included, or in all of the cells. The WMTS acknowledges the request by sending "204 No Content" response. Fig.2 shows the procedure.
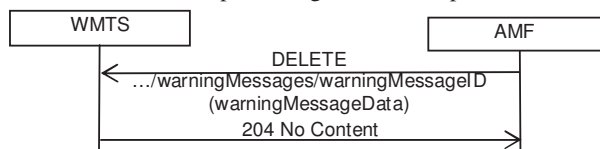
Fig.2 Flow of warning message broadcasting cancelation

The purpose of the PWS Failure Indication procedure is to inform the AMF that ongoing PWS operation for one or more cells of the NG-RAN node has failed. The purpose of the PWS restart indication procedure is to inform the AMF that PWS information for some or all cells of the NG-RAN node may be reloaded from the Cell Broadcast Center if needed. In order to receive indications, the AMF needs to make a subscription. When the AMF wants to receive PWS related indications it sends a POST request with the message body containing wmSubscriptionData data structure to the resource representing the respective subscription type. The subscription types may be one of the following: the pwsFailureSubscriptions type and pwsRestartSubscriptions type. The wmSubscriptionData data structure defines the event subscription, filtering criteria and the address where the AMF wants to receive indications. The WMTS sends "201 Created" response with the message body containing the address of the resource created and the subscribed PWS restart indication event subscription type. The above is illustrated in Fig.3.
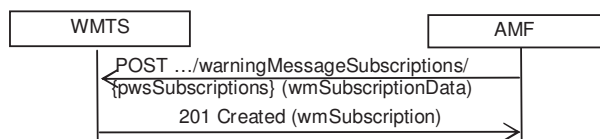
Fig.3 Flow of subscription for PWS related indications

The WMTS may define an expiry time for the PWS restart indication subscription, which is included in the response of subscription request. When subscription expiry time is used, prior the subscription expiry, the WMTS may notify the AMF, which in turn needs to update the subscription. The AMF updates the subscription by sending PUT request to the resource representing the respective subscription type with modified data structure specific to that PWS restart indication event subscription, as shown in Fig.4.

When the AMF does not want to receive indications about PWS broadcast related events anymore, it terminates the subscription. Fig.5 shows a scenario where the AMF uses REST based procedure to delete the subscription for indications about PWS broadcast indications.

Fig.4 Flow of subscription update for PWS related indications
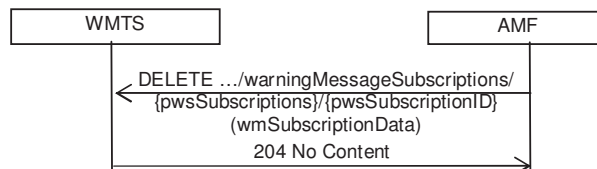
Fig.5 Flow of PWS restart indication subscription termination

When the NG-RAN node detects that ongoing PWS operation has failed for one or more cells, it sends a notification to inform the AMF. Fig.6 illustrates the scenario.
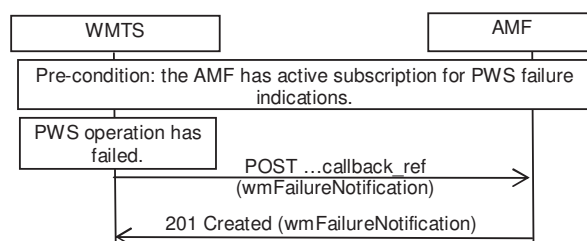
Fig.6 Flow of receiving indications about PWS failure

When the NG-RAN node detects that a PWS information for some or all cells may be reloaded from the Cell Broadcast Centre, the WMTS sends an indication to the AMF. Fig.7 illustrates the case.
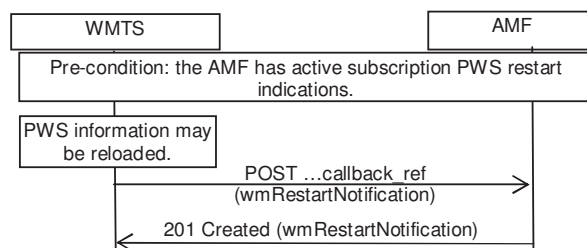

Fig.7 Flow of receiving PWS restart indications

Summarizing the WMTS use cases the following service operations may be defined:
- send_warning_message – the AMF invokes this operation to start broadcasting of warning message;
- cancel_warning_message – the AMF uses this operation to cancel the warning message broadcasting;
- start_pws_failure_indications and start_pws_restart_indications – the AMF uses these operations to create subscriptions for PWS failure and PWS restart indications respectively;
- stop_pws_failure_indications and stop_pws_restart_indications – the AMF uses these operations to terminate the subscriptions for PWS failure and PWS restart indications respectively;

- update_pws_failure_indications and update_pws_re-start_indications – the AMF uses these operations to update the subscriptions for PWS failure and PWS restart indications respectively;
- notify_pws_failure and notify_pws_restart – the WMTS uses these operations to send notification about PSW failure and PSW restart respectively.

III. INTERFACE DEFINITION

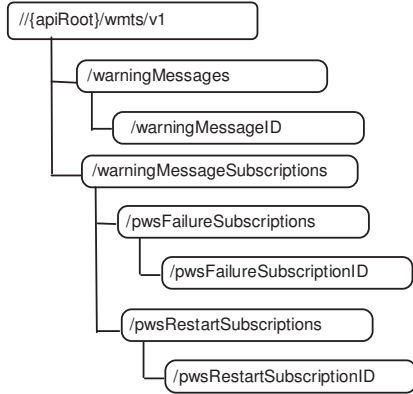Fig.8 presents the resource structure. All service resources have the following root: {apiRoot}/wmts/{apiVersion}/

Fig.8 Structure of resources supported by WMTS

The container resources warningMessages, warningMessage-Subscriptions, pwsFailureSubscriptions and pwsRestartSubscriptions support HTTP GET method which retrieves information about the resource, and HTTP POST method, which creates a new resource of the respective type. The leaf resources warningMessageID, pwsFailureSubscriptionID, and pwsRestart-SubscriptionID support HTTP GET the message method which retrieves information about existing resource, HTTP PUT method which updates information about existing resource, and HTTP DELETE method which deletes the existing resource.

IV. STATE MODELS

Deployment of WMTS in the network requires development of models, representing the warning message broadcast status as seen by the NG-RAN node and by AMF. Both models need to be synchronized.

Fig.9 shows the simplified model of the warning message broadcast status as seen by the AMF. In Broadcasted state, the warning message is broadcasted successfully or there is no warning message to be broadcasted. In Broadcasting state, the warning message broadcasting is ongoing. In PwsFailure state, the AMF has received a PWS failure indication.

Fig.10 shows a simplified model representing the warning message broadcast status supported by the NG-RAN node. In order to broadcast a warning message, the NG-RAN node needs to establish appropriate radio bearers. The purpose of Radio Resource Control (RRC) Reconfiguration procedure is

to modify an RRC connection, e.g. to establish/modify/release radio bearers [8].
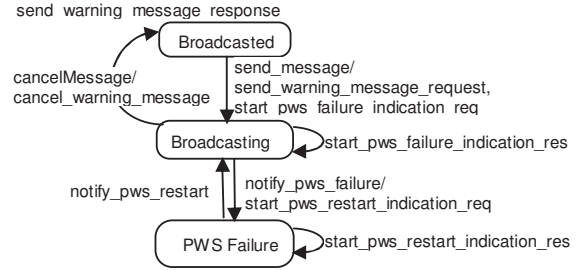
Fig.9 Model of warning message broadcast status, supported by AMF
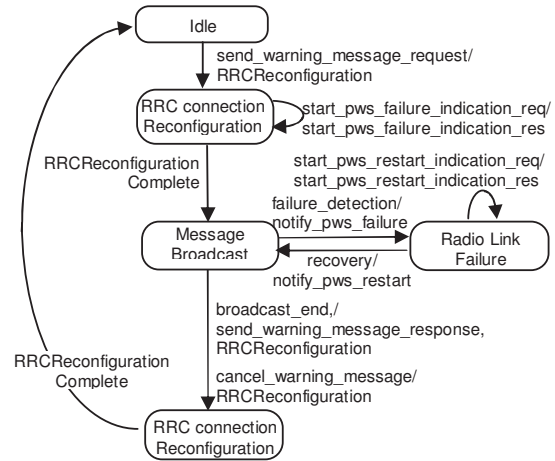
Fig.10 Model of warning message broadcast status, supported by NG-RAN node

In order to provide a more rigorous proof that both models are synchronized, i.e. expose equivalent behavior we formalize the models' descriptions. The model description is formalized using the notion of Labeled Transition System (LTS).

A Labeled Transition System is represented as quadruple of a set of states, a set of actions, a set of transitions, and a set of initial states.

By $T_{AMF} = (S_{AMF}, Act_{AMF}, \rightarrow_{AMF}, s_0^{AMF})$ it is denoted an LTS, representing the warning message broadcast state model supported by AMF, where:

- $S_{AMF} = \{$Broadcasted $[s_1^A]$, Broadcasting $[s_2^A]$, PWSFailure $[s_3^A]\}$;

- $Act_{App} = \{$send_message$[t_1^A]$, start_pws_failure_indica-tion_res$[t_2^A]$, notify_pws_failure $[t_3^A]$, notify_pws_re-start$[t_4^A]$, cancelMessage $[t_5^A]$, send_warning_me-ssage_response $[t_6^A]$, start_pws_restart_indication_res $[t_7^A]\}$;

- $\rightarrow_{AMF} = \{(s_1^A\ t_1^A\ s_1^A), (s_2^A\ t_2^A\ s_2^A), (s_2^A\ t_3^A\ s_3^A), (s_3^A\ t_4^A\ s_2^A), (s_2^A\ t_5^A\ s_1^A), (s_2^A\ t_6^A\ s_1^A), (s_3^A\ t_7^A\ s_3^A)\}$;

- $s_0^{\mathrm{AMF}} = \{ s_1^A \}$.

Short notations for states and actions are given in brackets.

By $T_{\mathrm{RAN}} = (S_{\mathrm{RAN}}, Act_{\mathrm{RAN}}, \rightarrow_{\mathrm{RAN}}, s_0^{\mathrm{RAN}})$ it is denoted an LTS, representing the warning message broadcast state model supported by NG-RAN node, where:

- $S_{\mathrm{RAN}} = \{$Idle [s_1^R], RRCConnectionReconfiguration [s_2^R],

MessageBroadcast [s_3^R], RadioLinkFailure [s_4^R]$\}$;

- $Act_{\mathrm{RAN}} = \{$send_warning_message_request [t_1^R],

start_pws_failure_indication_req[t_2^R], RRCReconfi-

gurationComplete [t_3^R], failure_detection [t_4^R], recovery

[t_6^R], start_pws_failure_indication_req [t_5^R],

broadcast_end [t_7^R], cancel_warning_message [t_8^R]$\}$;

- $\rightarrow_{\mathrm{RAN}} = \{(s_1^R\ t_1^R\ s_2^R), (s_2^R\ t_2^R\ s_2^R), (s_2^R\ t_3^R\ s_3^R),$

$(s_3^R\ t_4^R\ s_4^R), (s_4^R\ t_5^R\ s_4^R), (s_4^R\ t_6^R\ s_3^R), (s_3^R\ t_7^R\ s_2^R),$

$(s_3^R\ t_8^R\ s_2^R), (s_2^R\ t_3^R\ s_1^R)\}$;

- $s_0^{\mathrm{RAN}} = \{ s_1^R \}$.

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. The weak bisimilarity admits internal transitions to be ignored.

Proposition: T_{AMF} and T_{RAN} are weakly bisimilar.

Proof: As to definition of weak bisimulation, it is necessary to identify a relation between the states of both LTSs, such as for any transition from a state in one LTS there are respective transitions from states in the other LTSs.

By U_{AMFRAN} it is denoted a relation between the states of both LTS, where $U_{\mathrm{AMFRAN}} = \{(s_1^A, s_1^R), (s_2^A, s_3^R), (s_3^A, s_4^R)\}$. Then, the following relationship between the $\rightarrow_{\mathrm{AMF}}$ and $\rightarrow_{\mathrm{RAN}}$ exists:

1. The AMF initiates warning message broadcast and subscribes for PWS failure indications: for ($s_1^A\ t_1^A\ s_1^A$), ($s_2^A\ t_2^A\ s_2^A$)$\exists$ ($s_1^R\ t_1^R\ s_2^R$), ($s_2^R\ t_2^R\ s_2^R$), ($s_2^R\ t_3^R\ s_3^R$).

2. During message broadcasting a failure occurs and the AMF is notified. The AMF subscribes for PWS restart indications: for ($s_2^A\ t_3^A\ s_3^A$), ($s_3^A\ t_7^A\ s_3^A$) $\exists$ ($s_3^R\ t_4^R\ s_4^R$), ($s_4^R\ t_5^R\ s_4^R$).

3. The problem is fixed and the AMF is notified that the PWS broadcast can restart: for ($s_3^A\ t_4^A\ s_2^A$) $\exists$ ($s_4^R\ t_6^R\ s_3^R$).

4. The warning message broadcast completes successfully: for ($s_2^A\ t_6^A\ s_1^A$) $\exists$ ($s_3^R\ t_7^R\ s_2^R$), ($s_2^R\ t_3^R\ s_1^R$).

5. The AMF cancels warning message broadcasting: for ($s_2^A\ t_5^A\ s_1^A$)$\exists$($s_3^R\ t_8^R\ s_2^R$), ($s_2^R\ t_3^R\ s_1^R$).

Therefore T_{AMF} and T_{RAN} are weakly bisimilar, i.e. they expose equivalent behavior. ■

V. Conclusion

In this paper we propose an approach to design RAN application level functions as services. The research is based on functionality of the NGAP between the core network and NG-RAN. We focus on warning message transfer procedures and define a new service. The procedures are described as service interactions. The resource structure is presented and the respective APIs are defined. Service implementation is discussed by modeling the warning message broadcast status as seen by the NG-RAN and core network.

Service-based design improves modularity, facilitates the exposure of network functions to 3rd party applications, and the transition to cloud architecture.

Acknowledgement

References

[1] T. Do and Y. Kim, "Usage-Aware Protection Plan for State Management Functions in Service-Based 5G Core Network," *IEEE Access*, vol. 6, pp. 36906-36915, 2018.

[2] L. Ma, X. Wen, L. Wang, Z. Lu and R. Knopp, "An SDN/NFV based framework for management and deployment of service based 5G core network," *China Communications*, vol. 15, no. 10, pp. 86-98, Oct. 2018.

[3] C. Zhang, X. Wen, L. Wang, Z. Lu and L. Ma, "Performance Evaluation of Candidate Protocol Stack for Service-Based Interfaces in 5G Core Network," 2018 IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, 2018, pp. 1-6.

[4] Z. Chang, Z. Zhou, S. Zhou, T. Chen and T. Ristaniemi, "Towards Service-Oriented 5G: Virtualizing the Networks for Everything-as-a-Service," IEEE Access, vol. 6, pp. 1480-1489, 2018.

[5] J. He, W. Song, SOARAN: A Service-oriented Architecture for Radio Access Network Sharing in Evolving Mobile Networks,' pp.1-16, arXiv:1508.00306v1 [cs.NI], Aug 2015.

[6] W. Guan, X. Wen, L. Wang, Z. Lu and Y. Shen, "A Service-Oriented Deployment Policy of End-to-End Network Slicing Based on Complex Network Theory," *IEEE Access*, vol. 6, pp. 19691-19701, 2018

[7] 3GPP TS 38.413 Technical Specification Group Radio Access Network; NG-RAN; NG Application Protocol (NGAP), v15.2.0, 2019, Release 15.

[8] 3GPP TS 38.331 Technical Specification Group Radio Access Network; NR; Radio Resource Control (RRC) protocol specification, Release 15, v15.3.0, 2018.

[9] Chen, X. J., De Nicola, R. Algebraic characteristics of trace and decorated trace equivalences over tree-like structures, Theoretical Computer Science, vol.254, issue 1-2, 2001, pp.337-361.

IoT System for Monitoring Conditions in the Human Environment

Dejan Vujičić[1], Dušan Marković[2], and Siniša Ranđić[3]

Abstract – **This paper illustrates the technology impact on the quality of the human environment. The importance of the environment on human health and his working and living activities requires the constant tracking of the ecosystem changes. The aim is to identify possible causes of the environmental degradation and act upon them. IoT devices represent the ideal means for monitoring parameters that determine the quality of the environment. The possibility of their direct connection to the Internet is well suited for gathering information about the environmental status, as well as the time and place of its changes.**

Keywords – **Human environment, monitoring, IoT devices, Internet**

I. Introduction

The development of semiconductor technology has enabled the realization of many concepts in the area of computer science and telecommunications. At the same time, the Internet has brought a revolution to human communication. The Internet was preceded by the ARPA network [1]. Since the WWW (World Wide Web) Internet realization, the Internet has become the communication platform as it is used today [2]. Simultaneously, new concepts in computing have emerged. One of them is IoT (Internet of Things) [3][4]. IoT devices revolutionized the ways of data gathering and distribution to the end users. With this concept, it was made possible for different physical devices to become part of the Internet. The IoT devices possess relatively large processing power, small dimensions, and low power consumption.

The emphasis has to be made on the possibility of wireless connections. With this in mind, the IoE (Internet of Everything) concept has become popular [5]. IoE means that every device, including the human itself, can be connected to the Internet. By adding sensors to the IoT devices, they become systems for measurements and data acquisition. On the other hand, by adding actuators, the IoT devices become Internet controlled elements. For these devices to work as intended, it is necessary to have corresponding Internet applications. The IoT devices have made their application in the areas of health care, agriculture, environment protection, and industry [6].

The technological and industrial development also had some negative impact on the human environment [7]. The emission of large quantities of CO_2 gas is one of the main causes of the global warming process. The bad management in the application of modern technologies and lack of control have contributed to the high level of environmental pollution.

As the consequence of the large industrial activity, the extortion of the natural recourses, such as coil, oil, and wood is happening. In the process of burning off these resources, the large amount of toxic gases and materials is produced. At the same time, by cutting the forests, the possibility of toxic gases absorption is reduced, and the conditions for landfalls and floods are emerging.

The importance of environmental degradation requires making actions on its identification and prevention. With this in mind, the organizational and technical actions can be made. The organizational actions require the change in human behavior, as the most important factor in environment pollution. These would be the cheapest and the most effective actions, but the practice has shown that these changes are very slow. That is why technical actions are necessary. They include the identification of the factors affecting the environment the most, with the places where it is occurring.

IoT devices have shown to be the ideal platform for monitoring support of the parameters that define the environment quality [8][9]. The IoT devices characteristics enable the remote usage of these systems. It was made possible by implementing battery power and different communication capabilities. It is of great importance that IoT devices possess the possibility of wireless connection to the Internet.

This paper shows the example of the IoT device project that can be used for monitoring environmental parameters. The proposed system has a great level of generality, because of the possibility of adding different sensors to the same computer platform, without making changes to the device communication subsystem. This IoT device has capabilities for temperature, humidity, and CO_2 levels indoors monitoring.

The remainder of the paper is organized as follows. In the second section, the influence of CO_2 levels on the human environment is described. The main emphasis is given to the indoor conditions, and brief literature review with some conclusions and recommendations regarding allowed CO_2 levels are given. The third section describes the practical implementation of the IoT monitoring system, which is consisted of a Wi-Fi capable module with sensors for measurements of CO_2 levels, and temperature and relative air humidity. The final section brings concluding remarks.

[1]Dejan Vujičić is with the University of Kragujevac, Faculty of Technical Sciences Čačak, Sv. Save 65, Čačak 32000, Serbia, E-mail: dejan.vujicic@ftn.kg.ac.rs

[2]Dušan Marković is with the University of Kragujevac, Faculty of Agronomy Čačak, Cara Dušana 34, Čačak 32000, Serbia, E-mail: dusan.markovic@kg.ac.rs

[3]Siniša Ranđić is with the University of Kragujevac, Faculty of Technical Sciences Čačak, Sv. Save 65, Čačak 32000, Serbia, E-mail: sinisa.randjic@ftn.kg.ac.rs

II. THE INFLUENCE OF CO_2 LEVELS ON THE HUMAN ENVIRONMENT

The indoor CO_2 concentration has greater value than outdoor concentration since humans directly produce CO_2 in the process of respiration. The ventilation systems require additional power consumption. The CO_2 levels outdoors usually are around 380ppm (parts per million), while in urban surroundings this value is somewhat greater and is around 500ppm [10].

The indoor CO_2 concentration is beginning at the values for outdoor levels and up to several thousand ppm. The significant importance on human health have concentrations over 20,000ppm. For indoor values, it was common consideration that CO_2 levels up to 5,000ppm do not have significant importance on human health, perception, and working performance. However, the research conducted in Hungary, published in 2003, questioned this hypothesis [11]. These authors presented that controlled exposure to the CO_2 concentrations between 2,000ppm and 5,000ppm had small negative impacts on human activities, but the description lacked details. The aforementioned data have inspired authors of [12] to investigate the influence of the different CO_2 levels on human cognitive behavior.
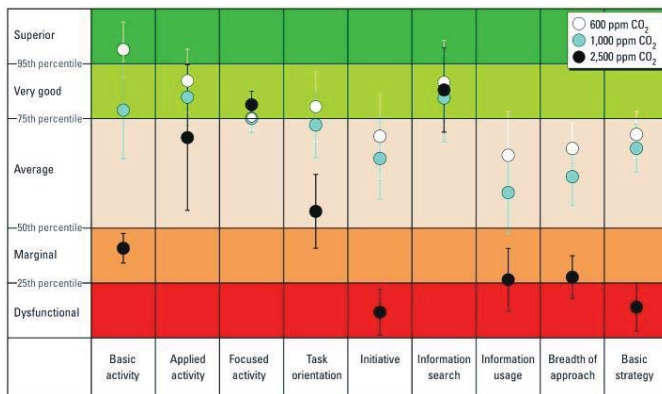
Fig. 1. Impact of CO_2 concentrations on human decision-making performance [12].

The results of this research are given in Fig. 1. The cognitive performances of the test subjects were diminished in 6 out of 9 observed parameters, with a CO_2 concentration of 1,000ppm, as opposed to the performance on 600ppm CO_2 levels. With CO_2 concentration of 2,000ppm, the 7 out of 9 parameters were diminished, as opposed to the performance on 600ppm CO_2 levels. The percentages on some parameters were degraded to the dysfunctional levels, shown in red in Fig. 1.

The research in [12] indicates that indoor CO_2 concentration can reach levels that can have a significant impact on human activities and decision-making performance. This is why is of great importance to track CO_2 levels in order to establish proper working and living conditions, and on the other hand, to determine optimal ventilation system support.

Indoor air quality not only affects human health but also human behavior. In [13], the laboratory conditions adapting to

the experimental needs were observed. The system for monitoring laboratory conditions was realized with adequate sensors and low-cost microcontrollers, and based on IoT system architecture. The wireless module used was ESP8266. The data acquisition system is supported by web and mobile applications. This system has the capability to notify the users that laboratory conditions have changed and need improvement.

The authors of [14] presented the system for IoT monitoring of CO_2 levels as air quality indicator. The presented system gave satisfactory estimations of the indoor air quality. It also supported the possibility of alarming if the air quality was unsatisfactory.

CO_2 monitoring was used to track elder persons' activity and indoor presence in [15]. Based on the CO_2 values, one can determine if the room is empty. If there is no one in the room, the CO_2 levels are near 650ppm, while if there is one person in the room, the CO_2 levels are around 850ppm. This change in CO_2 levels can be used to determine if a person is present in the room.

Fire prevention is another important factor for measuring indoor CO_2 levels. IoT based system for smart houses is presented in [16] and is based on the wireless sensor networks in order to early detect fire.

III. IoT MONITORING SYSTEM DESCRIPTION

There are many sensors for CO_2 measurement that use NDIR (nondispersive infrared) principle and that are sufficiently precise. Their price is near $40, while the sensors such as VERNIER CO2-BTA [17] have the price of near $340. Because of its good characteristics and low price, for the presented model, the MH-Z19 [18] sensor is used (Fig. 2). This module has small dimensions and uses NDIR technique for CO_2 air detection.
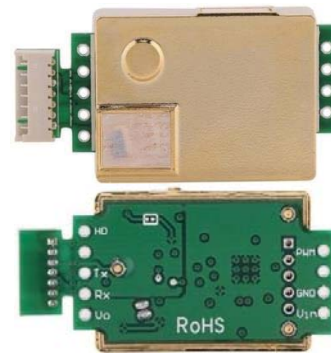

Fig. 2. MH-Z19 sensor, frontend and backend.

There are two variants of this sensor with different measurement ranges. One has the range from 0ppm to 2,000ppm, while another has the measurement range from 0ppm to 5,000ppm, with adequate temperature compensation. The working voltage of the module is from 4.5V to 5.5V, while the interface voltage level is at 3.3V, but also supporting the 5V levels. The dimensions of the module are 33mm × 20mm × 9mm (L×W×H), which is shown in Fig. 3.
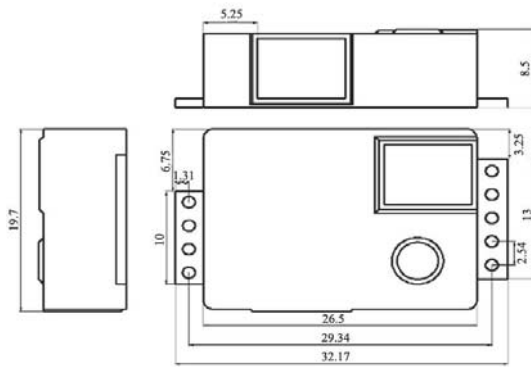
Fig. 3. MH-Z19 module dimensions with measures in millimeters.

The appearance of an IoT node used for monitoring environmental conditions is shown in Fig. 4. The system is based on ESP8566-12E module that has Wi-Fi connectivity to the Internet. This module is connected with the MH-Z19 CO_2 sensor and DHT21 sensor for measuring temperature and relative air humidity.
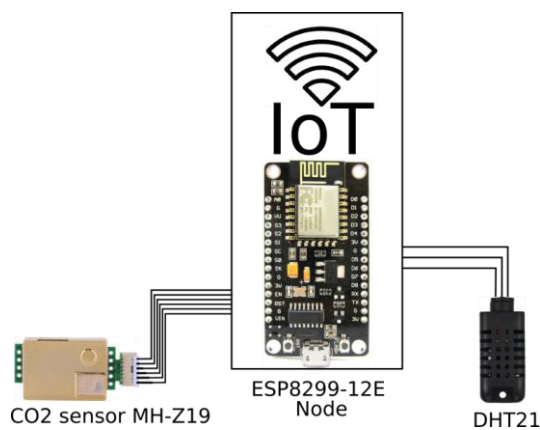

Fig. 4. IoT node with ESP8299-12E module, MH-Z19 CO_2 sensor, and DHT21 temperature and relative air humidity sensor.

The measurements of CO_2 levels, temperature, and relative humidity are taken at certain time intervals and distributed to the server on the Cloud. The data can be accessed with a corresponding web application where users can view chronological measurement data. Based on these data, users can take appropriate actions and recommendations. Depending on the type of the room, a number of persons in it and time spend indoor, the CO_2 concentration levels can be used to improve air quality in the room.

The acquired data from the system indicated that in ventilated rooms CO_2 concentration is near 640ppm, while if there are persons present in the room, the CO_2 levels can reach 1,200ppm. The goal of the realized system is to monitor the environmental conditions and gather information especially in cases where CO_2 levels reach 2,000ppm. This is the case when several persons are present in the non-ventilated room for some time.

The realized system can give clear indicators of the indoor air quality. This should be enough for users to take actions upon these results to ventilate the room. In this way, the optimal ambient conditions can be achieved, which may lead to a positive influence on working and living surrounding.

IV. CONCLUSION

The technological development has both positive and negative impact on human life. It is especially seen in environment quality degradation. Global warming and toxic gases emission are just some of the consequences of modern technologies usage. The monitoring of the parameters that identify the degradation of the environment has become practically imperative. Likewise, the possibility of locating the place where the environmental changes occurred is also very important.

The development of the intelligent sensors has greatly impacted on the environmental quality monitoring. The possibility of wireless communication can be here emphasized. The IoT systems further support environmental quality monitoring by being able to send raw data and make them publicly accessible on the Internet. With this in mind, the conditions have developed that lead to the open data system, for access by all the interested parties.

The realized example of IoT device, shown in this paper, has specifically emphasized the role of CO_2 monitoring, beside the temperature and relative air humidity. The CO_2 concentration levels are very important in determining the conditions in which people can live and work, without any hazards on their health and cognitive behavior. This IoT system is modular, with the possibility to add many other sensors in the future. This means that computing and communication subsystems rest the same, and only thing that is left to adjust is the software support for new sensors.

ACKNOWLEDGMENT

REFERENCES

[1] S. J. Lukasik, "Why the Arpanet was Built", IEEE Annals of the History of Computing, pp. 4 – 20, July – September 2011.

[2] T. Berners – Lee, "Information Management: A Proposal", Internal Report CERN, pp. 1 – 20, May 1990.

[3] K. K. Patel, S. M. Patel, "Internet of Things – IOT: Definition, Characteristics, Architecture, Enabling Technology, Application & Future Challenges", International Journal of Engineering Science and Computing, vol. 6, no. 5, pp. 6122 – 6131, May 2016.

[4] Q. F. Hassan, A. R. Kann, S. A. Madani ed., "*Internet of Things: Challenges, Advances and Applications*", 1st edition, Chapman and Hall/CRC, 2018.

[5] M. H. Miraz, M. Ali, P. S. Excell, R. Picking, "Internet of Nano – Things, Things and Everything: Future Growth Trends", Future Internet, pp. 1 – 28, 2018.

[6] V. Bhuvaneswari, R. Porkodi, "The Internet of Things (IoT) Applications and Communication Enabling Technology

Standards: An Overview", IEEE International Conference on Intelligent Computing Applications, pp. 324 – 329, 2014.

[7] D. S. Polaih, "Impact of Technology on Environment", International Journal of Engineering Science Invention, pp. 53 – 55, 2018.

[8] A. Sumithra, J. J. Ida, K. Karthika, S. Gavaskar, "A Smart Environmental Monitoring System Using Internet of Things", International Journal of Scientific and Applied Science, vol. 2, no. 3, pp. 261 – 265, March 2016.

[9] J. Shah, B. Mishra, "IoT Enabled Environmental Monitoring System for Smart Cities", International Conference on Internet of Things and Applications, pp. 383 – 388, 2016.

[10] A.K. Persily, "Evaluating building IAQ and ventilation with carbon dioxide", ASHRAE Transactions, vol. 103, no. 2, pp. 193-204, 1997.

[11] L. Kajtar, I. Herczeg, E. Lang, "Examination of influence of CO_2 concentration by scientific methods in the laboratory", Proceedings of healthy buildings 2003, pp. 176-181, Singapore: Stallion Press, 2003.

[12] U. Satish, M. J. Mendell, K. Shekhar, T. Hotchi, D. Sullivan, S. Streufert, W. J. Fisk, "Is CO_2 an Indoor Pollutant? Direct Effects of Low-to-Moderate CO_2 Concentrations on Human Decision-Making Performance", Environmental Health Perspectives, vol. 120, no. 12, pp. 1671-1677, 2012.

[13] G. Marques, R. Pitarma, "An Internet of Things-Based Environmental Quality Management System to Supervise the Indoor Laboratory Conditions", Applied Sciences, vol. 9, pp. 438, 2019.

[14] G. Marques, C. R. Ferreira, R. Pitarma, "Indoor Air Quality Assessment Using a CO_2 Monitoring System Based on Internet of Things", Journal of Medical Systems, vol. 43, no. 3, pp. 67, 2019.

[15] D. Zhang, W. Kong, R. Kasai, Z. Gu, Y. Minami Shiguematsu, S. Cosentino, S. Sessa, A. Takanishi, "Development of a low-cost smart home system using wireless environmental monitoring sensors for functionally independent elderly people", Proceedings of the 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp.153-158, Macau, China, 2017.

[16] F. Saeed, A. Paul, A. Rehman, W. H. Hong, H. Seo, "IoT-Based Intelligent Modeling of Smart Home Environment for Fire Prevention and Safety", Journal of Sensor and Actuator Networks, vol. 7, no. 1, article-no. 11, pp. 1-16, 2018.

[17] Vernier, CO2 Gas Sensor User Manual, https://www.vernier.com/manuals/co2-bta/

[18] WINSEN-Electronics, "Intelligent infrared CO2 module (Model: MH-Z19)", Users manual. Version 1.0, https://www.winsen-sensor.com/d/files/infrared-gas-sensor/mh-z19b-co2-ver1_0.pdf, 2016.

Modern Web-based Management System
for Administration of Scientific Conferences

Abstract – **This paper presents the web application named ConfEdit, which was created as a management system for the organization of scientific and academic conferences. The application was created using php Laravel framework and mySQL database and it is a tool that allows conference organizers to spend more time on the program and significantly less time administering. This application brings a number of improvements to nowadays applications used by the largest number of scientific-academic conferences. The application is multilingual and provides responsive design for different screen sizes.**

Keywords – **Scientific conference software, Paper review automation, Management system, Editor software.**

I. INTRODUCTION

Over the past two decades, demand for web and cloud applications has risen dramatically [1]. With the rising of demands, web technologies have been developed and they are exceeding desktop applications by functionality, performance and appearance. The inconsistency of the web browser, the lack of appropriate tools for creating a user interface, code-writer and debugging are now past. Even classical desktop applications, such as *Microsoft Office*, have their own cloud versions now, and the Serbian-American start-up company "Mainframe2" [2] has enabled to run any application software from a web browser. Today, web applications are accessible in all areas of life, as they allow access to data over the Internet from anywhere and at any time.

At the same time, with the development of the market for mobile applications and generally mobile devices, there is a need for multi-platform (cross-platform) solutions, i.e. applications that will work equally well on computer, tablet, mobile phone or any other device, which led to the creation of the rich Internet application-Rich Internet Application [3]. Technologies are moving very fast, and the big milestone has made the appearance of mobile phones and the ability to access the desired content from the smaller screens. Responsive design has become very important because it provides an uninterrupted user experience regardless of the size of the screen of the device from which it is accessed. Many applications developed earlier are more difficult to use on newer devices precisely because of the lack of responsive design.

Today there is a small number of web applications which support scientific conference management [4]. A large number of conferences for paper submission and review use *Easy Chair*

¹Slavimir Stošović is with College of Applied Technical Sciences in Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: slavimir.stosovic@vtsnis.edu.rs.
²Dušan Stefanović is with College of Applied Technical Sciences in Niš, Aleksandra Medvedeva 20, Niš 18000, Serbia, E-mail: dusan.stefanovic@vtsnis.edu.rs

[5] web application. This software is functionally very rich but it has not adequately improved the user interface comparing to similar modern web user interfaces.

In order to offer a solution that will enable to all participants in the conference a better user experience and easy use of the application from the mobile phone, the authors of this paper created the application as a conference management system.

The application was created using PHP language version 5.6.40 with Laravel framework [6][7] version 5.4 and mySQL database [8] and provides to its users modern user interface and easy using by mobile phone. PHP (Personal Home Page or PHP Hypertext Preprocessor) as a script running on a server and designed for use on the Web has become very popular in recent years. As an open source project, which means that developers are given free access to the source code, it can be used, modified, and still distributed, completely free of charge. PHP has many advantages: better performance, expandability, low cost, good support for object-oriented programming, portability, source code is available to everyone and embedded libraries to perform a large number of common Web application tasks. Because of all the above, PHP and MySQL are used to create this application.

II. USER ROLES AND PERMISSIONS

The application can be accessed via web browser at *www.demo.confedit.net* and application users are divided into the following roles:

- *Editor* - the conference editor with the highest level of privilege, has an insight into all the submitted papers, assigns reviewers, manages user accounts, performs application settings, creates accompanying documents for the conference. The account can only be created from the application.
- *Author* - participant at a conference who wants to publish the paper. The account can be handled by each author independently, and can be created by the editor from the application.
- *Reviewer* - person in charge of evaluating works. The account cannot be handled independently, but can only be created by the editor from the application.
- *Author and Reviewer* - united roles of Authors and Reviewers. Editor can also add the role of a Reviewer to the author or create an account with both privileges.
- *Printing manager* - person in charge of preparing the press release and only handles the papers that had got the status *ready for printing* after the acceptance.

After logging into the system, the user interface is adjusted to the privileges that a specific account has depending on the role. Below are detailed instructions given by user roles.

III. EDITOR SECTION

A. Management of the papers

The section *List of papers* lists all papers submitted to the conference divided by categories (topics of the conference). For each paper, the table shows paper title, authors' names, possibility to download last version of the paper, status, and quick add-on icon for detailed view of paper, Fig. 1.
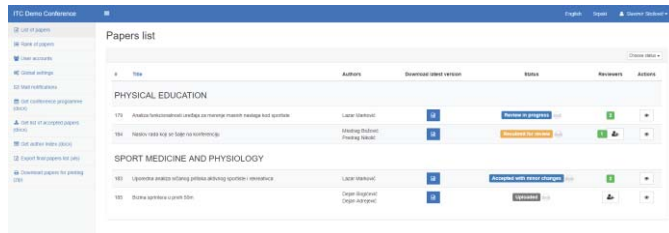

Fig. 1. Screen „Papers list"

By opening the option for a detailed overview of the paper, the Editor has all the necessary options for managing the current paper, as it is shown in Fig. 2.
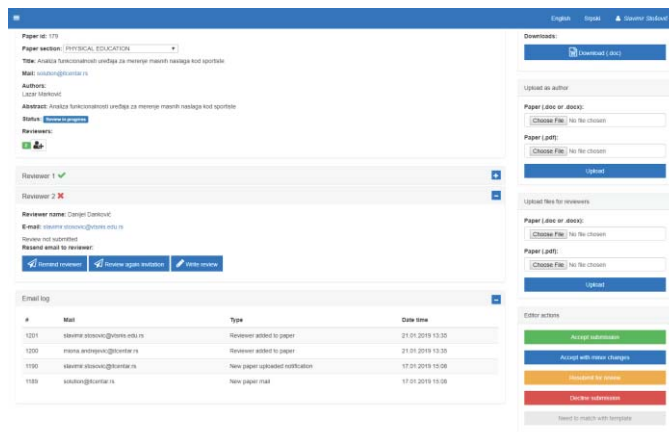
Fig. 2. Paper detailed view

In the *Paper section* Editor can change the category (section) of the work in the drop-down list. After changing the Category, Author is notified by e-mail about the change of category. In addition, the editor is available for a review. The editor has options to notify reviewer in different options:

• *Remind reviewer* - sending an e-mail reminder,

• *Review again invitation* - sending the reviewer an e-mail with a request to review the paper in the second and each subsequent review cycle,

• *Thank you Reviewer* - sending feedback to the reviewer for the completed review

In the section *Upload as author*, the editor can add a new version of the paper in .docx and .pdf format to the system, and in the section *Upload for reviewers*, the editor can add special versions of the paper in .docx and .pdf format (for example, with the author's name removed in order for the Reviewer to evaluate anonymous work).

In the section *Editor actions*, editor can make one of the following decisions:

- *Accepted as final* - paper is accepted as it is in the latest version
- *Accepted with minor changes* - paper is accepted, but it needs to be corrected in accordance to the recommendations, before it is finally accepted
- *Resubmit for review* - changes are required in accordance to recommendations, and then resubmitted for review
- *Reject* – Paper is rejected and cannot be resubmitted
- *Does not match the template* - if the paper submitted to the conference does not follow the instructions in the template, it can be returned to the author for editing. This status is active only in the first step, i.e. before assigning the paper to Reviewers.

B. Global Settings

The section for global system settings is for Editors to define timelines, work categories, and generic e-mail content, Fig. 3.

By defining the conference title and selecting the conference logo, each conference can be unified individually. Registration of new authors can be banned in a certain period, and the application can switch to the edit mode, whereby users receive a change mode message instead of the login screen. In the review form Editor can turn on or turn off the grading by selecting the appropriate option. If the rating regime is on, it is necessary to set the minimum and maximum grade as well as the step of changing the rating. In the Deadline section, the final date for the submission of papers for the conference is entered. After that date it is not possible to add paper through the system.

In the "Days for final version", "Days for correction", "Days for review" input fields, the number of days given to the author to complete the action is entered. Based on these values and the date of sending the e-mail to the author with certain requirements, the deadline is automatically calculated.

Creating text of generic e-mails is defined in a way that allows to the Editor combination of text and predefined values (values of variables) listed on the right. Adding a variable to a text is very simple; you need to place the cursor at the position where the variable is desired (for example, the access code) and click on the gray rectangle on the right side with the name of the desired variable (for example, the password).
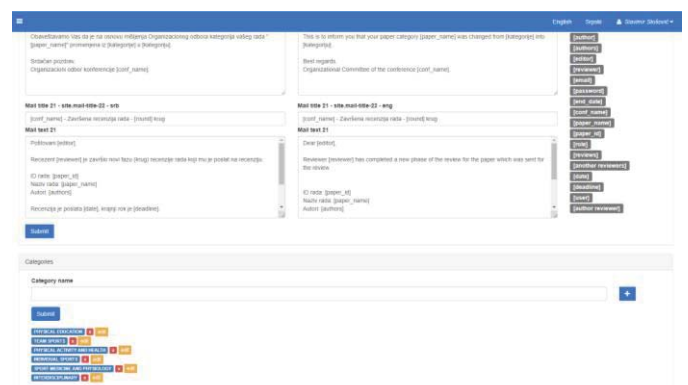

Fig. 3. Global settings screen

In the third part of Global Settings, there are options for adding / editing / deleting the conference paper categories. By entering

the category name and clicking the Submit button, a new category of paper will be added (conference area). Using the + option, multiple categories of papers can be added at the same time. The category of paper can be selected by the Author when registering for conference proceedings, but Editor can change it later.

C. Documents preparation

In order to optimize the time to prepare the necessary documentation, the Editor has the ability to create the following documents in the application:

Get a conference program - allows to automatically create a .docx file containing the names of the papers, the author's name and surname, divided by the category of paper.

Get a list of accepted papers - allows to automatically create a .docx file that lists the names of all accepted papers at the conference.

Get author index - allows to automatically create a .docx file that lists the authors' names of all accepted papers at the conference.

Export final paper list - allows to automatically create a .xls file listing the names of papers and authors' names of all papers in the conference that are in the final status and are ready to be printed in the proceedings.

Download papers for printing - enables automatic download of all papers that are in final status, past review and accepted for publication at the conference.

IV. Functionalities for Authors

The author can log in to the application by visiting the address www.demo.confedit.net using the web browser. If the Author does not create an account, he can do this by visiting the *Sign up for Author account* link and filling in all required fields (marked with star) for creating the account.

The author is obliged to enter the Name, Last Name, Title and e-mail address. The system sends an e-mail with detailed login information at the specified e-mail address. After clicking on the *Register* button, the account is created and the user is automatically logged into the system.

If any user has forgotten the access code, he can open the link to reset the access code. After entering the email address to which the account is registered, it is necessary to click the *Send Password Reset Link* button.

After logging into the application, the Author is automatically located in the *My papers* section where he can administer his conference work, as shown in Fig. 4.
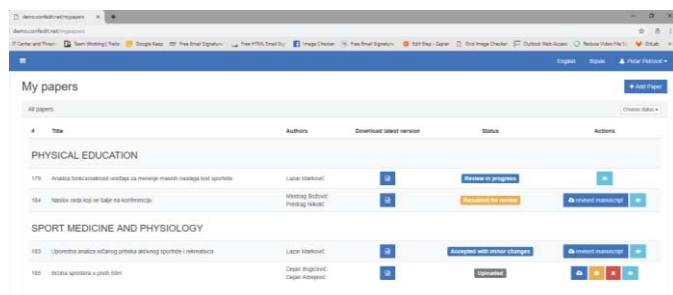
Fig. 4. Authors paper list

Each paper in the table lists the paper id, title, authors name, icons for downloading the latest version, the status of work and icons for certain activities whose availability depends on the current paper status.

By clicking on the icon for a detailed overview, user gets a screen with detailed information, reviews, status history, and paper history. By clicking the *Download template* link in the left-hand menu, the author can download the template for the preparation of the conference paper.

The Author can add new conference paper by clicking *Add* button, after which it is necessary to fill in the format shown in the following Fig. 5.

Fig. 5. Authors paper adding

It is also characteristic that the Author can send a link to download the poster and the link to the video presentation of the paper.

V. Functionalities for Reviewers

In the *Papers for Review* section, all papers assigned to the reviewer for grading by category are listed. For each paper, the

table shows the title of the paper, the download icon for the latest version, the status and the action icons. The display of the action icons depends on the current status of the paper and can be *Detailed view*, *submit review* (after assigning review paper by Editor) and *Resubmit review* (after evaluating the paper and before Editor sent the decision to the author).

By clicking on the detailed view icon, the reviewer gets an overview of detailed work information as well as a review of his previous review if he has already done so.

By clicking on the icon for reviewing a review *Submit review* or *Resubmit review*, a window will be opened with the form for evaluating the paper.
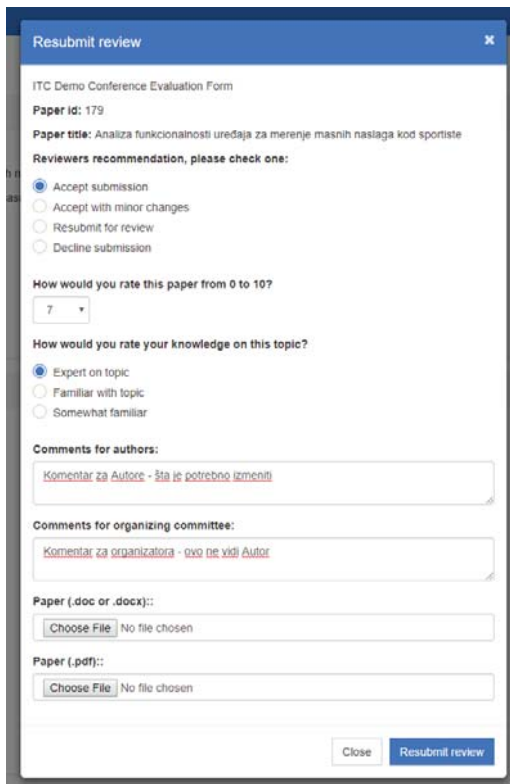
Fig. 6. Reviewers forms

In addition to the author's comments and the Organizing Committee, the Reviewer can evaluate the paper (if this option is included) and send the author an additional document in .docx and .pdf format as an explanation, as shown in the Fig. 6. The author's and the reviewer's account can be united so that the account obtains the common functionalities of the author's and reviewer's account. In doing so, the reviewer cannot review the paper where he is the author also.

VI. FUNCTIONALITIES FOR PRINTING MANAGER

In the *Papers for printing* section, all the papers that got the label *ready for printing* by category were shown in the table. The Prinitng manager's task is to take the papers that are accepted for the conference and prepare them for printing in the proceedings.

In the left menu the Printing manager has additional options: *Get a conference program, Get a list of accepted papers, Get a paper index, Export final paper list, Download papers for printing* for an easy way of preparing documentation.

CONCLUSION

Conference management system named ConfEdit and presented in this paper is a web-based application which supports scientific conference management with a lot of functionalities for Editors, Reviewer, Authors and Printing manager. Compared to other applications on the market this application provides modern user interface and responsive design for different screen sizes, sending conference call email to large number of recipients, ranking list based on different marks and few more things. Up to now, this application has supported three international conferences in Serbia, and authors plan to develop new functionalities in the future.

REFERENCES

[1] M. Kosanović, S. Stošović, "Bogate internet aplikacije – pregled stanja 4 godine kasnije", Zbornik radova, Infoteh, Jahorina, mart 2015, pp. 726 – 730, ISBN 978-99955-763-6-3

[2] https://fra.me/, visited on May the 1st

[3] S. Abolfazli, Z. Sanaei, A. Gani, F. Xia, L. T. Yang, "Rich Mobile Applications: Genesis, taxonomy, and open issues", Journal of Network and Computer Applications, Volume 40, April 2014, Pages 345-362, ISSN 1084-8045, http://dx.doi.org/10.1016/j.jnca., 2013

[4] Top 10 Abstract Management Systems for Academia: https://medium.com/@eventindustrynews/top-10-abstract-management-systems-for-academia-98c16cf17a15

[5] https://easychair.org/, visited on May the 1st

[6] N. Prokofyeva, V. Boltunova, "Analysis and Practical Application of PHP Frameworks in Development of Web Information Systems", Procedia Computer Science. 104. 2017, pp. 51-56. 10.1016/j.procs.2017.01.059.

[7] SocialCompare PHP frameworks comparison: http://socialcompare.com/en/comparison/php-frameworkscomparison.

[8] L. Welling, L. Thomson, "PHP i MySQL Razvoj aplikacija za Web", Pearson Education Inc., 2009.

Development of Machine Learning Models for Foreign Trade Volume Prediction

Ana Uzelac[1], Slađana Janković[2], Snežana Mladenović[3] and Stefan Zdravković[4]

Abstract – **There is a growing need to analyze supply chain data and find patterns that can be used to enhance them. In this paper we have presented how machine learning models can be used to analyze foreign trade dataset in order to predict the volume of the import and export of food products in the Republic of Serbia.**

Keywords – **Machine Learning, Prediction, Big Data Analytics.**

I. INTRODUCTION

Big data is a term used for massive data sets with complex structure. It refers to those datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze [1]. The four main characteristics defining Big Data are Volume, Velocity, Variety and Veracity [2]. As previously mentioned, Big Data exceeds the space of technical ability of storing, processing, managing, interpreting and visualizing of a traditional system [3].

Before the Big Data era, various data analytics technics were used to analyze data with the aim to find correlations between them. With the Big Data emergence, a great volume of data is generated every day creating a growing demand to investigate a greater amount of data in order to find useful patterns and correlations within. Big Data can be combined with analytics forming the Big Data Analytics (BDA). The term BDA can be defined as the application of advanced analytic techniques including data mining, statistical analysis, predictive analytics, etc. on big datasets as new business intelligence practice [4]. BDA has the ability to research massive amounts of data with the aim to reveal hidden patterns and secret correlations. Therefore, Big Data combined with analytics creates the ability to extract meaningful insights and turn data into information and intelligence [5]. BDA give firms competitive advantage by extracting significant value from massive amounts of data creating an imperative for business leaders in almost every industry sector: from healthcare to manufacturing.

One of the technics used in BDA is machine learning. Machine learning emerged many years before Big Data existed: in 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed" [6]. Tom Mitchell provides a more modern definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [7]. In general, any machine learning problem can be assigned to one of two broad classifications: supervised and unsupervised learning. If instances are given with known labels (the corresponding correct outputs) then the learning is called supervised, in contrast to unsupervised learning, where instances are unlabeled [8]. If we are trying to predict something that has a discrete value, that type of supervised learning is called classification. The other type of supervised learning is regression where the accuracy is measured by how close the estimate is to the actual value in the test dataset, rather than whether the predicted value is precisely right or not. Regression typically requires different algorithms than those used for classification. The aim of classification and regression is to predict something accurately. The part of the data is used for learning while the rest of the data is used for testing the result of the learning. In comparison, unsupervised learning is where we do not know the right answer ahead of time for any of the data—there is no prior basis to judge how good our result is. The goal of unsupervised learning is to find interesting and useful generalities within the data. A common form of unsupervised learning is clustering - given a collection of data, separate instances into two or more groups (clusters) based upon their similarities. Unsupervised learning has no objective measure of success, and therefore, all the data can be used as input to the algorithm. As previously seen, selecting the suitable machine learning technic depends mostly on the structure of the data and the goals of the study.

The aim of this paper is to explore possibilities to predict the volume of the foreign trade using supervised machine learning and to propose the development methodology and application of appropriate machine learning models. In the second section we have covered literature review. In the third section a methodology including development and the application of the machine learning model that predict the volume of import and export is presented. Some of the results obtained using proposed methodology on the dataset of foreign trade of food industry in the Republic of Serbia are shown in the fourth section. In the last section a conclusion about machine learning algorithms that have shown the best results in predicting the volume of the foreign trade on the available datasets is given.

II. LITERATURE REVIEW

Gartner estimates that by 2020 there will be around 26 billion devices in the supply chain. All of them generate a great amount

[1]Ana Uzelac is with the Faculty of Traffic and Transportation University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: ana.uzelac@sf.bg.ac.rs.

[2]Slađana Janković is with the Faculty of Traffic and Transportation University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: s.jankovic@sf.bg.ac.rs

[3]Snežana Mladenović is with the Faculty of Traffic and Transportation University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: snezanam@sf.bg.ac.rs

[4]Stefan Zdravković is with the Faculty of Traffic and Transportation University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: s.zdravkovic@sf.bg.ac.rs

of data every day. Therefore, there is a growing need to analyze huge amounts of data in Supply Chain Management (SCM). Scholars agree that BDA has the potential to transform the entire business process, by improving the various supply chain processes and logistics management [9]. Although the term "Big Data" is not new, there are not many applications of Big Data in the SCM field.

As machine learning is able to discover patterns in supply chain data, it has been identified ten ways how machine learning can revolutionize supply chain management: 1) improving demand forecast accuracy, 2) reducing freight costs, 3) improving SCM performance, 4) opening up many potential applications in physical inspection and maintenance of physical assets across an entire supply chain network, 5) lowering inventory and operations costs and getting quicker response times to customers, 6) forecasting demand for new products, 7) extending the life of key supply chain assets including machinery, engines, transportation and warehouse equipment, 8) improving supplier quality management and compliance, 9) improving production planning and factory scheduling accuracy, and 10) providing end-to-end visibility across many supply chains for the first time [10]. Artificial intelligence with machine learning can help the logistics industry fundamentally shift its operating model from reactive actions and forecasting to proactive operations with predictive intelligence [11]. Additionally, machine learning represents a new tool that can enable companies to better understand the impact of demand drivers such as media, promotions and new product introductions, and to then use that knowledge to significantly improve forecast quality and detail [12].

Artificial Intelligence is set to transform the foreign trade, for example; by reducing the cost of numerous processes throughout the trade lifecycle [13]. Furthermore, machine learning can help in reducing foreign exchange risk [14]. Moreover, machine learning can be used to build a meta-learning model that is able to detect the error in the foreign trade transactions [15].

Although machine learning models are recognized to be a great tool to discover patterns in large amounts of data with the aim to improve different parts of SCM, currently there are not many studies that investigate their practical applications. As foreign trade represents an important part in SCM, this investigation is of a great importance as it shows how machine learning models can be used in practice in order to predict the volume in the foreign trade.

III. Methodology

The research process consists of the following three phases: Data exploration, Data preprocessing and Predictive analytics.

During data exploration phase we examined some characteristics of the initial dataset, such as its volume, completeness, validity of data, potential relations between individual data elements, different ways raw data is organized and stored.

Data Preprocessing phase consists of standard ETL (Extract, Transform and Load) operations. During Data preprocessing we performed the following operations: data importing, data querying, data cleaning, data formatting end data exporting. As

a result of data processing, a new dataset on which it can be applied different machine learning technics is generated. Additionally, as the final step we divided the original dataset into two different datasets: testing and training datasets.

The aim of predictive analytics is to predict what will be happening or is likely to happen in the future by exploring data. It attempts to accurately predict the future events and discover the reasons. The aim of predictive analytics in this research was to predict the volume and structure of import and export of food products in the Republic of Serbia. In order to get the answer, we used different machine learning techniques. Machine learning process consists of the following steps: 1) data preprocessing, 2) model building, 3) model evaluation, 4) model testing, and 5) model deployment. Machine learning is an iterative process which is repeated until a satisfying performance is achieved.

The first step was to build a model. As we had labeled dataset, we could build only supervised machine learning model. First, we defined the goal of our model, then selected dependent variables (labels) and relevant attributes, performed necessary preprocessing of the dataset in order to prepare it to fulfill requirements of the selected algorithm. The next step was model tuning where we set hyperparameters that are specific for each type of the machine learning algorithm. The next phase was model training where we applied selected machine learning algorithm on the training dataset in order to obtain model parameters.

Since our dataset labels (Net weight [kg] and Amount [EUR]) are numeric continuous, we have chosen machine learning models based on the most popular regression algorithms: Linear Regression, k-Nearest Neighbors, Decision Tree, Support Vector Machine for Regression and Neural Network. Machine learning model results from learning algorithm applied on a training dataset.

We performed model evaluation using 10-fold cross-validation. To predict the performance of a model on a new dataset, we need to assess its performance measures on a dataset that played no part in model formation. This independent dataset is called the test dataset. We assume that both the training data and the test dataset are representative samples of the underlying problem. Comparing test vs. training performance allows us to avoid overfitting. If the model performs very well on the training data but poorly on the test data, then it is overfit. The success of numeric prediction was evaluated using various performance metrics, as they are: mean-squared error - Eq. (1), mean-absolute error - Eq. (2), root mean-squared error - Eq. (3), relative-squared error - Eq. (4), root relative-squared error - Eq. (5), relative-absolute error - Eq. (6) and correlation coefficient - Eq. (7). The total number of test instances is n; the predicted values on the test instances are p_1, p_2, …, p_n; the actual values are a_1, a_2, …, a_n; $\overline{p}$ and $\overline{a}$ are the average values of the predicted/actual values.

$$Mean-squared\ error = \frac{(p_1-a_1)^2+\cdots+(p_n-a_n)^2}{n} \quad (1)$$

$$Mean-absolute\ error = \frac{|p_1-a_1|+\cdots+|p_n-a_n|}{n} \quad (2)$$

$$Root\ mean-squared\ error = \sqrt{\frac{(p_1-a_1)^2+\cdots+(p_n-a_n)^2}{n}}\ (3)$$

$$Relative - squared\ error = \frac{(p_1-a_1)^2+\cdots+(p_n-a_n)^2}{(a_1-\bar{a})^2+\cdots+(a_n-\bar{a})^2} \quad (4)$$

$$Root\ relative - squared\ error = \sqrt{\frac{(p_1-a_1)^2+\cdots+(p_n-a_n)^2}{(a_1-\bar{a})^2+\cdots+(a_n-\bar{a})^2}} \quad (5)$$

$$Relative - absolute\ error = \frac{|p_1-a_1|+\cdots+|p_n-a_n|}{|a_1-\bar{a}|+\cdots+|a_n-\bar{a}|} \quad (6)$$

$$Correlation\ coefficient\ = \frac{S_{PA}}{\sqrt{S_P S_A}} \quad (7)$$

Where,

$$S_{PA}\ = \frac{\sum_{i=1}^{n}(p_i-\bar{p})(a_i-\bar{a})}{n-1} \quad (8)$$

$$S_P\ = \frac{\sum_{i=1}^{n}(p_i-\bar{p})^2}{n-1} \quad (9)$$

$$S_A\ = \frac{\sum_{i=1}^{n}(a_i-\bar{a})^2}{n-1} \quad (10)$$

We selected the best model comparing different criteria, such as performance (examining if the model has the best performance on the test dataset), robustness (if the model performs well across various performance metrics), consistency (if the model has one of the best cross-validated scores from the training dataset) and win condition (if it solves the original business problem).

IV. RESULTS

Previously described methodology is applied on the dataset of foreign trade of food products in the Republic of Serbia for the period from 2015 till 2017. For predictive analytics we used an open source data mining software called Weka 3.8.3. Weka is a collection of machine learning algorithms used in data mining. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

The available dataset consisted of 772517 instances, each having the following attributes: IE (Import/Export), ClearanceProcedure, RegistrationNumber, VATIN, CompanyName, CountryOfBuyer/Seller, CountryOfImport, CountryOfExport, CustomsTariff, CustomsTariffName, Year, Month, Quarter, TradeName, UnitOfMeasure, Quantity, NetWeightKG, and AmountEUR. Data preprocessing consisted of cleansing the dataset from the incomplete records, eliminating attributes that were uniquely dependent on the other attributes (such as VATIN, CompanyName, CustomsTariffName) and creating different SQL (Structured Query Language) on the dataset. Some of the queries performed only records grouping in different ways such as: Import/Export by Year and ClearanceProcedure, Import/Export by Year, Month and CountryOfImport/Export, Import/Export by Year, Month, RegistrationNumber and CountryOfImport/Export, Import/Export by Year, Month, RegistrationNumber, CountryOfImport/Export and CustomsTariff, etc. The second group of queries records were selected by different criteria, such as: Clearance Procedure, CountryOfImport/Export, Registration Number, Customs Tariff, while the third group of queries grouping and selecting

of records were performed. The results of these three categories of queries represent different datasets on which different machine learning models were built. Each of the obtained datasets were divided on the training and test dataset. Records related to the years 2015 and 2016 were used for creating training dataset, while records belonging to the year 2017 were used to make test dataset. SQL queries were created and training and test datasets were generated using MS Access 2016.

The following attributes were selected as labels (attributes which values will be predicted): Net weight [kg] and Amount [EUR]. On different datasets different machine learning models were created based on the application of the following algorithms: Linear Regression, Multilayer Perceptron (Neural Network), SMOreg (Support Vector Machine for Regression), IBk (k-Nearest Neighbors), M5P, Random Forest, Random Tree and REPTree.

Example: Training dataset: "COCA-COLA HELLENIC BOTTLING COMPANY-SERBIA" Export of Water to Montenegro by Year and Month and Custom Tariff - NetWeightT 2015-2016, for custom tariffs: 2201101100 and 2202100000; number of instances: 175; attributes: Month, CountryOfExport, CustomsTariff, SumOfNetWeightKG; Test mode: 10-fold cross-validation.

The performance of the first four different machine learning models created by using four different algorithms on this training dataset are shown in Table I.

TABLE I
PERFORMANCE MEASURES FOR THE FIRST GROUP OF PREDICTION MODELS

Machine learning algorithm	Linear Regression	Multilayer Perceptron	SMOreg	IBk
Correlation coefficient	0.7265	0.9477	0.6777	0.972
Mean absolute error	1050.97	404.73	950.78	231.16
Root mean squared error	1425.92	690.34	1609.56	489.84
Relative absolute error [%]	71.25	27.44	64.45	15.67
Root relative squared error [%]	68.56	33.19	77.39	23.55

The performance of the second four machine learning models created on the same training dataset are shown in Table II.

According to the results of the models shown in Tables I and II, the model that has the best performance was based on IBk (k-Nearest Neighbors) algorithm. Considering all performance measures this model shows the best performance. The second-place share two models with very similar performances: the

first is based on Random Forest and the second one on the Random Tree algorithm. The Fig. 1 shows relationships between actual values for the year of 2017 from test dataset, and the values predicted using the machine learning models that are selected as the best ones (IBk and Random Forest).

TABLE II
PERFORMANCE MEASURES FOR THE SECOND GROUP OF PREDICTION MODELS

Algorithm	M5P	Random Forest	Random Tree	REPTree
Correlation coefficient	0.9265	0.9691	0.9691	0.9585
Mean absolute error	448.98	256.98	248.62	321.73
Root mean squared error	791.97	514.14	516.88	592.76
Relative absolute error [%]	30.44	17.42	16.85	21.81
Root relative squared error [%]	38.08	24.72	24.85	28.50
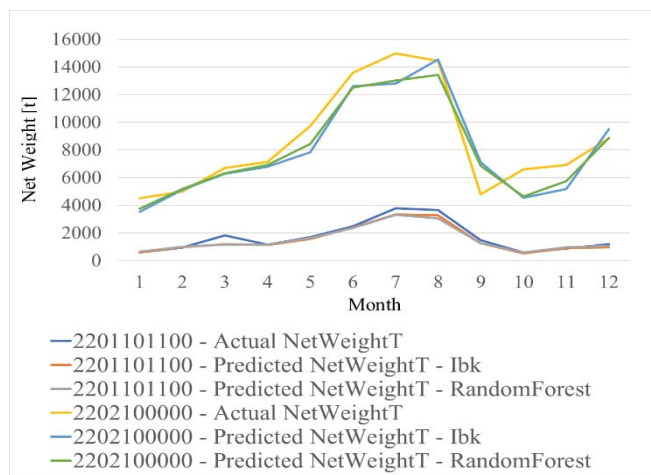
Fig. 1. Actual and predicted "Coca-Cola Hellenic Bottling Company-Serbia" export of water to Montenegro by month and custom tariff – net weight [t]

V. CONCLUSION

The most important conclusion of this research is that food foreign trade dataset can be used to build supervised machine learning models that can perform satisfying results in predicting of volume and the structure of import and export of the food products in the Republic of Serbia. Models that have shown the best performances were based on k-Nearest Neighbors, Random Forest and Random Tree algorithms. This means that the independence of the attributes of the observed dataset is better described by nonlinear machine learning algorithms and ensemble machine learning algorithms than by linear machine learning algorithms.

REFERENCES

[1] J. Manyika, M. Chui, B. Brown, and J. Bughin, "Big Data: The next Frontier for Innovation, Competition, and Productivity", McKinsey Global Institute, 2011. Available at: https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf

[2] J. P. Dijcks, "Big Data for the Enterprise", Oracle Corporation, 2011. Available at: http://resources.idgenterprise.com/original/AST-0054994_DW_US_EN_WP_BigData.pdf

[3] S., Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward". 46th Hawaii International Conference on System Sciences, pp. 995–1004, Wailea, 2013.

[4] P. Russom, "Big data analytics", TDWI Best Practices Report, Fourth Quarter, 2011.

[5] N. R. Sanders, "How to Use Big Data to Drive Your Supply Chain", California Management Review, vol. 58, no. 3, pp. 26–48, 2016.

[6] A. L., Samuel, "Some studies in machine learning using the game of checkers", IBM Journal of research and development, vol. 3, no. 3, pp. 210-229, 1959.

[7] T., Mitchell, *Machine Learning*. McGraw Hill, 1997.

[8] A. K., Jain, M. N. Murty, P. Flynn, "Data clustering: a review", ACM Comput Surveys, vol. 31, no. 3, pp. 264–323, 1999.

[9] E. Hofmann, "Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect", International Journal of Production Research, pp. 5108-5126, 2015.

[10] L. Columbus, "10 Ways Machine Learning Is Revolutionizing Supply Chain Management", 2011. Available at: https://www.forbes.com/sites/louiscolumbus/2018/06/11/10-ways-machine-learning-is-revolutionizing-supply-chain-management/#aff94263e370

[11] B. Gesing B., S. J. Peterson, D. Michelsen, "Artificial Intelligence in Logistics - A collaborative report by DHL and IBM on implications and use cases for the logistics industry", DHL Customer Solutions & Innovation, 2018. Available at: https://www.logistics.dhl/content/dam/dhl/global/core/documents/pdf/glo-artificial-intelligence-in-logistics-trend-report.pdf

[12] J. Shamir, "Machine learning: A new tool for better forecasting", CSCMP's Supply Chain, Quarter 4, 2014.

[13] Finextra, "Intelligent Machines and FX Trading", 2018. Available at: https://www.finextra.com/blogposting/15405/intelligent-machines-and-fx-trading

[14] Editorial Team, "How Machine Learning Can Help Reduce Foreign Exchange Risk", 2018. Available at: https://insidebigdata.com/2018/03/30/machine-learning-can-help-reduce-foreign-exchange-risk/

[15] M. N. Zarmehri, C. Soares, "Metalearning to choose the level of analysis in nested data: A case study on error detection in foreign trade statistics", International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland 2015.

ELECTRONICS

PWM and PFM Controlled Buck Converter designed for Wearable Electronic Devices

Tihomir Brusev[1], Georgi Kunov[2] and Elissaveta Gadjeva[3]

Abstract – **Integrated Pulse-Width Modulation (PWM) and Pulse-Frequency Modulation (PFM) controlled buck converter is proposed in this paper, designed for wearable electronic devices. The input voltage is equal to 3.6V and the average value of the output voltage is regulated to be 1.7V. The maximum efficiency η of the buck converter is 81.43%, when the load current I_{Load} is equal to 68mA. When I_{Load} is smaller than 8mA the efficiency of the PFM controlled buck converter is around 7% higher compared to the efficiency of the PWM controlled buck converter.**

Keywords – **Buck converters, Pulse-Width Modulation (PWM), Pulse-Frequency Modulation (PFM), Integrated circuits, Cadence.**

I. INTRODUCTION

Today, the number of wearable electronic devices incorporated in the Internet of Things (IoT) systems is growing very fast.
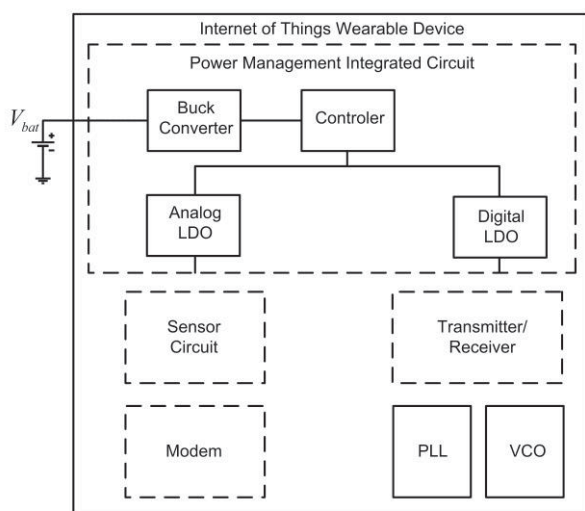
Fig. 1. Block diagram of power management integrated circuit (PMIC) in IoT wearable device [5].

Most of them are battery powered or they use energy from harvesting power sources [1], [2]. Long battery life is required

[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.

for many portable applications [3]. High efficient buck dc-dc converter integrated together with wearable electronic devices is necessary to be designed and implemented [4]. The block diagram of power management integrated circuit (PMIC) of IoT wearable device is shown in Fig. 1 [5]. The input voltage of the buck converter is equal to 3.6V, while the average output voltage is controlled to be equal to 1.7V [5]. On the other hand the output voltage of switching-mode converter is supply voltage for low drop regulators (LDO), which are shown in Fig. 1. The monolithic buck converter is necessary to occupy small silicon area and respectively to have small external filter inductor and capacitor components [6].

The PWM controlled integrated buck converter designed on CMOS 0.35 µm technology for IoT wearable device is presented in Section II. PFM control of the switching-mode regulator is proposed and presented in Section III. The efficiency η of the designed circuits, when both control techniques are used, is investigated as a function of the load current I_{Load}. The received results are compared and analyzed.

II. PWM CONTROLLED BUCK CONVERTER

The PWM controlled buck converter is designed for IoT wearable device with Cadence on CMOS 0.35 µm technology. The block circuit diagram is shown in Fig. 2.
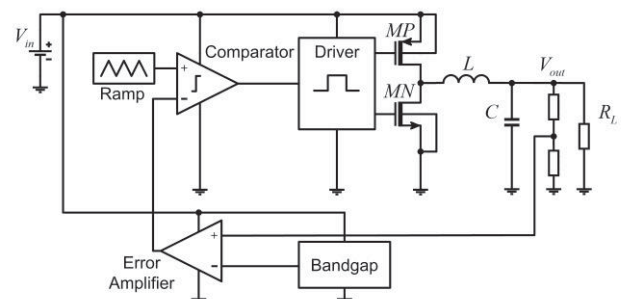
Fig. 2. Block diagram of PWM controlled buck converter.

The input voltage of the switching-mode regulator is equal to 3.6V and the average value of the output voltage is controlled to be equal to 1.7V [5]. The control system includes bandgap voltage reference, error amplifier, ramp generator and driver. The signal, which regulates the states of buck converter's power MOS transistors, is generated by comparing the voltage with repetitive waveform and error control voltage. The error signal is obtained, when difference between the actual output voltage of the whole system V_{out} and the output voltage level of bandgap reference is amplified. The frequency of the ramp generator defines the switching frequency f_s of buck converter. This frequency is constant for PWM control technique. The output signal of comparator controls the states of power buck

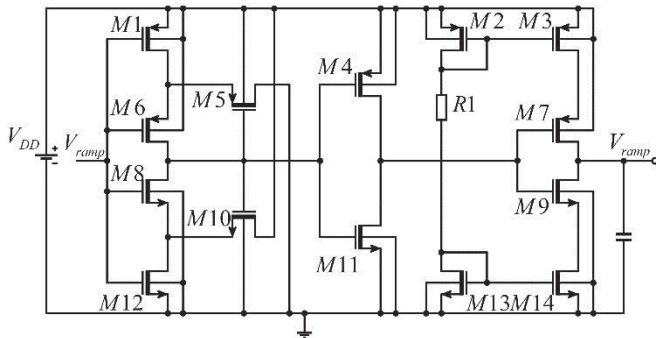converter's switches. The schematic of ramp generator is presented in Fig. 3.

Fig. 3. Ramp generator.

The output stage of the ramp generator consists of two current mirrors. They define the charging and discharging current of the output capacitor $C1$. The switching frequency f_s of the buck converter depends on the ramp capacitor's value and the current which flows through this component. The waveforms of the output signals of error amplifier, ramp generator and comparator are presented in Fig. 4.
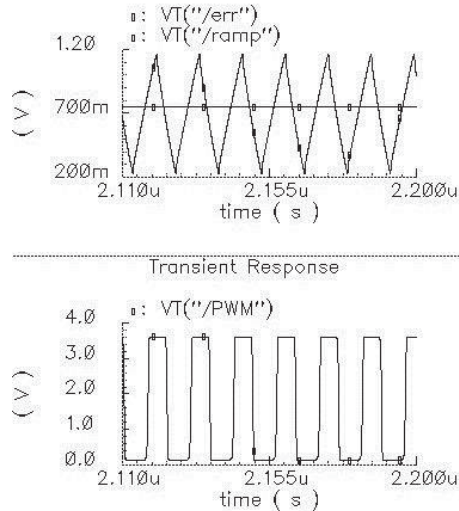

Fig. 4. The waveforms of the output signals of error amplifier, ramp generator and comparator.

The switching frequency f_s of the designed PWM controlled buck converter is equal to 80MHz.
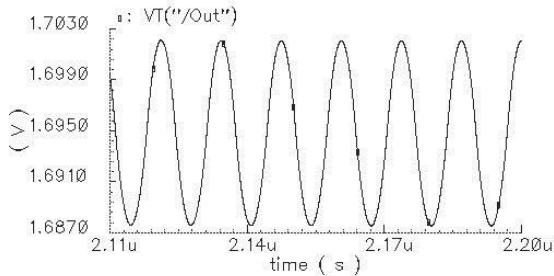
Fig. 5. The waveform of output voltage V_{out} of the designed PWM controlled buck converter.

The waveform of output voltage V_{out} of the designed PWM controlled buck converter is shown in Fig. 5. The values of the

filter inductor L and capacitor C, which are used in the low-pass filter, are equal to 250nH and 5nF respectively. The efficiency η of the buck converter as a function of the load current I_{Load} is investigated. The simulated results are presented in Table I.

TABLE I
PWM CONTROL
EFFICIENCY OF BUCK CONVERTER AS A FUNCTION OF I_{LOAD}

PWM Controlled Buck Converter	
I_{Load} [mA]	Efficiency [%]
100	75.6
80	80.64
68	81.43
50	80.81
30	79.07
20	74.69
10	62.04
8	56.9
5	45.7
2	28
1	17

The efficiency of the switching-mode regulator is calculated by formula:

$$\eta = \frac{P_{out}}{P_{in}}, \qquad (1)$$

where P_{out} is the average output power and P_{in} is the average input power of the circuit. The maximum efficiency of the PWM controlled buck converter is equal to 81.43%, when the load current is equal to 68mA. As it can be seen from the received results presented in Table I, the efficiency of PWM controlled buck converter is decreasing at light loads.
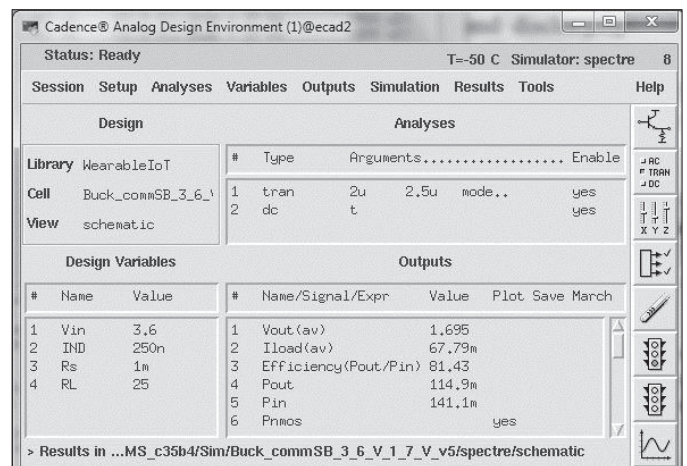
Fig. 6. Simulation results received in Cadence Virtuoso Analog Design Environment when I_{Load}=68mA.

The simulation results obtained in Cadence Virtuoso Analog Design Environment tool, when the investigated circuit indicates the highest efficiency result, are presented in Fig. 6 [7]. The output power of the converter in this particular case is equal to 115mW.

III. PFM CONTROLLED BUCK CONVERTER

The PFM controlled buck converter is designed with Cadence on CMOS 0.35 μm technology and the block diagram of the whole system is shown in Fig. 7. The input voltage of the buck converter is equal to 3.6V, while the output voltage is equal to 1.7V. The control system includes bandgap voltage reference, comparator with hysteresis, oscillator and driver stages. The power MOS transistors are regulated by oscillator with fixed 50% duty-cycle. When the actual output voltage of buck converter is higher than the desired level the control system works in sleep mode. In this case the load energy is delivered by filter capacitor C. The only stages which operate at sleep mode of the converter are bandgap and comparator with hysteresis. When the buck converter works in sleep mode of operation driver, oscillator and the power stage are disabled. Thus power losses in control system are minimized in light load conditions.
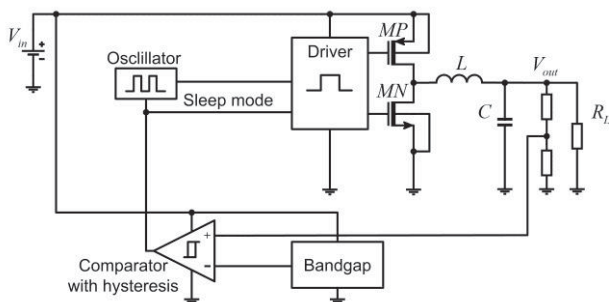
Fig. 7. Block diagram of PFM controlled buck.

The same bandgap voltage reference is used in this control as in the PWM controlled buck converter. If the output voltage V_{out} becomes smaller than the certain level, the comparator with hysteresis wakes up the whole system. This is the normal mode of operation of the switching-mode regulator. In the pictures bellow the operation of PFM controlled buck converter is presented at different load conditions.
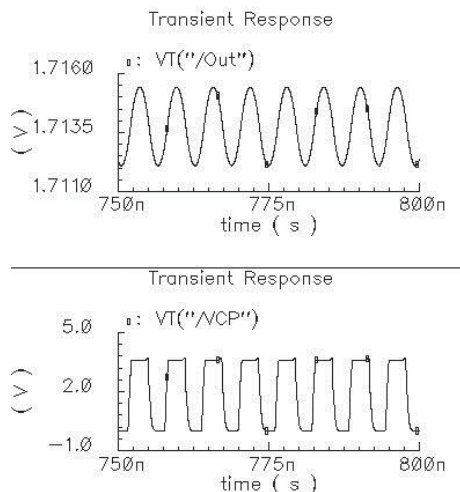
Fig. 8. The waveforms of V_{out} and control pulses of main power PMOS transistor V_{cp} when I_{Load}=8 mA.

The waveforms of output voltage V_{out} and control pulses of main power PMOS transistor V_{cp}, when the load current I_{Load} is equal to 8mA, are presented in Fig. 8.
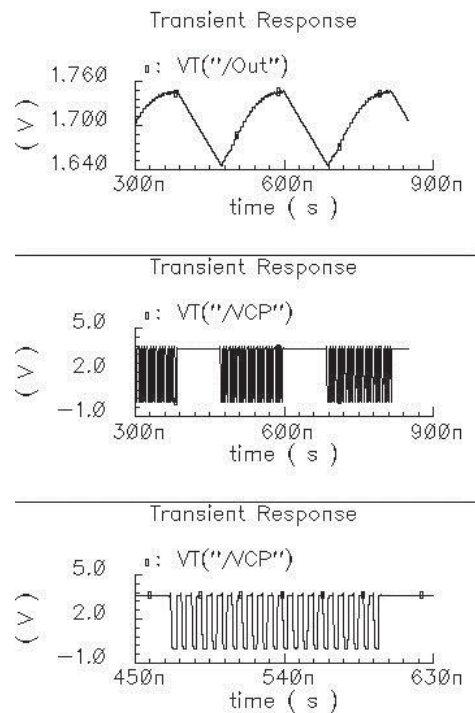
Fig. 9. The waveforms of V_{out} and V_{cp} when I_{Load}=5 mA.

The waveforms of the output voltage V_{out} and control pulses of main power PMOS transistor, when the load current I_{Load} is equal to 5mA, are presented in Fig. 9. As it can be seen from the pictures shown in Fig. 8 and Fig. 9 the designed PFM controlled buck converter works in proper manner and the average value of the output voltage V_{out} is equal to 1.7V. The waveforms shown in Fig. 9 prove that at light load the system operates longer in sleep mode.

For the oscillator with fixed 50% duty-cycle is used ring oscillator. The schematic of this ring oscillator is shown in Fig. 10.
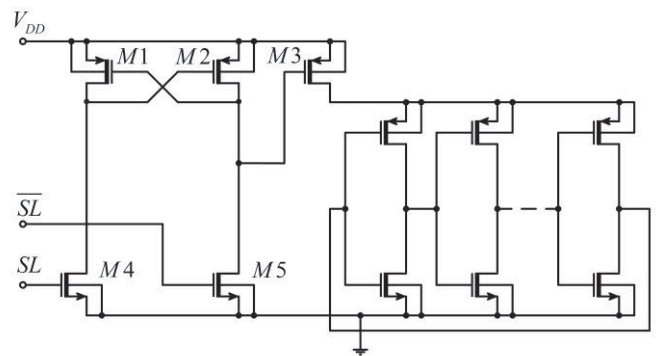
Fig. 10. Ring oscillator.

The power supply of this stage is switched-off in sleep mode in order to minimize the power losses in the PFM control system. Thus the overall efficiency of buck converter could be increased. If the output voltage of the buck converter V_{out} is higher than the desired voltage, the control signal "SL" has high voltage level, while "$\overline{SL}$" has low voltage level. In this case the transistor $M3$, which is illustrated in Fig. 10, is switched-

off. If V_{out} is smaller than the desired voltage, the control signal "$\overline{SL}$" has high voltage level and *M3* is switched-on.

The efficiency η results of the PFM controlled buck converter as a function of the load current I_{Load} are presented in Table II.

TABLE II
PFM CONTROL
EFFICIENCY OF BUCK CONVERTER AS A FUNCTION OF I_{LOAD}

PFM Controlled Buck Converter	
I_{Load} [mA]	Efficiency [%]
30	58
20	60
10	55
8	57
5	54.5
2	35
1	24

The efficiency results of PWM and PFM buck converter as a function of the load current I_{Load} are graphically presented in Fig. 11.
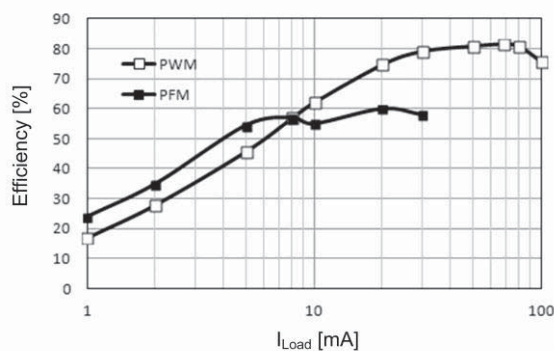
Fig. 11. Efficiency of PWM and PFM controlled buck converter as a function of I_{Load}.

The obtained results, illustrated in Fig. 11 and presented in Table I and Table II, show that the efficiency of the designed switching-mode regulator is higher at light loads if PFM control is used. If I_{Load} is higher than 8mA the PWM is more efficient control technique. When the load current is smaller than 8mA the efficiency of the PFM controlled buck converter is around 7% higher compared to the efficiency of the PWM controlled buck converter. The battery life of wearable electronic devices could be increased if PFM control for switching-mode regulator is used at light loads.

IV. CONCLUSION

Integrated PWM and PFM controlled buck converter designed for low power wearable electronic devices on CMOS 0.35 μm technology has been proposed in this paper. The input voltage is equal to 3.6V and the output voltage is regulated to be equal to 1.7V. The maximum efficiency η of the buck converter is 81.43%, when the load current is equal to 68mA. When the load current I_{Load} is smaller than 8mA, the efficiency

of the PFM controlled buck converter is around 7% higher compared to the efficiency of the PWM controlled buck converter. The PFM control technique can increase the battery life of wearable electronic devices used in IoT system, because they operate over a long period of time at light load conditions.

ACKNOWLEDGEMENT

REFERENCES

[1] C. Cheng, L. Lin, J. Lin, K. Chen, Y. Lin, J. R. Lin and T. Tsai, "A DVS-Based Burst Mode with Automatic Entrance Point Control Technique in DC-DC Boost Converter for Wearable Devices and IoT Applications", 2017 IEEE Asian Solid-State Circuits Conference (A-SSCC), Seoul, 2017, pp. 121-124.

[2] A. Paidimarri and A. P. Chandrakasan, "A Wide Dynamic Range Buck Converter With Sub-nW Quiescent Power", IEEE Journal of Solid-State Circuits, vol. 52, no. 12, pp. 3119-3131, Dec. 2017.

[3] N. Shafiee, S. Tewari, B. Calhoun and A. Shrivastava, "Infrastructure Circuits for Lifetime Improvement of Ultra-Low Power IoT Devices", IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 64, no. 9, pp. 2598-2610, Sept. 2017.

[4] A. Roy, A. Klinefelter, F. B. Yahya, X. Chen, L. P. Gonzalez-Guerrero, C. J. Lukas, D. A. Kamakshi, J. Boley, K. Craing, M. Faisal, S. Oh, N. E. Roberts, Y. Shakhsheer, A. Shrivastava, D. P. Vasudevan, D. D. Wentzloff, and B. H. Calhoun, "A 6.45 μW Self-Powered SoC with Integrated Energy-Harvesting Power Management and ULP Asymmetric Radios for Portable Biomedical Systems", IEEE Trans. Biomed. Circuits Syst., vol.9, no. 6 pp. 862-874, Dec. 2015.

[5] Y. J. Park, J. H. Park, H. J. Kim, H. Ryu, S. Y. Kim, "A Design of a 92.4% Efficiency Triple Mode Control DC–DC Buck Converter With Low Power Retention Mode and Adaptive Zero Current Detector for IoT/Wearable Applications", IEEE Transactions on Power Electronics, vol. 32, no. 9, pp. 6946-6960, Sept. 2017.

[6] M. Spasova, D. Nikolov, G. Angelov, R. Radonov, N. Hristov, "SRAM Design Based on Carbon Nanotube Field Effect Transistor's Model with Modified Parameters", 2017 40th International Spring Seminar on Electronics Technology (ISSE), pp. 1-4, 2017, Bulgaria.

[7] Virtuoso Analog Design Environment, www.cadence.com.

Experimental Module for Contactless Measurement of Electrical Current

Nikola Draganov[1] and Lyubomir Spasov[2]

Abstract – **Magnetic field sensors have been widely used in household and engineering practice. Their important advantage is their ability to measure different electrical parameters only by measurement of a magnetic field based on advanced sensors for electric current measurement by a magnetic field.**

A sensing module for measuring electrical current has been developed.

This article presents an experimental measuring module realized on the basis of a magnetosensitive integrated circuit with a Hall element. The block diagram, a general electrical circuit and results of experimental studies are described.

Keywords – **Galvanomagnetic sensors, Current sensors, Electromagnetic sensors, Hall sensors, Electric current measurement.**

I. INTRODUCTION

With the development of newer electronics technologies, in particular, sensor technology – Hall effect sensors are becoming increasingly popular in engineering environments as the primary means of measuring electrical current with high reliability, performance, sensitivity and accuracy. They are successfully applied for measurement and control of the parameters of electrical circuits. Today, Hall's components are widely used in almost all measuring device and household appliances in light and heavy industries.

Interest for sensors technologies is the contactless measurement of the parameters of electrical signals and more precisely, of the electrical parameters of power circuits. This method has many advantages the most important among them is the galvanic separation of the two circuits – the measured and measuring [1-10].

There are various sensor transducers that allow contactless measurement of circuitry parameters. They work on different physical principles – optical, inductive, magnetic, galvanomagnetic. The last are widespread in modern electronics and automation. They have some very important advantages over others, namely easy signal processing, repeatability, high sensitivity to broadband frequency signals, reliability, and more [1-6].

The purpose of the present study is to design, develop and study a real-time sensor module for electrical current measurement based on a modern magnetosensitive integrated circuit with a Hall element.

[1]Nikola Draganov is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadji Dimitar Str., Gabrovo 5300, Bulgaria, E-mail: ndrag@abv.bg.

[2]Lyubomir Spasov is with the Faculty Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadji Dimitar Str., Gabrovo 5300, Bulgaria

II. PRESENTATION

The block diagram of the realized module is shown in Figure 1.

The main unit is the magnetosensitive integrated circuit (MIC), which converts the magnetic field into an electrical signal, a voltage stabilizer (VS) providing the required DC mode and an interface circuit (IC) that forms the signal in a suitable form for follow processing.
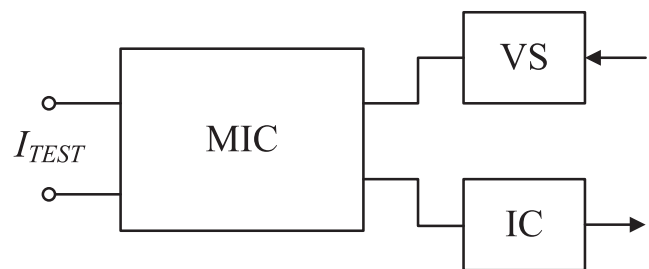

Fig. 1. Block diagram of the measureming module

The developed model is a single-PCB module with detached terminals, connecting the measured circuit to the current terminals of the magneto-sensitive IC. The last is a monolithic silicon chip with a magnetosensitive element and a processing circuit mounted on a current-carrying bus the ends of which are the IP + and IP- current measuring terminals of the integrated circuit. When a current flows through the bus, a magnetic field is formed around it which is detected by the sensor integrated in the package, and it converts the magnetic field into a proportional output voltage.

Figure 2 shows the circuit diagram of the developed module.

The stabilizer block is built with an integral voltage stabilizer of type LM7805 (IC_1). The developed device is of an experimental type. The supply current I_S that is required for it does not exceed 50mA. Of course, when installing the device into portable measurement and battery-powered systems, it is necessary to replace it with an appropriate low dropout voltage regulator.

The magnetosensitive integrated circuit (IC_2) of type ACS710 for electrical measurement with Hall element is produced by Allegro Micro Systems Company [9]. It provides a precise solution to convert the size of alternating and direct current in output voltage. Due to its small dimensions, it can be embedded in the package of most electronic devices for electrical currents measurement in current protections, instantaneous values of electrical signals, control and diagnostics of power transducers and electrical equipment.

In this type of IC, the current terminals (IP + and IP-) must be connected to the measured current loop in a given circuit. Current terminals IP+ and IP- of the integrated circuit have respectively four physical terminals internally integrated into a common copper conductor with $R_{IP+IP-}=1m\Omega$. They provide the carrying out and measurement of electrical current up to 25A. The sensor signal processing scheme allows to measure positive and negative values of electrical current, but it is necessary to observe the polarity of the IP+ and IP- terminals when entering the circuit. The selected magnetosensitive integrated circuit is equipped with Fault and Fault_EN functions. It can provide additional feedback to detect the maximum value of the measured current. This improves the functionality and reliability of the sensor device. The speed of the sensor

can be set by selecting a suitable C_F capacitor.

Figure 3 shows the dependence $F=f(C_F)$. It can clearly be seen that at higher capacitance values the operating frequency sharply decreases. Another application of C_F is to adjust the noise steadiness. The low noise steadiness is observed by low capacitance value. So it is necessary when a sensor is used in an apparatus to take into account these two parameters – noise steadiness and speed of work [2]. Figure 3 shows an exemplary selection of the capacitor value C_F limiting the transmit frequency to 45 kHz. This is necessary for further development.

The output of the MIC is at terminal 12 (U_{IOUT}). Its voltage changes in proportion to the magnitude of the measured current I_{TEST}.

The schematic diagram of the experimental setup is shown in figure 4.

A water rheostat with NaCl water solution has been used for
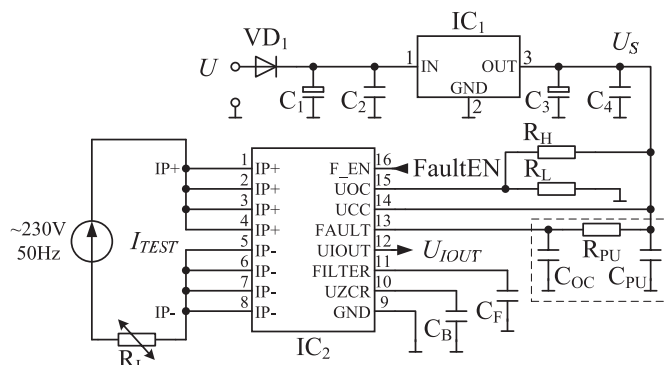
Fig. 2. The circuit diagram of the developed module

Fig. 3. Dependence of the frequency of operation on the value of the $F = f(C_F)$

Fig. 4. Schematic diagram of the experimental setup

Fig. 5. Experimental transmission characteristic $U_{IOUT} = f(I_{TEST})$

carrying out the experimental tests. So it is possible to control fluel the current I_{TEST} through measuring module and contactless ammeter.

For the purpose of collecting data for further studies the temperature of the magnetosensitive integrated circuit package T_P and that of the water rheostat T_W solution are also measured. Output data are measured by a two channel electronic oscilloscope (EO).

The obtained experimental transmission characteristics are shown in Figure 5.

They reflect the variation of the output voltage U_{IOUT} from the set current I_{TEST} passing through the sensor module. The analysis of the results shows that the transmission characteristic is linear across the current range ($I_{TEST}=0\div20A$). From the obtained results the absolute sensitivity S_A of the developed device is determined. It is given by the following formula:

$$S_A = \frac{\Delta U_{IOUT}}{\Delta I_{TEST}} \tag{1}$$

where: ΔU_{IOUT} – change of the output voltage from the magnetosensitive integrated circuit; ΔI_{TEST} – the change of the set current, or:
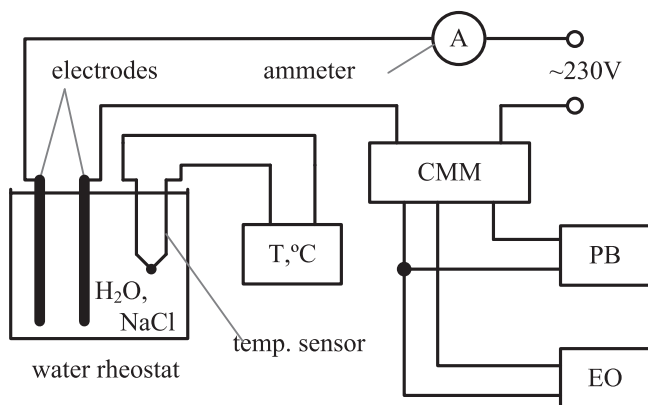
$$S_A = \frac{90,7}{4} = 22,66 mV/A \tag{2}$$

From the calculations for the S_A it is found that the absolute sensitivity of the test module was kept constant
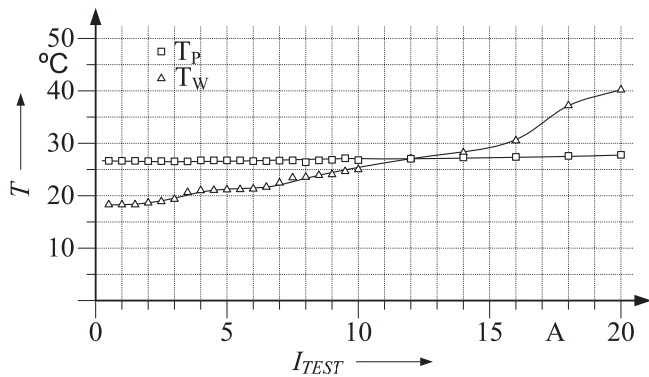

Fig. 6. Graphic dependencies $T_P = f(I_{TEST})$ and $T_W = f(I_{TEST})$

S_A=22,66mV/A. For additional information which can help the used module modeling are obtained the characteristics $T_W = f(I_{TEST})$ and $T_P=f(I_{TEST})$. They reflect respectively the water temperature of the water rheostat and that of the magnetosensitive integrated circuit package. The temperature of the chip has a minimal change over the entire operating range, due to the very low resistance of the built-in low resistive copper conductor. The dependencies are presented in the general graph of Fig. 6.

III. CONCLUSION

Sensor module with modern magnetic-sensitive integrated circuit with a Hall element of type ACS710 produced by Allegro Microsystems Company it is developed.

Magnetic-sensitive elements are becoming more widely used. The development of new sensor components on their basis extends the functional capabilities of the device in which they are built.

The developed measuring module is characterized by easy and inexpensive craftsmanship. The use of a magnetic-sensitive integrated circuit allows to measure electrical signals with a wide range of variation of the parameters, such as amplitude, frequency, rise time, and others.

Block diagram of the device consisting of a integrated sensor and power supply circuit is given.

A electrical circuit diagram of a module for measuring electrical current is synthesized, realized and studied.

With simple schematic solutions the overvoltage protection of the supply voltage poles is greater than the supply voltage ($U_S = 5 \div 15V$).

The sensor used in the module enables of bi-directional measurement of direct and alternating currents up to $I_{TMAX} = \pm 25A$.

The module possibility of feedback realizing when maximum current is reached through external pins of integrated circuit.

Low self-consumption power consumption of the module (I_S<50mA), both in measurement mode and standby mode, and the small dimensions allow him to be built integrated in various electronic devices where it is necessary to measure and monitor the magnitude of the electrical current.

The capability of the conversion feature of the module enables the output electrical signal to be directly coupled to the ADC input thereby simplifying its processing. It is determined absolute sensitivity. It is $S_A = 22,66mV/A$.

An experimental set-up was realized (Figure 4). The water rheostat configuration allows the measurement of the electrical current flowing through the sensor to be readily measured.

Experimental studies have been carried out and have been reflect the measuring module work.

Transformation characteristics $U_{IOUT} = f(I_{TEST})$ and these reflecting temperature regimes during operation $T_P= f(I_{TEST})$, $T_W= f(I_{TEST})$ at different values of the measured current $I_{TEST} = 0,1 \div 20A$ are built in sequence.

Developed and tested device is intended for installation in a variety of instrumentation and devices allowing its output signal to be used to measure and build feedbacks monitoring of various electrical parameters of the power circuits.

REFERENCES

[1] Draganov, N. Contact-Less Galvanomagnetic Ammeter for AC and DC. Journal of „Mechanical engineering and electronics". Vol. 3, Sofia, 2014, ISSN 0025-455X, стр.40-43

[2] Draganov, N. SENSORS. Principle, structure, technologies, characteristics, parameters and applications, Vol. 1. Publishing House X-Press Gabrovo, Bulgaria, 2014.

[3] Draganov, N. Galvanomagnetic contactless protection of AC induction motor. Journal of the Technical university of Gabrovo, Vol. 46, Gabrovo, Bulgaria, 2013, ISSN1310-6686, pp.84-88

[4] Draganov, N. D. Experimental Digital Three-Phase Check Electrical Energy Meter. Part 1. Proceedings of papers of international scientific conference ICEST 2014, 25-27 Juny, Niš, Serbia, 2014, ISBN 978-86-6125-109-2, pp.429-432

[5] Draganov, N., T. Draganova. Based of AMR Sensors Device for Multicannal Contactless Measurement of AC Current. Journal of the Technical university of Gabrovo, Vol. 41, Gabrovo, Bulgaria, 2011, ISSN1310-6686, pp.84-87

[6] Penchev, P. Measuring in electronics. Vasil Aprilov Print, Gabrovo, Bulgaria, 2014

[7] Lozanova, S., L. Altunyan, S. Noykov, A. Ivanov, C. Roumenin. Temperature influence of Hall sensor nonlinearity. International Scientific Conference UNITECH 2017, 17-18 Nov. 2017, Gabrovo, Bulgaria, pp.341-345

[8] Roumeni, C. Solid State Magnetic Sensors (Handbook of Sensors and Actuators). Vol. 2. Elcevier, 1994

[9] Allegro Microsystems, www.allegromicro.com data sheet of sensors 20.11.2018.

[10] Digikey, www.digikey.com/en/articles/techzone/2012/sep/the-basics-of-current-sensors 20.05.2019.

Contactless energy transmission and information data through a common inductive link

Nikolay Madzharov[1], Lyudmil Petkov[2]

*Abstract – **One of the actual direction of the contactless energy transmission systems, is the parallel two-way communication between the transmitting and receiving side using the same inductive link. This leads to the removal of the additional modules for the transmission of information and control signals and contributes to the improvement of the technical and operational parameters of the wireless power and data transfer systems. This report presents the results of the frequency multiplexing analysis. A computer simulation of the developed system was carried out. The analytical results are verified by computer and real experiments with a developed laboratory stand.***

*Keywords – **Contactless transmission, Frequency multiplexing, Two-way communication, Wireless power transfer, Control signals***

I. INTRODUCTION

In recent years, a number of companies and research teams have been working on transmitter-to-receiver communication in systems for contactless energy transmission [1,2,3,10,11,12,13,15]. The bi-directional transfer of information and control signals through the power module was shaped as the current direction. There are different methods and schematic solutions to realize the parallel transfer of energy and data [2, 4, 5, 6, 7, 9]. To reduce the mass parameters and to improve the technical parameters, systems with a common inductive connection are used. One way to transmit power up to several kW and data up to 500 kBit / s is the frequency multiplier induction method [7,9,12]. The power transmission frequency is in the range of dozens of kilohertz, and the data transfer rate is in the order of megahertz to minimize the impact of one frequency to the other. This problem is very topical and in this report are presented analytical and computer studies at different frequencies and capacities of the two modules - power and information.

II. FREQUENCY MULTIPLEXISING SCHEME

The equivalent scheme for transmitting data from the transmitter to the receiver, based on the developed frequency multiplexing method [7,9], is presented in Fig.1. The transmitter module is composed of a high frequency generator with voltage U_d and frequency ω_d. With a unit in the transmitted data the voltage is supplied to the transformer L_3. It, together with the C_R capacitor, forms a resonant circle set to frequency ω_d:

$$\frac{1}{\sqrt{L_3 . C_R}} = \frac{1}{\sqrt{L_4 . C_S}} = 2\pi f_d = \omega_d \tag{1}$$

[1]Nikolay Madzharov is with the Faculty of Electronics at Technical University of Gabrovo, 4 H. Dimitar, Gabrovo, 5300, Bulgaria, E-mail: madjarov@tugab.bg.

[2]Lyudmil Petkov is with the Faculty of Electronics at Technical University of Gabrovo, 4 H. Dimitar, Gabrovo, 5300,, Bulgaria, E-mail: ludmil_p@abv.bg.

Similarly, in the receiving side, the receiving module is comprised of the inductance L_4 and the capacitor C_S, which form a resonant circle, set to the carrier frequency ω_d, used to transmit the data. In this way, the data acquisition channel acts as a narrow line filter, which minimizes the impact of the power transmission channel.

Transformer T_3 performs multiplexing of the information signals and the frequency of energy transmitted to the load. C_{L1} and C_{L2} are parasitic capacities of the transmit and receive coils.
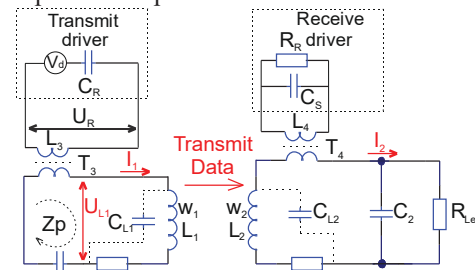

Fig. 1. An equivalent scheme, used to transfer data from the power transmitter to the receiver

Multiplexing is performed by adding the frequency used to transmit data ω_d to the energy transfer frequency. When in the data has a logical unit, on the frequency for the transfer of energy is superimposed the frequency used for data transmission ω_d. This algorithm is presented in the graphics shown of Fig. 2.
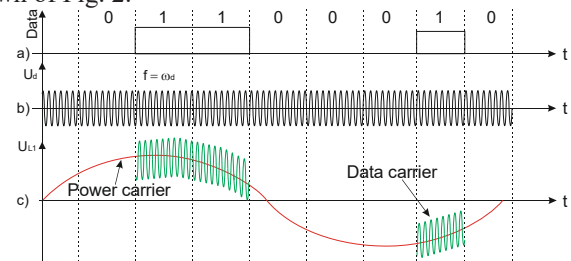

Fig.2. a) –the chart of the transmitted information signal; b) - the used data carrier frequency ω_d ; c) - the received signal from the overlaid frequency for transmitting energy and the frequency of data transmission.
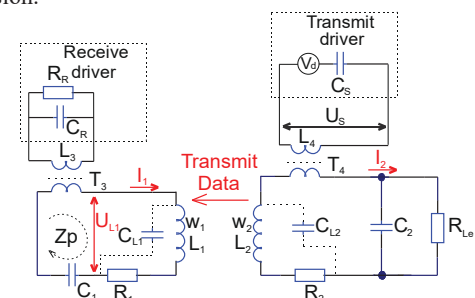

Fig.3. An equivalent scheme used to transfer data from the power transmitter to the receiver

In Fig.3 the equivalent data transfer scheme is presented in the opposite direction - from the receiving side to the transmitting side of energy. Analogous to the data transmission from the transmitter to the receiver, here the resonant circuits is composed of L_3 - C_R and L_4 - C_S and are set to the frequency ω_d used for transfer of data.
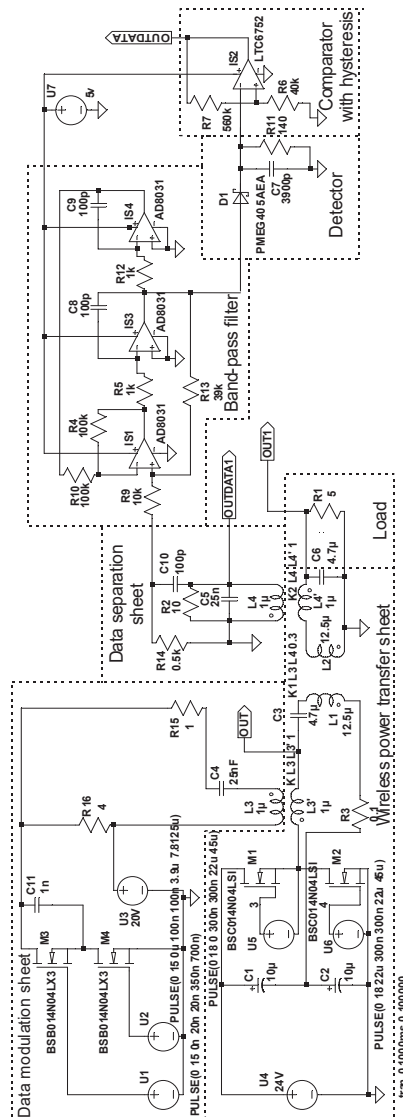

Fig.4. Scheme used for computer simulation with program LTSPICE XVII

The scheme used for computer simulation is shown in Fig. 4. The results for data transmission are presented in the following two figures. Fig. 5 shows data transmission in the direction from the power transmitter to the receiver. In order to simulate data transmission in the direction from the receiver to the transmitter it was exchanged the circuits connected to L3 and L4.

In the Fig. 6 shows the transmitting data from the receiver to the transmitter.

From the simulation and the results obtained in Fig. 5 and Fig. 6, it can be concluded that the developed frequency multiplexing method is applicable for power transmission [14,15] and bidirectional transfer of information signals at a power and control voltage ratio in both channels up to 20 times.
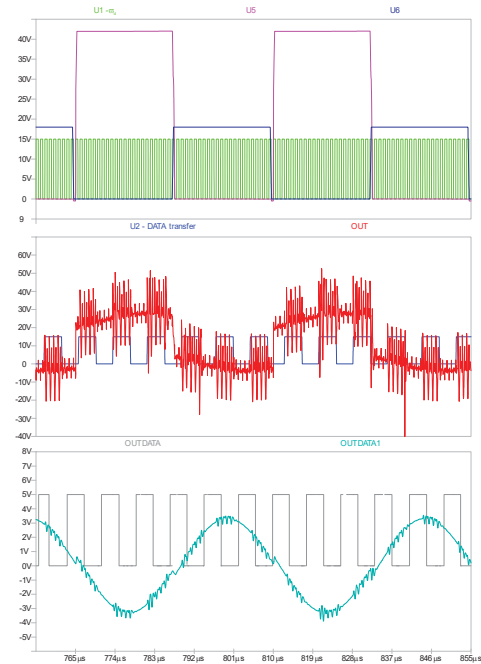
Fig.5. Computer simulation results at data transfer from the transmitter to the power receiver: U1 – data transmission voltage with frequency of ω_d; U5 and U6 - the transistor M1, M2 gate voltage control; U2 - the voltage of the transmitted data; OUT - the voltage at the resonant inverter circuit; OUTDATA1 - the voltage at the receiving data coil; OUTDATA - the voltage of the data formed in the receiving part.
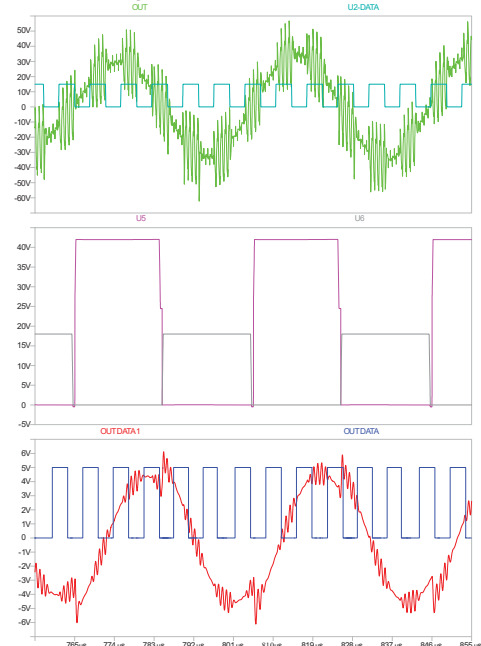
Fig.6. Computer simulation results at data transfer from the receiver to the transmitter: U2-DATA-voltage of transmitted energy from the receiver, OUT – voltage of transmitting data winding; U5 and U6 - transistor M1, M2 gate voltage control; OUTDATA1 - strip filter; OUTDATA - voltage of received data.

III. PRACTICAL EXPERIMENTS

In order to verify the developed algorithm for parallel transmission of electrical energy and information signals by frequency multiplexing, a laboratory mockup has been

developed. It is a scheme for powering and controlling the direction and speed of rotation of a DC motor. It visualizing the commands sent to the receiver and returned results and contains two principal schemes. The first one is an energy transmitter and a transmitter and data receiver Fig.7. In this circuit the PIC16F1713 processor controls the power transfer inverter and consists of U4, C5, C6, C7, C16, C18, D1, M1, M2, C17, L1, and T1. It also transfers the data to the energy receiver and receives the data from it, the scheme being implemented with Q1, Q2, Q3, Q6, Q7, Q8, VT2, VD1, VZ1, D2, U2, U3, R3, R4, R6, R7, R8, R9, R10, R12, R13, R14, R15, R19, R20, R24, R25, C8, C9, C11, C12 и C15. The 1MHz high frequency used to transmit the data is generated by the processor. To visualize sent and received data a graphical LCD display 2x16 - DD1 is used. To select the transmitted ones, the UP and DOWN keys are used, and the ENTER button is used to send the selected command. ICs IS1 and IS2 provided the necessary additional supply voltages.
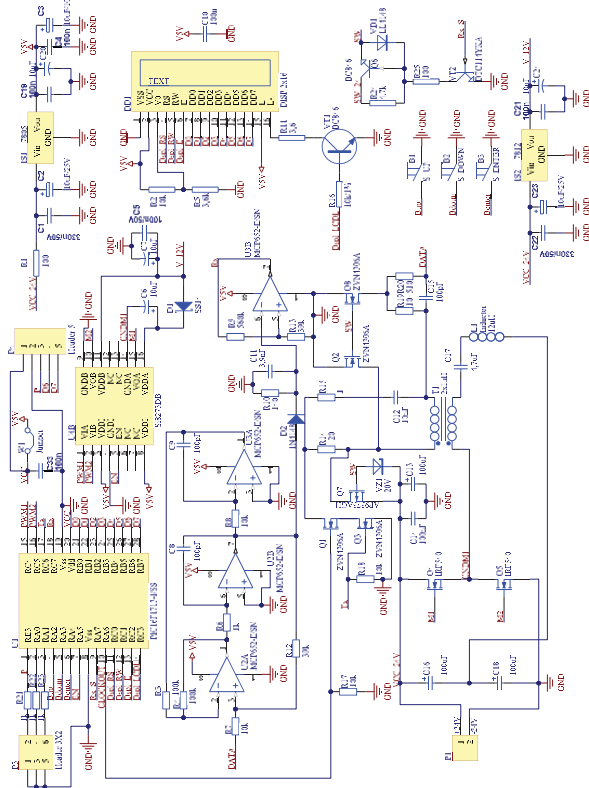
Fig.7. Scheme of the wireless power transfer and the data transmitter and receiver

The scheme of the wireless power receiver and the data receiver and transmitter is shown in Fig.8. In this scheme the PIC16F1713 processor is used for communication, direction control and speed, as well as voltage and current measurement of the DC motor. The communication scheme [8] is implemented with U2, U3, Q1, Q2, Q3, Q4, Q16, Q17, VT8, D1, VD2, VZ1, T1, R5, R6, R8, R9, R10, R11, R13, R14, R15, R17, R18, R19, R20, R52, R53, R59, R64, C12, C13, C14, C15, C16. The voltage is measured with the divider R65, R66, C34, which is connected to the rectifier D2, D8, D9, D10, and the current is measured with R61, R62, R63 and C18. Selecting the direction of rotation is controlled by VT6, VT7, Q5, Q13, Q14,

Q15, R54, R55, R56, R57, R58, R60. By pulse width modulation delivered to Q14 or Q15, the motor speed is controlled. To visualize the received commands and the measured parameters, a graphical LCD display 2x16 - DD1 is used. The buttons are used to select the desired visualization parameter. IS1 provide power to the processor.
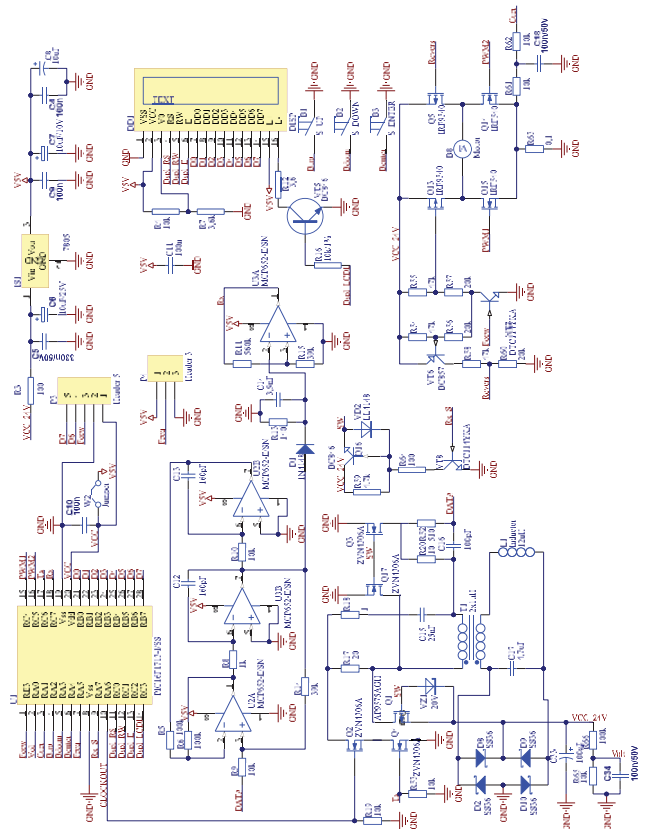

Fig.8. Scheme of the energy receiver, and the data transmitter and receiver

The experiments were performed in data transfer from the power transmitter to the receiver at a supply voltage of 24V and a distance of 2 cm between the windings of wireless module. The carrier frequency used to transmit the data is 998,780 kHz and the energy transfer rate is 20,810 kHz. The 4-channel adjustable stabilizer Twintex TP-4305 is used for power supply. To download the results, a 4-channel digital oscilloscope Tektronix TDS2014C was used. The developed model is presented in Fig. 9.
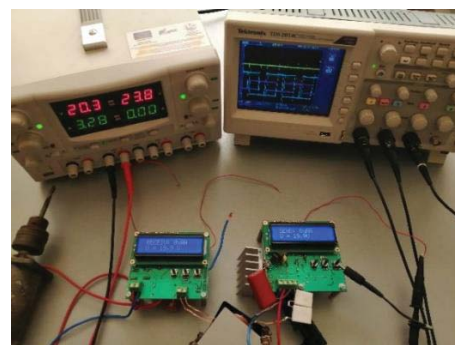

Fig. 9. Wireless mockup.

The results are presented in Fig.10 a), b), c). The Fig.10 a) presents the high frequency ω_d voltage - CH1, the gate control voltage of the MOS transistors of the inverter - M1 - CH2 and M2 - CH3. On Fig.10 b) is shown the voltage of the transmitted data - CH1 and the voltage at the point between the transformer T1 and the condenser C17 - CH2. In Fig.10 c) represents the voltage at the input of the strip filter in the energy absorption scheme - CH3 and the voltage of the output of the comparator coming to the processor - CH1.

a)

b)

c)

Fig.10. Experimental results

IV. CONCLUSION

From the applied scientific work, it can be concluded that the frequency multiplexing method can be used as a reliable tool for non-contact parallel transmission of energy and control signals. The computer and practical experiments with the developed laboratory model showed that we have a coincidence of more than 90%. The insignificant error between the results obtained from the two tests is due to differences in the components parameters used in the simulation and the limitations in the formed data parameters of the processor used.

The method can be used successfully at power up to several KW and power and control voltage ratio up to 20 times.

REFERENCES

[1] G. A. Covic and J. T. Boys, *"Inductive Power Transfer"*, *Proceedings of the IEEE,* vol. 101, no. 6, pp. 1276-1289, June 2013.

[2] Hr.Daskalova. On the choice of wireless data transmission systems. T. Atanasova (Editor), Compendium *"Modeling and Management of Information Processes"*, Sofia, Bulgaria, 2009, Sofia, 2009, ISBN: 978-954-9332-55-1, pp. 18-29.

[3] Kraev, G., Hinov, N., Arnaudov, D., Rangelov, N., Gilev, B., *"Serial ZVS DC-DC converter for supercapacitor charging"*, 19th International Symposium on Electrical Apparatus and Technologies, SIELA 2016; Bourgas; Bulgaria; DOI: 10.1109/SIELA.2016.7543018.

[4] Harakawa, Kenichi (2014), *"Wireless power transmission at rotating and sliding elements by using the capacitive coupling technology"*. 2014 ANSYS Electronic Simulation Expo October 9–10, 2014, Tokyo. ExH Corporation.

[5] U. K. Madawala, J. Stichbury, and S. Walker, *"Contactless power transfer with two-way communication"*, *in proc. IEEE* Conf. Ind. Electron., Busan, Korea, Nov. 2004, pp. 3071-3075.

[6] W. P. Choi, W. C. Ho, X. Liu, and S. Y. R. Hui, *"Bidirectional communication techniques for wireless battery charging systems & portable consumer electronics"*, *in proc. IEEE Conf. Applied Power Electron.,* Palm Springs, CA, 2010, pp. 2251-2257.

[7] Yong Zeng, Bruno Clerckx, and Rui Zhang, *"Communications and Signals Design for Wireless Power Transmission"*, in proc. IEEE Transactions on Communications (Volume: 65 , Issue: 5 , May 2017),pp. 2264 – 2290.

[8] George Clayton and Bruce Newby, *"Operational amplifiers"*, Technique, Sofia, Bulgaria, 1997.

[9] Jiande Wu, Chongwen Zhao, Zhengyu Lin, Jin Du, Yihua Hu and Xiangning He,*"Wireless Power and Data Transfer via a Common Inductive Link Using Frequency Division Multiplexing"*, IEEE Transactions on Industrial Electronics (Volume: 62 , Issue: 12 , Dec. 2015), pp.7810-7820.

[10] Vuchev, A.S., Grigorova T.G*.," Investigation of Snubber Capacitors Influence on the Operation of a Phase-Shift Controlled Series Resonant DC/DC Converter with Zero-Voltage Switching"*, 2019, 9th National Conference with International Participation, ELECTRONICA 2018-Proceedings.

[11] Vuchev, A.S., Grigorova T.G.*," Extending the Zero Voltage Switching Range of a Phase-Shift Controlled Series Resonant DC/DC Converter"*.,2018, 9th National Conference with International Participation, ELECTRONICA 2018 – Proceedings

[12] Rankovska, V., *"Using Wireless Interfaces in a Smart Home Model"*, ICEST 2016 Conference Proceedings, Ohrid, Macedonia. ISBN-10 9989-786-78-X, ISBN-13 978-9989-786-78-5, EAN 9789989786785. pp. 325-328.

[13] Arnaudov D., N. Hinov, I. Nedyalkov *"Resonant converter for charging a supercapacitor stack "*, IEEE International Power Electronics and Motion Control Conference (PEMC) 2016, Festival and Congress Centre Varna, Varna, Bulgaria, 25 - 28 September, 2016 978-1-5090-1797-3, pp. 192 -197

[14] GORANOV Goran, "*Generator, controlled by code in the structure of ALL DPLL realized in a programmable logical environment*", Machine and Electronics, Issue 1, 2013 pages 38-41, ISSN 0025-455x.

[15] GORANOV Goran, "*Digital PLL in the Transistor Inverter Structure - Development and Experimental Studies"*, E + E, 2011, N1-2,pp.29-33. ISSN 0861-4717

Design and Realization of Interleaved PFC Converter with GaN FETs and SiC Diodes

Zoran Zivanovic[1] and Vladimir Smiljakovic[2]

Abstract – **This paper presents the straightforward design of interleaved Power Factor Correction converter using the Gallium Nitride transistors and Silicon Carbide diodes. The operating principles are briefly explained, including the simplified schematics. The prototype has been built and tested through lab measurements to verify the design. Experimental results are presented to support theoretical analysis and to demonstrate the converter performance. The goal was to demonstrate the capabilities of GaN transistors and SiC diodes in the high power applications.**

Keywords – **Boost, efficiency, GaN, interleaved, PFC, SiC.**

I. INTRODUCTION

Most off-line power supplies have a front end section made by a rectification bridge and a filter capacitor. Such power supply draws a current pulse during a small fraction of the each half-cycle duration. Pulsed current waveform produce non efficient RMS currents affecting the real power available from the mains [1]. In order to utilize the full line power and to reduce line current harmonics Power Factor Correction (PFC) circuits are required. The most popular PFC converter is a boost converter (Fig. 1). Mostly because the boost converter can have continuous input current that can be modulated with average current mode control to force the input current to track changes in the line voltage.
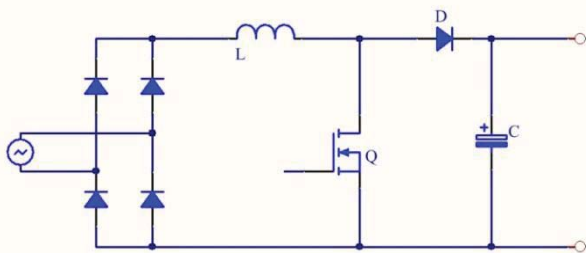
Fig. 1. PFC converter

II. INTERLEAVED PFC CONVERTER

The two phase interleaved PFC converter shown in Fig. 2 is practically made from two boost converters (phases) operating

[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.

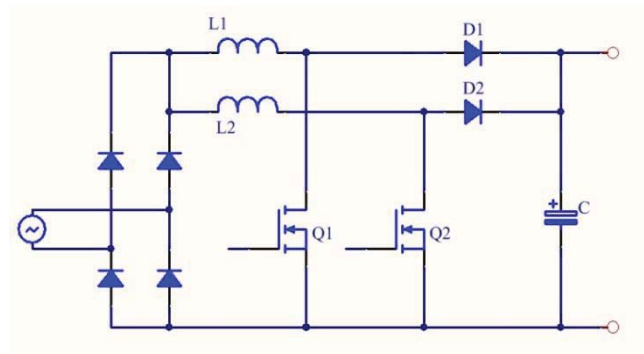at the same frequency but 180° out of phase [2,3].

Fig. 2. Interleaved PFC converter

The input current is the sum of the two inductor currents. These currents are out of phase and tend to cancel each other, reducing the input ripple current caused by the boost inductors. The output bulk capacitor current is the sum of the two diode currents reduced by the load current. As a result of interleaving its RMS value is reduced and the size of the bulk capacitor is also reduced. By splitting the current into two paths, the conduction losses of power switches Q_1 and Q_2 and inductors L_1 and L_2 are reduced. The inductors magnetic volume is lower, because the energy storage requirement of the two interleaved inductors is two times smaller relative to the single inductor. The performances are improved at the cost of additional power switches, inductors and output diodes.

III. DESIGN AND ANALYSIS

To design 1200W interleaved PFC converter we will start from the design specifications given in Table I.

TABLE I
DESIGN SPECIFICATIONS

		Min	Typ	Max	
Input voltage (AC)	V_{IN}	173	220	265	V
Output voltage	V_O		400		V
Output current	I_O		3		A
Output current limit	I_{OCL}		3.3		A
Full load efficiency	η	96			%
Switching frequency	f_{SW}		100		kHz

Converters duty cycle at the peak at low line operation is given by

$$D = \frac{V_{OUT} - V_{IN\_MIN} * \sqrt{2}}{V_{OUT}}$$

Ripple current cancelation – ratio of input current to inductor ripple current at the peak at low line operation is

$$K = \frac{\Delta I_{IN}}{\Delta I_{L1}} = \frac{1 - 2 * D}{1 - D}$$

Inductor ripple current at the peak at low line operation can be calculated as

$$\Delta I_L = \frac{P_{OUT}\sqrt{2} * 0.3}{V_{IN\_MIN} * \eta * K}$$

Minimum inductance of the boost inductors is

$$L_{1\_MIN} = L_{2\_MIN} = \frac{V_{IN\_MIN} * \sqrt{2} * D}{\Delta I_L * f_{SW}}$$

Peak current of FETs and boost diodes is

$$I_{PK} = \left(\frac{P_{OUT} * \sqrt{2}}{2V_{IN_{MIN}} * \eta} + \frac{\Delta I_L}{2}\right) * 1.2$$

RMS current of FETs is determined by

$$I_{RMS} = \frac{P_{OUT}}{2 * \eta * V_{IN\_MIN} * \sqrt{2}} * \sqrt{2 - \frac{16 * V_{IN_{MIN}} * \sqrt{2}}{3 * \pi * V_{OUT}}}$$

Diodes average current is determined by

$$I_D = \frac{P_{OUT}}{2 * V_{OUT}}$$

Calculated values of the basic parameters of the each phase are given in Table II.

TABLE II
BASIC PARAMETERS OF THE EACH PHASE

		Max	Typ	Min	
Duty cycle	D			0.37	
Inductor ripple current	I_{PP}	7.88			A
Inductor RMS current	I_{LRMS}	4.23			A
Inductance	L			120	µH
FET RMS current	I_{RMS}	2.58			A
FET peak current	I_{PK}	11.12			A
Diode average current	I_{DAVG}	1.53			A
Diode peak current	I_{DPK}	11.12			A

As a power switch we will choose between a 650V MOSFET IPW65R095C7 and GaN FET TPH3205WSB [4] whose characteristics are given in Table III. GaN FET obviously has a better figure of merit (FOM) defined as

$$FOM = (Q_{GD} + Q_{GS}) * R_{DS}$$

The reverse recovery charge causes the boost diode to look

TABLE III
POWER SWITCH CHARACTERISTICS

	Q_{GD}	Q_{GS}	C_{OSS}	R_{DS}	FOM
IPW65R095C7	15nC	12nC	33pF	84mΩ	2268
TPH3205WSB	6nC	10nC	135pF	49mΩ	784

like a short to the FET until the charge is cleared from the diode. This is a large source of switching loss in the PFC pre-converter. Those losses are dependent on the peak operating current and diode temperature. A silicon carbide diode C3D08065I [5] is chosen because it has no reverse recovery charge and therefore zero reverse recovery losses.

The output capacitor is selected based on holdup requirements and can be calculated as

$$C_{OUT} \geq \frac{2 * P_{OUT}}{f_{LINE}[(V_{OUT})^2 - (0.65 * V_{OUT})^2]}$$

IV. REALIZATION

The interleaved PFC converter has been realized on two layer FR-4 substrate, with a thickness of 1.6mm and 70µm copper with footprint 165 x 140mm. The boost inductors are wounded on two stacked iron powder toroids. Low side driver UCC27511 is used. The output capacitor is made of three 270µF electrolytic capacitors. Using isolation mains transformer and high power resistive load we have measured efficiency, power factor and also recorded the waveforms at the point of interest. Efficiency is over 98%, for 20 to 100% of the output power (Fig. 3). The power factor is presented in Fig. 4. Rectified line voltage and drain voltages are given in Fig. 5. and Fig. 6. Gate and drain voltage of the GaN FET are given in Fig. 7. Input line current with the respect to the line voltage is given in Fig. 8. Limited inrush current is recorded in Fig. 9. Output line ripple is given in Fig. 10. Output voltage startup into full load is given in Fig. 11. The converter prototype with top heatsink removed is presented in Fig.12. Finaly, the experimental setup in the lab is given in Fig. 13.
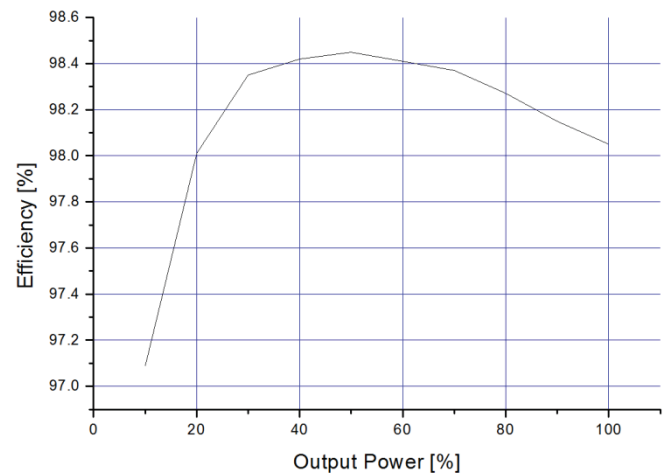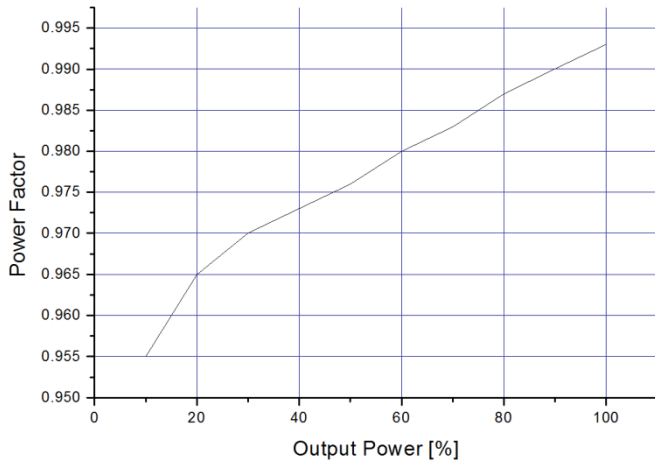

Fig. 3. Efficiency at 220V input
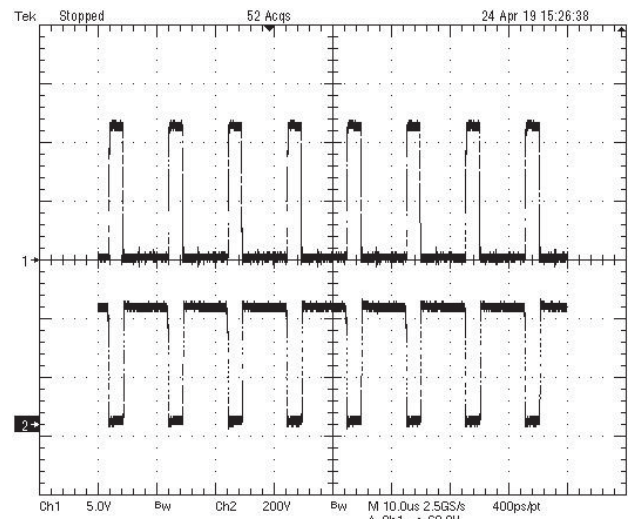
Fig. 4. Power factor
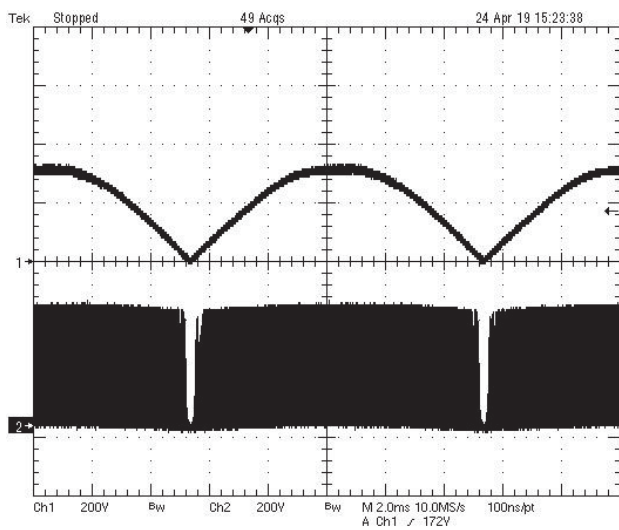

Fig. 7. Gate and drain voltage

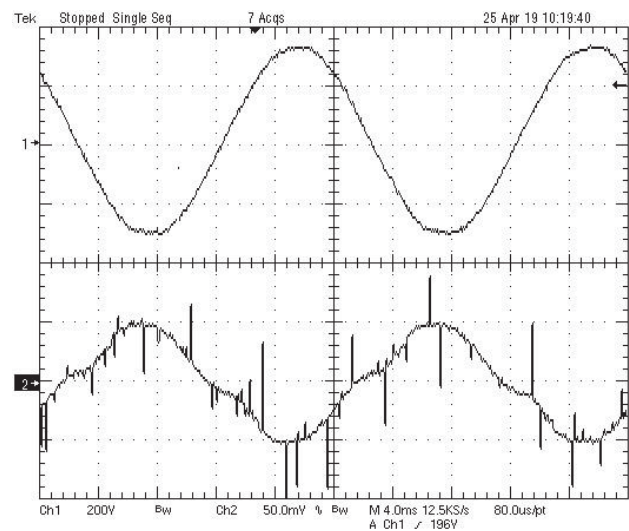Fig. 5. Rectified line voltage and drain voltage
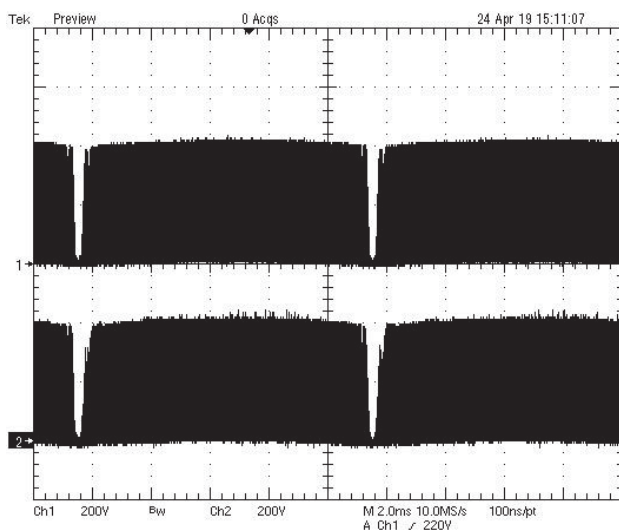

Fig. 8. Input line current
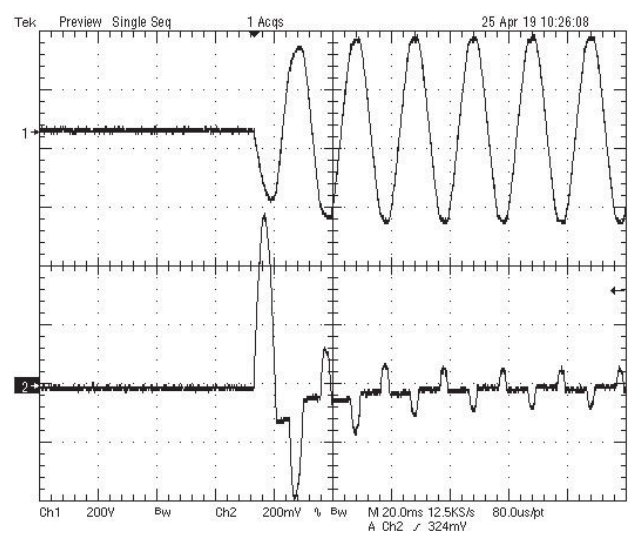

Fig. 6. Drain voltage simetry
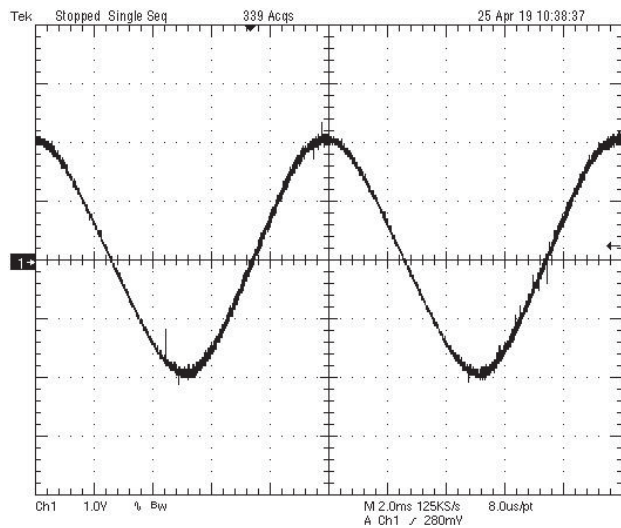

Fig. 9. Inrush current
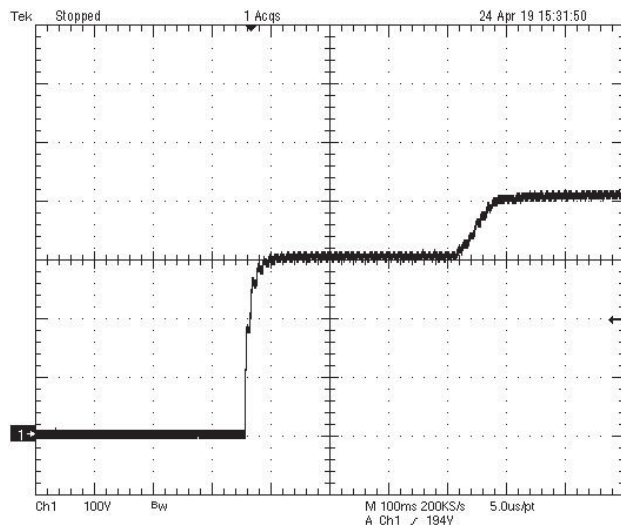
Fig. 10. Output ripple

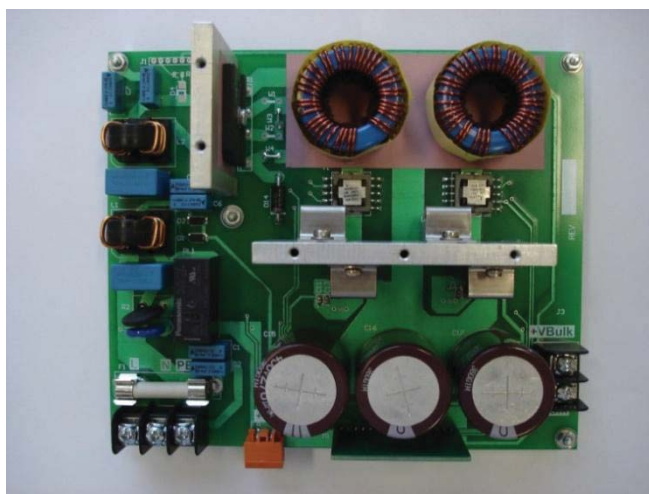Fig. 11. Output voltage start up into full load

Fig. 12. The converter prototype

Fig. 13. Experimental setup in the lab

V. Conclusions

In this paper the design and realization of a two phase interleaved PFC converter with GaN FETs and SiC diodes is presented. Calculations and experimental results are presented. The prototype was built and tested. The results verified that the efficiency can go over 98%. Further increase in efficiency is possible by switching to Bridgeless PFC converter topology [6], in order to eliminate the losses of the rectification bridge.

Acknowledgement

References

[1] STMicroelectronics, "Understanding power factor", Application note, www.st.com

[2] Klaus Raggl, Thomas Nussbaumer, Gregor Doerig, Juergen Biela and Johann W. Kolar, "Comprehensive Design and Optimization of a High-Power-Density Single-Phase boost PFC", IEEE Transactions on Industrial Electronics, July 2009, Vol. 56, No. 7, pp. 2574-2587

[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] Transphorm, "TPH3205WSB 650V GaN FET in TO-247", Data sheet, www.transphormusa.com

[5] Cree, "C3D08065I Silicon Carbide Schottky Diode", Data sheet, www.wolfspeed.com

[6] Fariborz Musavi, Wilson Eberle and William G. Dunford, and Xiaoyong Ren, "A High-Performance Single-Phase Bridgeless Interleaved PFC converter for Plug-in Hybrid Electric Vehicle Battery Chargers", IEEE Transactions on Industry Applications, July 2011, Vol. 47, No. 4, pp. 1833-1842

Microchannel Plate (MCP)

I. Zlatković[1], A. Stanković[2], R. Nikolov[3], B. Brindić[4] and D. Pantić[5]

Abstract – **Microchannel plate (MCP) is the main component in image intensifier tubes (IIT), which role is to amplify the input signal without distortion. MCP shows high efficiency in amplifying electrons, has good resolution, response time, S/N ratio, and with such characteristics are widely used in imaging and non-imaging applications. In this paper operating principle and manufacturing of MCP is depicted and explained.**

Keywords – **MCP, Image intensifier tube, generations of IIT**

I. Introduction

Microchannel plate is a wafer (thin plate) with a thickness of 0.3-0.4mm, round or rectangular in shape. MCP consists of millions of very thin, conductive glass capillaries (4 to 25 micrometers in diameter) fused together and sliced into a thin plate. Each capillary or channel works as an independent secondary-electron multiplier which together forms a two-dimensional secondary-electron multiplier. MCP shows high efficiency in amplifying electrons, more than any other type of electron multiplier. It is compact, lightweight, has good transit time due to the low channel length, high amplitude, excellent amplitude distribution. With these characteristics is used as a key component in the image intensifier tube (IIT) for night vision devices, SEM (Secondary Electron Microscope), MCP-detector, Photomultiplier tube…

Night vision devices require a very good electronic image at the output of the device. The quality of MCP is crucial for image quality and it needs to meet stringent production criteria. The main parameters for a good electronic image are gain, resistance, signal-to-noise ratio, resolution and absence of defects.

MCP application can be distinguished as imaging and non-imaging applications. The split is based on the requirements for the quality of the electronic image, i.e. more specifically for use in image intensifier tube used in night vision devices.

Image intensifier tube is an electro-optical vacuum tube that enhances the intensity of the available light, allowing use in low-light conditions, such as the night. The night scene has a very low level of light whose spectrum falls into the infrared portion of the spectrum. Therefore, photocathodes, as an input window in the image intensifier tube should be sensitive to this part of the spectrum. Usually, the so-called multialkali photocathode is used for the second generation, or GaAs photocathodes for the 3rd generation image intensifier tube.

A. Generation of the image intensifier tube

As technology progressed, so were generations of night vision devices alternated in order to meet new requirements. The following are the typical characteristics of each generation of image intensifier tube, as well as generations that come in between major generations that appeared on the market.

Generation 0- Typically use the S-1 photocathode (bialcal) with a pick sensitivity in the blue-green area of the spectrum, a typical sensitivity of 60µA/lm, using focusing by electrostatic field, as well as the acceleration of the electrons to achieve gain.

Generation 1- S-10 or S-20 photocathode (multi-alkaline) are mostly used, typical sensitivity of the photocathode is 1200-200µA/lm. Also uses focusing by electrostatic field, and acceleration of the electrons to achieve gain. Gen 0 is the first true passive image intensifier tube with amplification, resolution 25-30lp/mm, MTTF 1000hrs. It's characterized by flash-resistance, so-called blinding and image distortion between the center and the periphery. Due to insufficient amplification, 3 modules are connected to one another in order to obtain the required amplification.

Generation 2- Standard type of photocathode is multialkali S-25 (extended red), alkali Na, K, Sb, Cs. Typical sensitivity of photocathodes 240-350µA/lm with a microchannel plate as an electron multiplier. Good features at low levels of light and very slight distortion of the image. Resolution at center is 28-36lp/mm, typical SNR- 11, MTTF 2.000h.

Generation 2+- Based on the Gen 2 characteristics but with a higher sensitivity of the photocathode 300-450uµA/lm, resolution at the center 32-40lp/mm, typical SNR-13, MTTF 2,000hrs.

Super Gen- Based on Gen 2 technology but with further increase in cathode sensitivity 500-600µA/lm, resolution at center 45-54lp/mm, typical SNR- 18-21, MTTF 10.000hrs.

Generation 3- Photocathode GaAs wafer and microchannel plate with ion barrier. Excellent sensitivity to minimum levels of light due to high cathode sensitivity, 1600-2000µA/lm. Resolution at center 55-72lp/mm, typical SNR- 20-28, MTTF 15.000hrs.

Figure 1 shows the design of image intensifier tube generations.

[1]I. Zlatković is with the Department of Microelectronics, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: ivanzlatkovic90@hotmail.com.

[2]A. Stanković is with the Department of Microelectronics, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

[3]R. Nikolov is with Photon Optronics, Bulevar Svetog Cara Konstantina 80-82, 18000 Niš, Serbia.

[4]Branislav Brindić is with Sova HD, Bulevar Svetog Cara Konstantina 80-82, 18000 Niš, Serbia.

[5]D. Pantić is with the Department of Microelectronics, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.
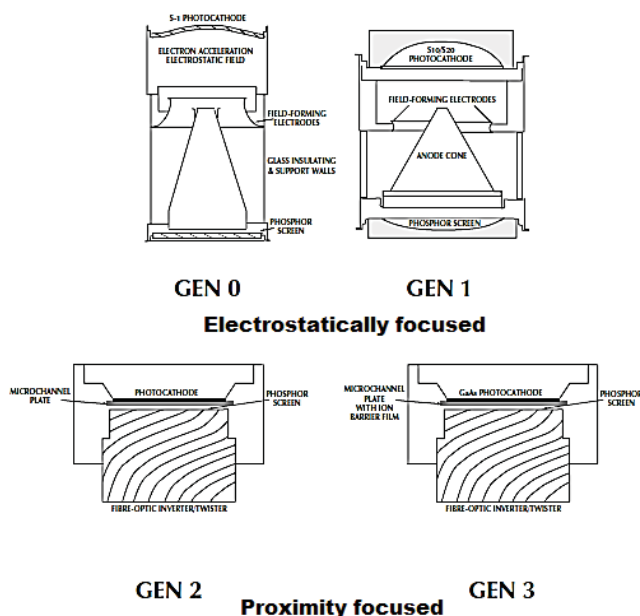
Fig. 1. Overview of image intensifier tube generations.[1]

B. Application of microchannel plates

Image intensifier tubes are widely used in the defense industry, starting with monocular, night vision gun riflescope, binoculars, pilot goggles, and various combat armored vehicles. Civilian applications have a lot of possibilities, in traffic, border and airport surveillance, firefighting, medicine, science. Microchannel plate as an electron multiplier is used in an electronic microscope. As far as non-imaging applications are concerned, a microchannel plate is used as a detector of various charged particles, gamma, x-ray radiation.

As far as imaging is concerned, it is very important that the image from the input to the output is enhanced without distortion, i.e. the introduction of additional defects. The ideal picture at the output of the image amplifier would be the same as the input image only amplified. The real picture is somewhat different, the structure itself and the production technology cause some defects, the most common are black spots, image nonuniformity, dark or light hexagonal structure that can cause the problem in the picture quality itself.

II. STRUCTURE AND OPERATION OF MICROCHANNEL PLATE

The main component of the image intensifier tube is MCP, composed of several million channels that can be 5, 6, 8, 10 or 12μm in diameter. The MCP consists of an active area where the channels are located with a semiconductor property and the ability to multiply electrons. The peripheral part is called RIM and gives the tiles mechanical strength during manipulation and MCP built into the device. In the image intensifier tube, that is, the vacuum tube from the beginning to the end, we have a strong electric field, potential difference up to 6-7kV depending on the type, so we have a straight line motion of the electron from the photocathode to the microchannel plate. A several million channels form an active area of the microchannel plate, independently acts as a mini multiplier of the electrons. Operating principle of the MCP is based on the secondary electron emission (Fig. 2a). The channels are positioned under a slight angle in relation to the

direction of the electrons in order to secure a necessary collision of electrons with the walls of the channel. For the 25-10 MCP channel angle is 12° and for 18-6 it is 4-5°. In each collision, the electron from the channel wall excites the new 3-8 electrons, which further move accelerated by the influence of a strong electric field, striking the channel wall and exciting new electrons. This is a cascade process which continues through the channels, eventually causing the avalanche of so-called secondary electrons at the end (output) of a channel. This phenomenon of multiplication of an electron under the action of an electric field is called the secondary electron emission. The input signal is increased about 10^3-10^4 times, depending on the operating mode of the MCP and applied voltage.

At the end, such amplified signal hits the phosphor screen and creates a monochromatic image at the output through the ocular.
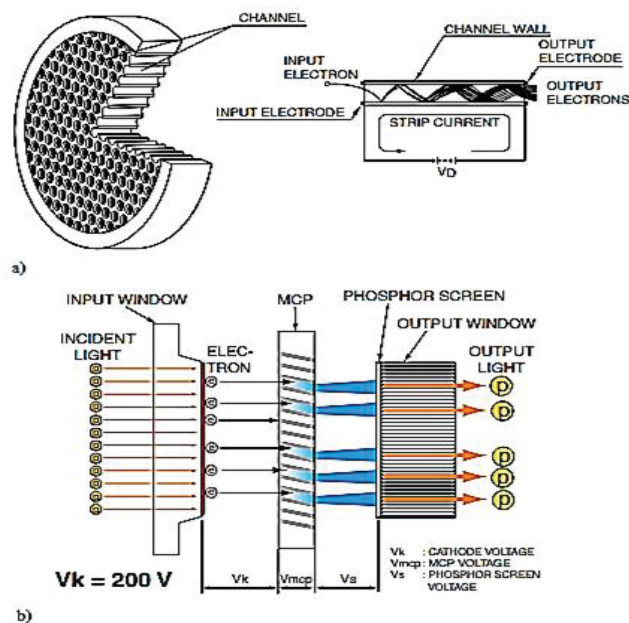

Fig. 2. The principle of operation of the microchannel plate (a) and structure of the image intensifier tube (b). [2]

In Figure 2(b), we can see the working principle of the image intensifier tube: incident light strikes photocathode, that converts photons into electrons. Once released by the photocathode, these photoelectrons are accelerated and focused by a high electric field towards the MCP. For each electron that enters MCP, approximately one thousand electrons are generated and accelerated from the output of MCP to the phosphor screen, which converts the electrons back into photons. All of these components are integrated into the final device- image intensifier tube. The very principle of image intensifier tubes is based on technological processes through which the components are individually processed, starting from the technological processes for the production of photocathodes, MCP as an amplifier and the phosphor screen used to display this image.

A. MCP Operating Background

It is well known that when silica glass with a high amount of lead oxide is treated in hydrogen at elevated temperatures, a semiconductor surface layer is formed. The surface layer of

the glass appears black due to the formed dispersion of metallic lead particles during the hydrogen reduction of PbO.[3] On the surface of a cross-section of the channel wall there is a secondary emission layer mostly composed of silicon dioxide and rich in alkali metals, that easily releases electrons. The thickness of the emission layer formed in the process of hydrogen reduction is 10-20 nm. A tiny emission layer formed at the channels wall surface determines the MCP gain, i.e., how many electrons will be emitted from that layer, and a lifetime of MCP.
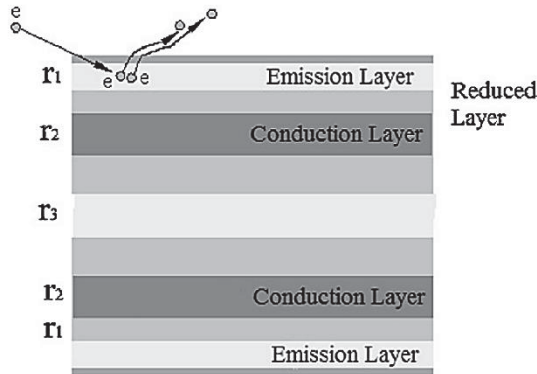

Fig. 3. Channel wall structure. [4]

The next layer is a conductive (resistant) layer of 50-100nm thickness. This layer is placed deeper in the channel and it is rich in lead.[3] The conductive layer determines another electrical characteristic of MCP and that one is its electrical resistance. The composition of the glass and the conditions of the reduction process (time and temperature) affect the resistance of MCP. This resistant layer contains clusters (small groups and aggregates) of lead that transfers electrons. In the first phase of the hydrogen reduction process, small lead grains are formed and after that, they group up. At temperatures of 450-500°C, grains begin to the grouping into clusters, and the principle of conducting the current is the principle of skipping the electron from the clusters to next clusters of lead. The function of the conductive layer is to donate the electrons to the emission layer for the secondary emission by the support of the strip current. When the electron leaves the lattice of potassium silicate, a hollow left in the lattice is occupied with an electron from a conductive, lead layer and that makes the MCP current ("strip current"). Usually, the saturation of MCP starts when the output current is up to 10% of the strip current.

$$R = Umcp/Istrip \qquad (1)$$

III. Manufacturing of MCP

The production of microchannel plates (MCPs) belongs to a group of multidisciplinary technologies. Production is organized in the high special area, in the so-called clean rooms of the ISO6 class of cleanliness according to ISO 14644-1 or Class 1000 according to the FED US 209E standard and very strict operating standards of the operator. The necessary prerequisites are the control of the microclimate (temperature and humidity), which significantly affect the quality of the process. However, technology depends mostly on raw material, special semiconductor glasses: lead glass- cladding (rich in alkalies) and boron silicate-core.

Here is briefly presented technological process in the manufacturing of MCP. Microchannel plate begins as a glass tube (cladding) fitted with a solid rod (core) - together they are called single fiber preform and drown via fiber optic techniques to form a single fiber. This kind of technology is used for producing optic fibers (for communications), etc. A number of these fibers are then stacked in a hexagonal array, so-called bundle- multy-fiber preform; the entire assembly is then drowned again to form multy-fibers. In the same time, the hex rode (RIM) is drawn in the same way as single fibers, they serve as a support for MCP. The multy-fibers and hex-rim fibers are then stacked together (multy-fibers are in the center of MCP forming active area surrounded with rim hex-fibers) and fused at high temperature to form boul. As it's shown in the picture below.
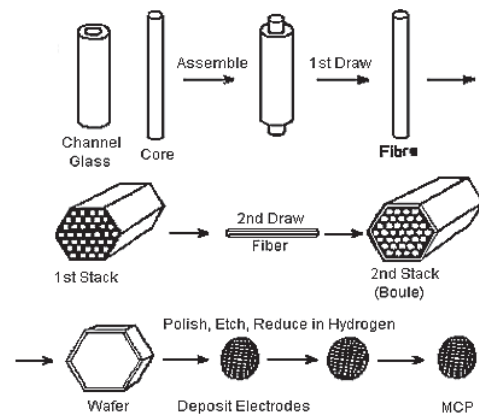

Fig. 4. Fabrication of MCP. [5]

The boul is sliced to wafer with required bias angle, edged to size and then ground and polished. And this is the first part of technology. After that, the second part of technology includes next steps. The individual wafers are chemically processed to remove the solid core material, leaving the structure of millions of tiny holes. Then, through reduction at elevated temperatures, conductive and secondary emission properties are given to the glass wafers. Finally, a thin metal electrode (usually chromium, Inconel or Nichrome) is vacuum deposited on both input and output surfaces of the wafer to electrically connect all the channels in parallel. Finished products are then tested under certain criteria in the tester.

IV. Conclusion

In terms of further improvement and new trends, the following directions are possible in order to improve the characteristics of MCP:

1. Upgrading the traditional MCP manufacturing technique by:

1.1 Miniaturization of the existing model in terms of reducing the diameter of the channel (e.g. 3μm), in order to get a better resolution.

1.2 Introduction of additional physical and chemical treatment of plates in order to reduce the impact of ion feedback (applying ion barrier film- IBF).

1.3 Optimizing the chemical composition of the glass to achieve the above-mentioned results.

2. A new approach of MCP manufacturing based on Atomic Layer Deposition (ALD), which combines glass capillary array substrates with thin film deposition techniques that provide the necessary resistive and secondary emissive properties of an MCP. [6,7,8]

In the table below we can see the advantages and disadvantages of these two technologies.

TABEL I
ADVANTAGES AND DISADVANTAGES OF CONVENTIONAL MCP TECH-
NOLOGY OVER ALD TECHNOLOGY.

Technology	Advantages	Disadvantages
Conventional	-Well-known technology -Row material- known. -Lifetime. -The equipment is inexpensive, relatively.	-Structure defects. -Defects in the electronic image. -Limit by IBF.
ALD	-Simpler procedure. -Possibility of independent choosing glass, resistive and emissive properties-layer control. -Mechanically robust glass and may be selected to have a low content of radioactive material for reduction of background activity. -Better image quality.	-Expensive equipment. -Not proven technology.

REFERENCES

[1] Ientilucci, Emmett. (2019). Synthetic Simulation and Modeling of Image Intensified CCDs (IICCD).

[2] https://www.hamamatsu.com/resources/pdf/etd/II_TII0007E.pdf

[3] A.M. Then, C.G. Pantano, "Formation and behavior of surface layers on electron emission glasses", Journal of Non-Crystalline Solids, vol. 120, no. 1-3, pp. 178-187, Apr. 1990.

[4] J. Pan, J. Lv, R. Zhang, Q. Yu, H. Zhao, S. Sun, X. Hanb, "Formation mechanism of 'memory' Phenomenon of micro-channel plate image Intensifier", Optik International Journal for Light and Electron Optics, vol. 124, no. 20, pp. 4509-4513, Oct. 2013.

[5] F. Sauli (2010). Photon detectors [PDF Presetation]. Retrieved from:https://indico.cern.ch/event/80223/sessions/112102/attach ments/1061990/1514408/4_PHOTON_DETECTORS_CHIPP1 0.pdf

[6] A. Mane, Q. Peng, M. J. Wetstein, R. G. Wagner, H. J. Frisch, O. H.W. Siegmund, M. Minot, B. W. Adams, M. Chollet, J. W. Elam, „An Atomic Layer Deposition Method to Fabricate Eco-nomical and Robust Large Area Microchannel Plates for Photo-detectors", Physics Procedia, vol. 37, pp. 722-732, 2011.

[7] M. A. Popecki, B. Adams, C. A. Craven, T. Cremer, M. R. Fo-ley, A. Lyashenko, A. O'Mahony, M. J. Minot, M. Aviles, J. L. Bond, M. E. Stochaj, W. Worstell, J. W. Elam, A. U. Mane, O. H. W. Siegmund, C. Ertley, L. M. Kistler, M. S. Granoff, "Mi-crochannel plate fabrication using glass capillary arrays with Atomic Layer Deposition films for resistance and gain", JGR Space Physics vol. 121, no.8, pp. 7449-7460, Avg. 2016.

[8] C. Ertley, O. Siegmund, T. Cremer, C. Craven, M. Minot, J. Elam, and A. Mane, "Performance studies of atomic layer de-posited microchannel plate electron multipliers", Nuclear In-struments and Methods in Physics Research Section A: Acceler-ators, Spectrometers, Detectors and Associated Equipment, vol. 912, pp. 75-77, Dec. 2017.

ENERGY SYSTEMS AND EFFICIENCY

Power Network Reliability Estimation Using Fuzzy Set Theory

Natasha Dimishkovska[1] and Atanas Iliev[2]

Abstract —**The use of fuzzy logic for network reliability estimation is justified by its simplicity which enables its use in complex systems and creation of models that are easily applied. In this paper an upgraded method for power network reliability estimation using fuzzy logic considering the daily power load is presented. Also, a case study of two parallel transformers is reviewed. The network reliability is calculated as a complex system consisting of substation and network components each with different unavailability.**

Keywords: **Network reliability, fuzzy logic, switchgear, uncertainty**

I. Introduction

A strong power network is defined by its reliability and quality of energy supplied. Building a reliable and secure network is of technical and economic importance for the power distribution companies. Switchgear is part of the power system that consists of electrical equipment for protection, control and transmission of electrical energy. This equipment is directly linked to the electrical energy supply.

Network reliability is evaluated by its ability to maintain continuous service to the customers [1]. Each interruption in supply or fault decreases the reliability of the network. Since there are no absolutely reliable networks in practice, the problem arises as how to measure and quantify the reliability and to determine the fault as fast as possible.

The equipment in switchgears is usually from different manufacturers, which leads to big insecurity and uncertainty during the component reliability estimation process [2]. Therefore, finding a proper method for calculating the unavailability is a big challenge. Fuzzy logic is an approach of understanding the world based on degrees of truth. According to fuzzy logic, the world is not only "true or false", as the modern technology works, but there is something in between. Fuzzy logic works much likely as the human brain works.

From the presented, it can be concluded that the fuzzy logic is an appropriate mathematical basis for network reliability evaluation. The mathematical expressions are simple to use and can be easily applied in the modern technology.

In the following will be explained the theory of fuzzy logic and fuzzy sets. Also an approach for calculating substations' reliability based on a daily power load data presented as a triangle fuzzy number will be demonstrated.

[1] Natasha Dimishkovska and [2]Atanas Iliev are with the Faculty of Electrical Engineering and Information Technologies, University of Skopje, Rugjer Boshkovikj, Skopje 1000, North Macedonia
E-mails: [1]dimishkovska.n@gmail.com, [2]ailiev@feit.ukim.edu.mk

The case study of two parallel transformers in 400 kV and 110 kV network will be reviewed, considering the unavailability of the components in the switchgears.

II. Fuzzy Set Theory

The fuzzy set theory was based in 1965 by Lotfi A. Zadeh. A fuzzy set is a mathematical tool for determining whether the statement is part of the set or not. In traditional binary logic, for instance, a statement can be true or false (1 or 0) and nothing in between. In set theory, an element can either belong to a set or not, or in optimization a solution can be feasible or not. It is determined by the degree of membership.

A fuzzy set is a set which has no crisp, clearly defined boundary. It helps coping with unclear boundaries, and it is considered to be an alternative way to deal with uncertainties. Fuzzy set is a set of information which consists of membership functions, and ordered pairs. Its elements only have partial degree of membership, using numbers from 0 to 1.

a. Membership functions

A membership function for a fuzzy set **A** on the universe of discourse **U** is defined as μ_A: **X** → [0, 1], where each element of **U** is mapped to a value between 0 and 1. This value, called membership value or degree of membership, quantifies the grade of membership of the element in **U** to the fuzzy set **A** [3].

If **U** is a collection of objects denoted generically by x, then e fuzzy set **A** in **U** is a set of ordered pairs:

$$A = \{(x, \mu_A(x))| \quad x \in U\} \tag{1}$$

b. Real numbers presented as fuzzy numbers

A fuzzy number M is a convex normalized fuzzy set M of the real line $\mathbb{R}$ such that

1. It exists exactly one $x_0 \in \mathbb{R}$ with $\mu_M(x_0) = 1$ (x_0 is called the mean value of M).
2. $\mu_M(x)$ is piecewise continuous.

Fuzzy numbers basic mathematical operations include the standard algebraic operations addition (+), subtraction (-), multiplication (·), division (:), except that they have different form $\oplus$, $\ominus$, $\otimes$, $\oslash$ [3].

Real numbers can be presented as a fuzzy number of LR-type, meaning that the numbers left and right from the presented number are determined with a certain expression,

and the presented number is in that range. A fuzzy number A is of LR-type if there are reference functions L (for left), R (for right), scalars $\alpha > 0, \beta > 0$ (left and right spreads) and m, the mean value of A. The number is defined as:

$$\mu_A(x) \begin{cases} L\left(\frac{m-x}{\alpha}\right) \ for \ x \leq m \\ R\left(\frac{x-m}{\beta}\right) \ for \ x \geq m \end{cases} \qquad (2)$$

c.   Uncertainty modelling

Uncertainty is a feature of the real system caused by unpredicted events, such as lack of information, complexity of the system and human factor [3]. For instance, if a fault occurs in switchgear or in some point of the network, there is electricity not supplied, which leads to decrease of the reliability. It is unable to predict when and where exactly the fault will occur, but it can be taken in consideration that it will occur in some point of the time. Fuzzy logic provides analysis of uncertain events even if the probability is in a certain interval.

For instance, if the variable value is in the following range $[a_1; a_2]$, the membership function of the variable x is defined as:

$$\mu_A(x) = \begin{cases} 1 \ for \ a_1 \leq x \leq a_2 \\ 0 \ for \ x < a_1 \ \wedge x > a_2 \end{cases} \qquad (3)$$

Generally the fuzzy interval is represented with two end points a_1 and a_3, and a peak point a_2. If there is an α- cut included (in this case it is a safety level), then the fuzzy interval is presented as in eq. (9).

$$A = [a_1; a_2; a_3] \qquad (4)$$

$$A_\alpha = [a_1^{(\alpha)}; a_2^{(\alpha)}; a_3^{(\alpha)}] \qquad (5)$$

The mathematical operations applied for fuzzy numbers can also be applied for fuzzy intervals.

Graphically, the presentation of fuzzy intervals is shown on Fig 1.
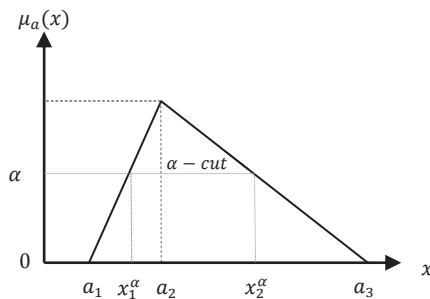

Fig 1 Triangular fuzzy number

II.   NETWORK RELIABILITY ESTIMATION USING FUZZY LOGIC

Substations are considered to be one of the most reliable parts of the power system, although they consist of many other components. The components in the substations can be connected in serial, in parallel or combined manner.

If K is the number of components connected in a series, as shown on Fig 2, and if $n_i(x, \mu_i(x))$ is the unavailability, then the equivalent unavailability of a serial union of components for a given α- cut (safety level) is defined as [2]:

$$n_e^{(\alpha)} = \sum_{i=1}^{K} n_i^{(\alpha)} - \sum_{i=1}^{K-1}\left(n_i^{(\alpha)} \cdot \sum_{j=i+1}^{K} n_j^{(\alpha)}\right) + \cdots \cong \sum_{i=1}^{K} n_i^{(\alpha)} \qquad (6)$$

If the unavailability is given with an interval fuzzy set $n_i = [n_{1i}; n_{2i}; n_{3i}]$, then the equivalent unavailability will also be a fuzzy set [2].

$$n_e = [n_{e1}; n_{e2}; n_{e3}] = [\sum_{i=1}^{K} n_{1i}; \sum_{i=1}^{K} n_{2i}; \sum_{i=1}^{K} n_{3i}] \qquad (7)$$
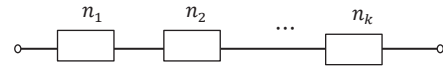

Fig 2 Serial connection of the components

In case when the components are connected in parallel, as shown on Fig 3, then the equivalent unavailability of the system is defined as:

$$n_e^{(\alpha)} = \prod_{i=1}^{K} n_i^{(\alpha)} \qquad (8)$$

And when the unavailability of the components is given as a fuzzy set interval, then the equivalent unavailability is not a fuzzy number, but a fuzzy polynomial.

$$n_e^{(\alpha)} = [n_{1e}^{(\alpha)}, n_{2e}^{(\alpha)}] = [\prod_{i=1}^{K}[n_{1i} + (n_{2i} - n_{1i})\alpha]; \prod_{i=1}^{K}[n_{3i} + (n_{2i} - n_{3i})\alpha]] \qquad (9)$$
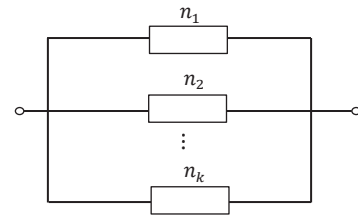

Fig 3 Parallel connection of the components

If the components in the system are connected both in serial and in parallel connection, then the equivalent unavailability is calculated with the combination of the equations (6-9).

III.   THE PROPOSED APPROACH

Using triangular fuzzy numbers and membership functions to represent the daily power load provides simpler energy loss analysis. In [4] a method for reliability evaluation of composite power systems is presented, using load flow results in different conditions. Based on that, first the

maximum, medium and minimum load points are determined (Fig 4). The load membership functions' borders are defined as in Eq. (10) and Fig 5.
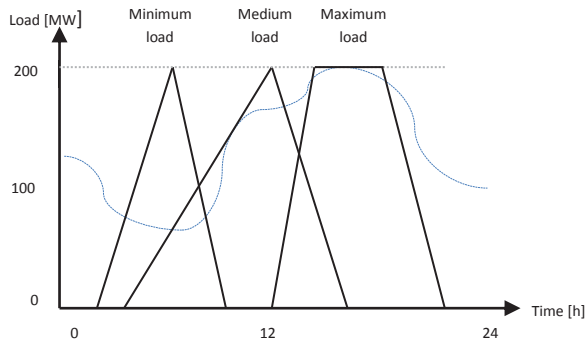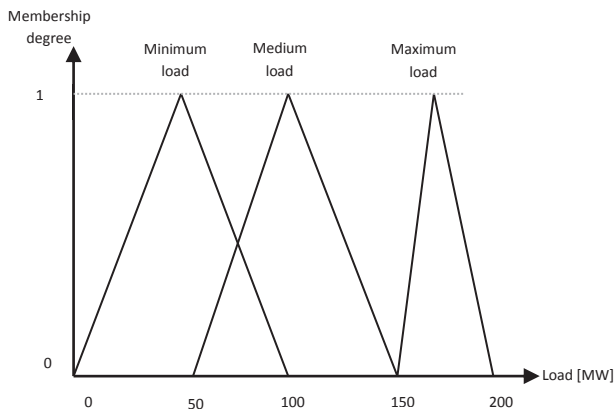

Fig 4 Load membership functions

Fig 5 Definition of the power load membership functions

$$P^{max}: \qquad 150 \leq P \leq 200$$
$$P^{mdl}: \qquad 50 \leq P < 150 \qquad (10)$$
$$P^{min}: \quad 0 < P \leq 100 \ \wedge P^{min} < P^{med}$$

where, P^{max} is the maximum daily power load, P^{med} is the medium value the daily power load, P^{min} is the minimum daily load and P is daily load in the ith hour.

The membership function of the daily load consists of the minimum, medium and maximum load point, Eq. (11). Time duration function consists of the time duration of each of the load points, Eq. (12).

$$P_{DL} = [P^{min}; P^{mdl}; P^{max}] \qquad (11)$$

$$t_{DL} = [t^{min}; t^{mdl}; t^{max}] \qquad (12)$$

where, $t^{min}, t^{mdl}, t^{max}$ are the time durations of the minimum, medium and maximum load.

Total energy not supplied (ENS) is calculated in the following way described.

a) First the daily power loss is determined, which is a number between zero and the power of the failed transformer. If

the power load is less than or equal to the power of the failed transformer, and it can be supplied by other transformers in the network, the subtraction is quantified as zero. And, it is multiplied with its time duration.

$$\Delta P_{dl} = \begin{cases} P_{DL} - P_{TRo,} & if \ (P_{DL} - P_{TRo,}) > 0 \\ 0, & otherwise \end{cases} \qquad (13)$$

$$W_{dl} = \sum_{i=1}^{365} \sum_{j}^{min,mdl,max} \Delta P_{dl,i}^{j} \cdot t_{DL,i}^{j} \ [MWh] \qquad (14)$$

where, ΔP_{dl} is the daily power loss, $P_{TRo,}$ is the power of the failed transformer, and W_{dl} is the total energy loss.

b) Next, the total energy loss is multiplied with the equivalent unavailability, i.e. probability of component failure in the system.

$$ENS = W_{dl} \cdot N_{eqv} \ [MWh] \qquad (15)$$

c) Last, total energy not supplied is sum of the energy not supplied of all possible events' combinations.

IV. TEST EXAMPLE

In [2] a method for network reliability estimation and energy not supplied rate calculation is presented using triangular fuzzy numbers. In this paper that method for network reliability is used and upgraded for daily load data processing. The case study of two transformers connected in parallel in 400/110 kV/kV substation, is reviewed. The system is shown on Fig 6.
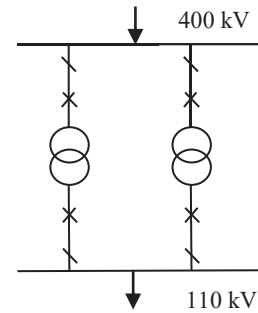

Fig 6 A two parallel transformers system

Each of the branches is equipped with two circuit-breakers (one on 400 kV side, and the other on 110 kV side), two disconnectors (one on 400 kV side, and the other on 110 kV side) and one transformer. Each of the components in the substations has a certain rate of unavailability, outage duration, depending on the voltage rate, as shown in the Table 1(data from [5]).

TABLE 1
SUBSTATION COMPONENTS' UNAVAILABILITY

Components	Voltage (kV)	Unavailability
Circuit breaker	400	$[2,02; 2,32; 2,62] \cdot 10^{-5}$
Disconnector	400	$[4,01; 4,51; 5,01] \cdot 10^{-6}$
Circuit breaker	110	$[3,21; 3,60; 4,09] \cdot 10^{-5}$
Disconnector	110	$[1,06; 1,20; 1,33] \cdot 10^{-6}$
Power transformer	400/110	$[3,50; 3,70; 4,05] \cdot 10^{-4}$

The calculated value of the equivalent unavailability of each of the branches is the following fuzzy set interval:

$$N_{eqv1} = N_{eqv2} = [3,78; 4,03; 4,42] \cdot 10^{-4}$$

The equivalent unavailability of the system calculated with the fuzzy logic is $N_{eqv} = [1,43; 1,62; 1,95] \cdot 10^{-7}$.

Another way to calculate the equivalent unavailability is with Eq. (9). For achieving higher accuracy, it is assumed that the α- cut is $\alpha = 0,9$. The equivalent unavailability in that case is $N_{eqv} = [1,60; 1,65] \cdot 10^{-7}$. For higher safety level, in this case, the results are more accurate than those using fuzzy numbers.

In this paper, it is assumed that the power of each of the transformers is $P = 100$ MW. The first case reviewed is failure of one branch. That means that the consumption greater than 100 MW won't be supplied. The other case reviewed is failure of two branches. In this case, no power is supplied. The power load and the energy not supplied due to one branch failure are presented in Fig 7. For more accurate computation of the energy not supplied, the power not supplied from each day is multiplied with the duration of the power load and the possibility that event occurs, as in Eq. (14) and Eq. (15).
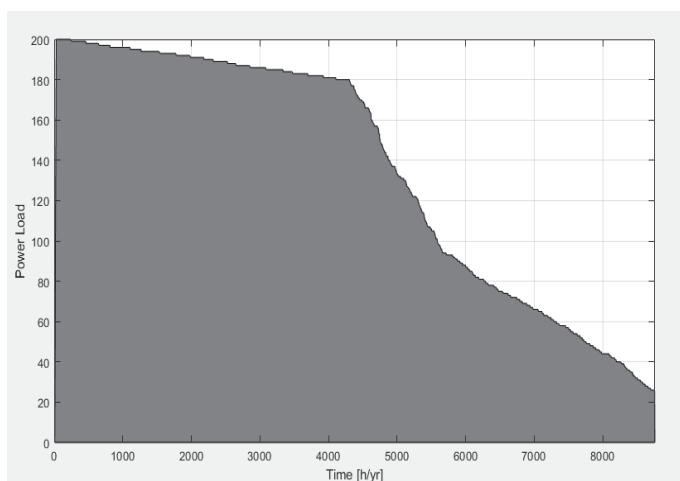

Fig 7 Year power load curve

In Table 2 the equivalent unavailability of the system of all possible combinations of branch failures and the energy not supplied from all the possible combinations is shown.

The total energy not supplied in one year due to one and

two branches failure using the proposed method is:

$$ENS = [332,43; 353,55; 387,95] \text{MWh}$$

V.   CONCLUSION

Building a secure and a hundred percent reliable network is still a challenge for researches and power engineers. Creating a highly accurate method for ENS estimation is important from both technical and economical point of view.

From the presented, it can be concluded that the fuzzy logic based methods can be easily implemented. Therefore, the proposed approach is an expansion of the classical and fuzzy logic based methods for reliability estimation. Considering the fact that power load data is measured daily, it wouldn't be difficult for this method to be applied. Accuracy of the results depends on the power load measured data and the measured time duration of the power load.

VI.   REFERENCES

[1] Nahman, J. (1997). *Fuzzy Logic Based Network Reliability Evaluation.* Belgrade, Republic of Yugoslavia : Faculty of Electrical Engineering.

[2] Iliev, A., & Zlatanoski, M. (2004). Reliability Evaluation of Electric Substations Considering Uncertainty of The Input Data. *Mako Cigre.* Ohrid.

[3] Zimmermann, H.-J. (2001). *Fuzzy Set Theory-and Its Applications, Fourth Edition.* New York: Springer Seience+Business Media .

[4] Fotuhi, M., & Ghafouri, A. (2007). Uncertainty Consideration in Power System Reliability Indices Assessment Using Fuzzy Logic Method. *IEEE Xplore*.

[5] Sharma, M. K. (2017). Possibility and Probability Aspect to Fuzzy Reliability Analysis of a Network System. *Global Journal of Pure and Applied Mathematics*, pp. 3641-3655.

[6] Working Group A3.06. (2012). *Final Report of the 2004-2007 International Enquiry on Reliability of High Voltage Equipment* . CIGRE.

TABLE 2
RELIABILITY ANALYSIS

#	1st branch	2nd branch	Probability	Power not supplied [MW]	ENS [MWh]
1	1	1	/	0	0
2	1	0	$[3,78; 4,03; 4,42]$ $\cdot 10^{-4}$	0-100	$[166,13; 176,68; 193,86]$
3	0	1	$[3,78; 4,03; 4,42]$ $\cdot 10^{-4}$	0-100	$[166,13; 176,68; 193,86]$
4	0	0	$[1,43; 1,62; 1,95]$ $\cdot 10^{-7}$	0-200	$[0,17; 0,19; 0,23]$

Optimal Power Injection Placement in Radial Distribution Systems using Mixed Integer Second Order Cone Programming

Jovica Vuletić, Jordančo Angelov and Mirko Todorovski[1]

Abstract – **Nowadays power injections that originate from distributed energy resources play an important role in every power distribution system. Their proper location, sizing and operation contribute to power systems loss reduction, improved voltage profile, increased reliability and deferral of system upgrades. Many methods deal with the complex combinatorial and non-linear problem of distributed generation optimal placement and sizing. Most of these methods present acceptable and near optimal results but at the expense of using pre-processing procedures based on complex mathematical modelling techniques. This paper presents an alternative solution to that problem. Instead of constructing an algorithm that hopefully, deals best with the issue at hand, the proposed approach places its focus on proper problem shaping, formulation and solution utilizing the powerful benefits of today's solvers. The proposed approach is tested on a 69-bus distribution system and it yields far better results. It also outperforms other methods in terms of simplicity and easy implementation.**

Keywords – **Distributed generation, Power injection, Losses, Voltage profile, Distribution system.**

I. INTRODUCTION

Distributed energy resources play an important role in today's power systems. Their optimal placement and sizing contributes to several beneficial aspects such as decrease in power system losses, voltage profile improvement, increased reliability and ultimately deferral of system upgrades. The trade-off for utilizing these benefits is their optimal placement and sizing. [1]

Many technologies harness the energy from distributed resources. All of them possess some specifics in terms of power conversion, efficiency, operation etc. However, from power system point of view, all of these can be regarded as some sort of (active, reactive and/or apparent) power injection (PI), depending on the technology. In order to utilize the benefits these technologies provide, their proper sizing, location and operation regime is of utmost importance [2].

A plethora of methods exists that deal with the combinatorial and non-linear problem of optimal PI placement and sizing. They can be categorized in several groups, i.e. analytical [4],[5], heuristic and meta-heuristic [6],[7] and mathematical programing algorithms [8],[9]. They all possess some method specific advantages and disadvantages. They treat the

[1]Jovica Vuletić, Jordančo Angelov and Mirko Todorovski are with the Department of Power Systems, Faculty of Electrical Engineering and Information Technologies, "Ss. Cyril and Methodius" University in Skopje, Rugjer Boskovic 18, 1000 Skopje, Republic of Macedonia, E-mail: jovicav@feit.ukim.edu.mk

aforementioned problem in ways that introduce simplifications/complications, linearization, natural process imitation's, coding, decoding etc. The primary focus in these approaches is the method itself and its proper shaping in order to address the problem at hand. This paper proposes a different approach. Rather than spending time on proper method shaping, a suitable reallocation in problem shaping is presented instead. Utilizing the benefits of today's powerful optimization solvers, i.e. YALMIP [11] and CPLEX [12], the problem is solved in a way that disburdens the user from complex mathematical formulations. The obtained results show superiority compared to other methods and approaches that deal with the highly complex non-linear problem of optimal placement and sizing of PI's.

II. PROBLEM FORMULATION

The problem of optimal placement and sizing of PI's is quantified through an objective function that minimizes power system losses, i.e. Eq. **Error! Reference source not found.**:

$$\min F = \sum_{N\text{branch}} R_{\text{branch}} \cdot I_{\text{branch}}^2 \tag{1}$$

This objective function is subject to two subsets of constraints. The first subset C_{system} refers to a set of conditions that describe power system's performance. Instead of using conventional and distribution system appropriate load flow techniques [13]-[15], the power system here is described with set of equations that model and quantify its behavior, i.e. branch power flows and bus voltage profile [14], i.e. Eqs. (2) - (8):

$$V_1^2 = V_{\text{slack}}^2 \tag{2}$$

$$V_{to}^2 = V_{from}^2 - 2 \cdot \left(R_{\text{branch}} \circ P_{\text{branch}}^{from} + X_{\text{branch}} \circ Q_{\text{branch}}^{from} \right) + \left(R_{\text{branch}}^2 + X_{\text{branch}}^2 \right) \circ I_{\text{branch}}^2 \tag{3}$$

$$P_{\text{branch},from} - P_{\text{branch},to} = R_{\text{branch}} \circ I_{\text{branch}}^2 \tag{4}$$

$$Q_{\text{branch},from} - Q_{\text{branch},to} = X_{\text{branch}} \circ I_{\text{branch}}^2 \tag{5}$$

$$I_{\text{branch}}^2 \circ V_{from}^2 \geq \left(P_{\text{branch},from}^2 + Q_{\text{branch},from}^2 \right) \tag{6}$$

$$I_{\text{branch}}^2 \geq 0 \tag{7}$$

$$V^2 \geq 0 \tag{8}$$

Eq. (2) strictly defines the slack bus voltage. Eqs. (3) - (5) describe the voltage profile and branch power flow of the power system [14]. Eqs. (6) - (8) define non-negativity

conditions since those are required from the optimization solver in an explicit form [12]. Constraint (6) is written in a relaxed form and it is converted from an equality constraint, which it really is, to an inequality constraint. In such a way, with (2)-(8) we have defined second order cone programing problem, which is easily solved by state of the arts solver such as CPLEX. Inequality (6) is proven to converge to an equality in any optimal solution [10], so that the final solution will satisfy all network constraints in their original form. The vectors *from* and *to* refer to branch's sending and receiving end indices accordingly and the operator "∘" denotes a Hadamard's product or element wise multiplication of vectors.

The second subset of constraints C_{PI} refers to placement and sizing of different types of PI's, i.e. Eqs. (9) - (16):

$$P_{slack} + \mathbf{C}_P \cdot P_{PI} - P_{demand} = \mathbf{A}_{from} \cdot P_{branch,from} + \mathbf{A}_{to} \cdot P_{branch,to} \quad (9)$$

$$Q_{slack} + \mathbf{C}_Q \cdot Q_{PI} - Q_{demand} = \mathbf{A}_{from} \cdot Q_{branch,from} + \mathbf{A}_{to} \cdot Q_{branch,to} \quad (10)$$

$$0 \leq P_{PI,i} \leq a_{Pi} \cdot P_{PI,max@bus} \quad (11)$$

$$\sum_{i=1}^{N_P} a_{Pi} \leq N_P \quad (12)$$

$$\sum_{i=1}^{N_P} P_{PI,i} \leq P_{PI,total} \quad (13)$$

$$0 \leq Q_{PI,i} \leq a_{Qi} \cdot Q_{PI,max@bus} \quad (14)$$

$$\sum_{i=1}^{N_Q} a_{Qi} \leq N_Q \quad (15)$$

$$\sum_{i=1}^{N_Q} Q_{PI,i} \leq Q_{PI,total} \quad (16)$$

Eqs. (9) and (10) define power balance for active and reactive power. $\mathbf{C}_P$, $\mathbf{C}_Q$, $\mathbf{A}_{from}$ and $\mathbf{A}_{to}$ are generator-bus and branch-bus connection sub-matrices. Sub-matrices are derived from the appropriate connection/incidence matrices for the power system.

Eqs. (11) - (13) introduce limitations on active PI's size per bus and total system PI. Variable a_{Pi} is a binary variable that describes whether there will be an active PI at a certain bus. The binary variable is an element from a vector of length N_P where N_P is a maximum number of locations for active PI placement. $P_{PI,max@bus}$ and $P_{PI,total}$ are maximum and total active PI per bus and in system accordingly.

Similar to the previous triplet of limitations, Eqs. (14) - (16) introduce the same but for reactive PI's. Variables refer to the same quantities using appropriate indices for reactive PI's in this case.

Subset's C_{PI} number of constraints is variable depending on the type of PI placed:
- In case of apparent PI, all Eqs. (9) - (16) apply.
- In case of purely active PI, Eqs. (14) - (16) are omitted from the subset and the second term on the left hand side in Eq. (10) is zero.
- For purely reactive PI, Eqs. (11) - (13) are omitted from the subset and the second term on the left hand side in Eq. (9) is zero.

It is worth mentioning that using YALMIP, the problem can be described with symbolic equations just as the reader can see them in this paper. The latter makes the aforementioned optimization toolbox extremely suitable for problem shaping and optimization. YALMIP also offers possibilities for using other more efficient solvers for various types of problems [10].

III. CONVERGENCE PROPERTIES AND ROBUSTNESS

The proposed approach has no convergence and robustness issues as the utilization of solvers inherently eliminates the infeasible aspects of the optimization process. Computation time mostly depends on the set of possible locations for PI placement. Worst-case scenario suggests a set of locations that contains all buses apart from the slack. It is a user's choice to relax the optimization process by introducing restrictions to this set, i.e. reducing the number of candidate locations.

IV. CASE STUDIES

The proposed approach is applied to the well-known 12.66 kV 69-bus distribution system [16]. Three scenarios are developed and analyzed, i.e. placing of purely active, reactive and apparent PI's at one to three locations accordingly. Obtained results are compared to those from recent studies.

Base case values for this distribution system are $\Delta P_0 = 225.00$ kW and minimum voltage $U_{min@65} = 0.9092$ pu.

For optimization purposes, the following input variables and values are initialized:
- All buses apart from the slack bus (index 1) are potential candidates for PI placement of any type, i.e. set of 68 buses indexed from 2÷69 for potential placement of one to three PI's of the same type, depending on the scenario, i.e. $L_P = L_Q = 2÷69$.
- Maximum active and reactive power per bus is set to three MW/MVAr accordingly, i.e. $P_{PI,max@bus} = 3$ MW, $Q_{PI,max@bus} = 3$ MVAr.
- Maximum total active and reactive power injection in the system is set to five MW/MVAr accordingly, i.e. $P_{PI,total} = 5$ MW, $Q_{PI,total} = 5$ MVAr.
- Number of locations for PI placement varies from one to three locations depending on the scenario, i.e. $N_P = 1÷3$, $N_Q = 1÷3$.
- Vector a_P/a_Q of length L_P/L_Q with binary variables a_{Pi}/a_{Qi} accordingly, that defines whether there will be a PI of active/reactive type in some bus. Maximum number of ones in these vectors corresponds to number of locations for PI placement, i.e. combination of N_P ones in a pool of L_P positions for a_P and N_Q ones in a pool of L_Q positions for a_Q.

TABLE I
COMPARISON OF RESULTS FOR ACTIVE PI PLACEMENT

Location N^0	NH [6] Size@bus (kW)	HPSO [7] Size@bus (kW)	Proposed method Size@bus (kW)
1	$1823^{@61}$	$1810^{@61}$	$1872.7^{@61}$
ΔP (kW)	83.30	83.40	83.22
2	$1733^{@61}$	$1733^{@61}$	$1781.4^{@61}$
	$520^{@17}$	$520^{@17}$	$532.3^{@18}$
ΔP (kW)	71.80	71.80	71.68
3	$1689^{@61}$	$1670^{@61}$	$1719.0^{@61}$
	$312^{@21}$	$380^{@17}$	$381.1^{@18}$
	$471^{@12}$	$510^{@11}$	$526.5^{@11}$
ΔP (kW)	69.70	69.60	69.43

Table I presents comparison of results for active PI placement with two other methods, Novel Heuristic (NH) [6] and Hybrid Particle Swarm Optimization (HPSO) [7]. Results show that for all three considered scenarios, the proposed approach presents better results in terms of system power losses. PI's are of same magnitude order for all considered scenarios and in all three approaches. Locations differ in all three methods when placing three PI's.

TABLE II
COMPARISON OF RESULTS FOR REACTIVE PI PLACEMENT

Location N^0	NH [6] Size@bus (kVAr)	HPSO [7] Size@bus (kVAr)	Proposed method Size@bus (kVAr)
1	$1310^{@61}$	$1290^{@61}$	$1330.0^{@61}$
ΔP (kW)	152.10	152.10	152.04
2	$1224^{@61}$	$1240^{@61}$	$1275.1^{@61}$
	$356^{@17}$	$350^{@18}$	$361.2^{@17}$
ΔP (kW)	146.50	146.50	146.44
3	$1210^{@61}$	$1190^{@61}$	$1232.5^{@61}$
	$226^{@21}$	$250^{@18}$	$231.4^{@21}$
	$320^{@12}$	$330^{@11}$	$412.6^{@11}$
ΔP (kW)	145.30	145.20	145.12

Table II presents comparison of results for reactive PI placement. Comparisons are made to the same methods referenced in Table I. PI size again slightly differs between methods. Proposed approach offers better results and slightly different set of locations when placing three PI's.

Table III compares results from the proposed approach to Improved Analytical (IA) [4]. Apart from the case of single PI placement where IA outperforms the proposed approach, in all other cases it is vice versa. PI's differ in terms of active and reactive PI size. Set of locations is different between methods for three locations. IA presents a same set of buses for the active/reactive part of the apparent PI placement, while proposed approach presents two different sets for the active and reactive part of apparent PI placement that only differs for the second injection.

TABLE III
COMPARISON OF RESULTS FOR APPARENT PI PLACEMENT

Location N^0	IA [4] P@bus+jQ@bus (kVA)	Proposed method P@bus+jQ@bus (kVA)
1	$1531.6^{@61}+j1638.7^{@61}$	$1828.6^{@61}+j1300.7^{@61}$
ΔP (kW)	22.62	23.17
2	$1498.8^{@61}+j1603.6^{@61}$	$1735.3^{@61}+j1239.0^{@61}$
	$450.0^{@17}+j481.4^{@17}$	$522.3^{@17}+j353.4^{@17}$
ΔP (kW)	7.25	7.20
3	$1415.5^{@61}+j1514.5^{@61}$	$1674.4^{@61}+j1195.5^{@61}$
	$424.7^{@17}+j454.4^{@17}$	$379.2^{@17}+j230.5^{@21}$
	$566.1^{@50}+j605.6^{@50}$	$494.3^{@11}+j374.8^{@11}$
ΔP (kW)	4.95	4.26

V. CONCLUSION

This paper presents an approach for optimal siting and sizing of power injections in distribution systems utilizing the benefits of powerful optimization solvers. The approach focuses on appropriate problem shaping instead of proper method shaping. There are several advantages to this approach, amongst which is the user's disencumbrance from complex mathematical formulations. Problem description is intuitive and simple and does not require any pre-processing, which is not the case in other methods. Equations and constraints are written in an understanding and readable way. The need for load flow sensitivity analysis before the optimization begins in order to detect suitable placement locations is completely eliminated through the utilization of a power system model represented with a system-subset of constraints. The approach presents better results in terms of power system losses compared to other methods that deal with the same problem.

REFERENCES

[1] Griffin, T. Tomsovic, K. Secrest, D. Law, A. "Placement of dispersed generation systems for reduced losses", Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000

[2] Pesaran, M. Huy, P. D. Ramachandaramurthy, V. K. "A review of the optimal allocation of distributed generation: Objectives, constraints, methods, and algorithms", Renewable and Sustainable Energy Reviews, 2017, 75, 293 - 312

[3] Ehsan, A. Yang, Q. "Optimal integration and planning of renewable distributed generation in the power distribution networks: A review of analytical techniques", Applied Energy, 2018, 210, 44 - 59

[4] Hung, D. Q. Mithulananthan, N. "Multiple Distributed Generator Placement in Primary Distribution Networks for Loss Reduction", IEEE Transactions on Industrial Electronics, 2013, 60, 1700-1708

[5] Naik, S. N. G. Khatod, D. K. Sharma, M. P. "Analytical approach for optimal siting and sizing of distributed generation in radial distribution networks", IET Generation, Transmission & Distribution, Institution of Engineering and Technology, 2015, 9, 209-220 (11)

[6] Bayat, A. Bagheri, A. "Optimal active and reactive power allocation in distribution networks using a novel heuristic approach", Applied Energy, 2019, 233-234, 71 - 85

[7] Kansal, S. Kumar, V. Tyagi, B. "Hybrid approach for optimal placement of multiple DGs of multiple types in distribution networks", International Journal of Electrical Power & Energy Systems, 2016, 75, 226 - 235

[8] Rueda-Medina, A. C. Franco, J. F. Rider, M. J. Padilha-Feltrin, A. Romero, R. "A mixed-integer linear programming approach for optimal type, size and allocation of distributed generation in radial distribution systems", Electric Power Systems Research, 2013, 97, 133 - 143

[9] Nojavan, S. Jalali, M. Zare, K. "Optimal allocation of capacitors in radial/mesh distribution systems using mixed integer nonlinear programming approach", Electric Power Systems Research, 2014, 107, 119 - 124

[10] M. Farivar, R. Neal, C. Clarke and S. Low, "Optimal inverter VAR control in distribution systems with high PV penetration, " 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, 2012, pp. 1-7.

[11] J. Lofberg, "YALMIP : a toolbox for modeling and optimization in MATLAB" 2004 IEEE International Conference on Robotics and Automation, New Orleans, LA, 2004, pp. 284-289

[12] CPLEX. http://www.ilog.com/products/cplex.

[13] Luo, G. X. Semlyen, A. "Efficient load flow for large weakly meshed networks", IEEE Transactions on Power Systems, 1990, 5, 1309-1316

[14] Baran, M. E. Wu, F. F. "Network reconfiguration in distribution systems for loss reduction and load balancing", IEEE Transactions on Power Delivery, 1989, 4, 1401-1407

[15] Rajicic, D. Ackovski, R. Taleski, R. "Voltage correction power flow", IEEE Transactions on Power Delivery, 1994, 9, 1056-1062

[16] M. E. Baran and F. F. Wu, "Optimal capacitor placement on radial distribution systems" in IEEE Transactions on Power Delivery, vol. 4, no. 1, pp. 725-734, Jan. 1989

Optimal Locations of Energy Storage Devices in Low-Voltage Grids

Jordancho Angelov [1], Jovica Vuletik[1] and Mirko Todorovski [1]

Abstract – **In this paper, we present a methodology for optimal placement of storage devices in low voltage grids. We use CIGRE low voltage benchmark grid as a study case for solving the problem of optimal location and sizing of storage devices. The objective is to maximize the penetration of renewable energy sources (photovoltaic or wind) in a local residential area, while satisfying all grid constraints. The problem is defined as a mixed-integer quadratic programming, with an objective function equal to the total annual costs. The costs include annualized investments in storage devices, operation and maintenance costs and electricity cost separated into categories: bought/sold electricity and network losses. In the model we consider the variations in energy production and consumption by using daily variation curves for each month of the year. We show that the proper placement and sizing of storage devices enables maximum usage of renewable energy sources within the analyzed network keeping the costs at minimum even in cases with unfavorable feed-in tariffs.**

Keywords – **Storage devices, power systems, low-voltage grid, photovoltaic, optimization.**

I. INTRODUCTION

As climate changes become an issue [1], people are turning to embrace ideas for clean energy, such as wind and sun energy [2]. This idea is implementing on a large scale in transmission and distribution networks, as well as, in nowadays into low-voltage grids (LVG).

The main problem with renewable energy sources is their availability. Usually they are available partially throughout the day (for ex. sun energy), and may not match with the energy peaks, during the day. For that reason, we are trying to find a way to storage this "free" energy and use it in times when we need it the most. In that manner, energy storage devices become an attractive idea for storing surplus energy generated from renewable sources in order to be used when we have the greatest electrical or economic benefit.

As storage devices we may use batteries, gas compression facilities, liquid based storage, etc. In particular, Battery Storage (BS) in LVG is considered to be a promising technology for that matter [3]-[4].

The main idea of using storage devices in LVG is to make the consumers more independent from the main source and in that manner to allow lowering their electricity costs. This is particularly attractive in net-metering tariff schemes, as well as,

[1]Jordancho Angelov, Mirko Todorovski and Jovica Vuletik are with the Department of Power Systems, Faculty of Electrical Engineering and Information Technologies, "Ss. Cyril and Methodius" University in Skopje, Rugjer Boskovic 18, 1000 Skopje, Republic of Macedonia, E-mail: jordanco@feit.ukim.edu.mk, mirko@feit.ukim.edu.mk and jovicav@feit.ukim.edu.mk.

in LVG with limited capacity. In some cases, they can be used for balancing short-term fluctuations and for alleviating grid congestion.

II. PROBLEM FORMULATION

A. Problem Statement

In this section we define the problem that we aim to solve. The basic assumption is that we already have the locations of the residential houses were the renewable sources are installed, and for that reason they are not part of this analysis. As we have stated before, the main goal is to find the best location for installing storage devices and define their size in LVG in such a manner that residential houses shall be more independent from the main energy source. In that way, they shall gain their energy "independence" and lower their energy bills, and also lowering grid losses.

Finding best location for placing and sizing of the elements in a network is a combinatorial problem and for that purpose we shall use OPF method to gain the solution. For the economics of the problem at hand we observe that large portion of the investment is associated with equipment costs, while remaining portions covers costs related to grid connection, foundations, buildings, losses, etc.

Assuming that this kind of projects are finances by debt, we should annualize the capital costs using an appropriate Capital Recovery Factor (CRF), which is defined as:

$$CRF(I,N) = \frac{I \cdot (1+I)^N}{(1+I)^N - 1},\qquad(1)$$

where I is the interest rate in decimal fraction and N is the loan term in years.

The largest portion of the costs are for the equipment purchase, which we refer to as capital costs:

$$C_{\text{capital}} = \sum_{q=1}^{n_{loc}} Nbat_q \cdot C_{\text{battery}}\ [\$],\qquad(2)$$

where $Nbat_q$ is the number of batteries installed at location q and C_{battery} is the cost for each battery. In each location we are allowed to install a certain number of batteries of predefined unit size. Because the project is financed by debt, we must take into consideration the loan L which is given as a percentage of the capital cost:

$$L = \frac{C_{\text{loan}}}{100} \cdot C_{\text{capital}}\ [\$],\qquad(3)$$

with C_{loan} being the percentage of the capital costs payed by the loan.

Operations and maintenance costs (OM) include regular maintenance, repairs, insurance, administrations, etc. In general, OM costs depend not only of yearly equipment usage, but also from their age. To find a levelized cost estimation for the energy delivered by the batteries, we must divide the annual costs by annual energy delivery. To find annual cost, we must spread the capital costs out over the projected lifetime using an appropriate factor and then add in an estimate of annual OM. OM's costs are defined as:

$$OM = \frac{OM_{econom}}{100} \cdot C_{capital} \ [\$] \tag{4}$$

where OM_{econom} is OM rate of costs, expressed as a percentage.

One segment of the annual costs are the annual payments for the debt and they are defined as:

$$C_{payment} = L \cdot CRF(I, N) \ [\$]. \tag{5}$$

The equity return is defined as percentage of the difference between capital costs and the loan:

$$ER = \frac{ER_{rate}}{100} \cdot \left(C_{capital} - L \right) [\$], \tag{6}$$

where with ER_{rate} is defined the percentage rate for the desired equity return.

Other costs that must be taken into consideration, are the costs due to the power losses. This costs are defined by the network model (see subsection B). The total grid loss costs are defined for multi-period as hourly snapshot per day are taken into consideration, in period of one year:

$$C_{loss} = \sum_{l=1}^{ns} \left(\sum_{m=1}^{nt} \left(\sum_{k}^{nb} Grid\_loss_k \cdot C_{loss\_cost} \right)_m \right)_l [\$], \tag{7}$$

where with $Grid\_loss_k$ is defined the k-th branch losses (see subsection B) and C_{loss_cost} is the cost of the power losses expressed as \$/kWh. With nt and ns we define the number of periods per day, which in our case is $nt = 24$, and number of characteristics days/snapshots per year, which is $ns = 12$.

If the residential house is pulling energy from the main energy source, that will contribute to costs rise. This costs are defined as energy buy in each period of the day for all snapshots in the year:

$$C_{EB} = \sum_{l=1}^{ns} \left(\sum_{m=1}^{nt} \left(energy\_buy \right)_m \cdot snapDay \right)_l [\$]. \tag{8}$$

In certain time intervals when the renewable sources in the network are producing more than the energy needs of the load demand there will be a possibility to sell energy upstream into the medium voltage network, which shall introduce revenue for the residential houses where storage devices are installed. This revenue is defined by (9).

The energy_buy and energy_sell in Eqs. (8) and (9) are costs/revenue expressed in \$/kWh, for buying/selling energy from/into the grid.
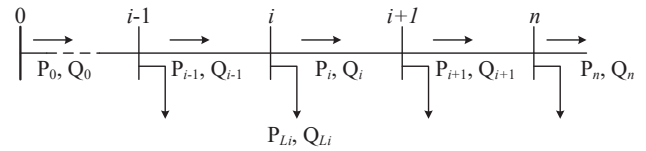

Fig. 1 One line diagram of radial grid

$$C_{ES} = \sum_{l=1}^{ns} \left(\sum_{m=1}^{nt} \left(energy\_sell \right)_m \cdot snapDay \right)_l [\$]. \tag{9}$$

Taking into consideration all above mentioned costs, the objective function that shall lead us to the solution of the location and sizing problem of storage devices is defined as:

$$C_{annual} = C_{payment} + ER + OM + C_{loss} + C_{EB} + C_{ES} \ [\$]$$
$$F_{obj} = \min C_{annual} \ [\$] \tag{10}$$

B. Network Model

Low-voltage grids normally operates as radial networks, however, their configuration could be changed during operation for some reasons (load transfer, loss reduction, overloads relieves, etc.). Mainly, grid reconfiguration are done to lower the grid losses, so in some way grid model affects upon the total costs. Reconfiguring LVG was simply-made in the past, as they were treated as passive ones, but nowadays, they become more active as result of renewable source penetration.

To define the grid losses and accordingly to quantify their costs, we costs, we need to use a set of power flow equations, [5]. To illustrate them, consider the grid depicted in

Fig. 1 The lines impedances are represented by $z_i = r_i + jx_i$, and loads as constant power sinks, $S_{Li} = P_{Li} + jQ_{Li}$. The power flow in radial grid can be described by set of recursive equations, that use real power, reactive power and voltage magnitude at the sending end of a branch (P_i, Q_i, U_i) respectively to express the same quantities at the receiving end of the same branch as follows:

$$P_{i+1} = P_i - r_i \frac{P_i^2 + Q_i^2}{U_i^2} - P_{Li+1}$$

$$Q_{i+1} = Q_i - x_i \frac{P_i^2 + Q_i^2}{U_i^2} - Q_{Li+1} \tag{11}$$

$$U_{i+1}^2 = U_i^2 - 2\left(r_i P_i + x_i Q_i \right) + \left(r_i^2 + x_i^2 \right) \frac{P_i^2 + Q_i^2}{U_i^2}$$

Hence, if the set of real power, reactive power and voltage magnitude are known for the first node (P_0, Q_0, U_0), then the same quantities could be calculated at the other nodes, by applying Eq. (11).

Having the network model, now we can express the power loss in terms of system variables, [6].

To set the proper grid configuration, that represents the condition of minimal grid losses, we need to minimize the total $i^2 r$ losses in the system, which can be calculated as follows:

$$Grid\_loss = \sum_{i=0}^{n-1} r_i \frac{P_i^2 + Q_i^2}{U_i^2} \qquad (12)$$

The set defined by Eq. (11) could be simplified by noting that the quadratic terms in the equations represent the losses on the branches and hence they are much smaller than the branch power terms P_i and Q_i. Therefore, by dropping these second order terms we can get a new set of branch equations of the following form.

$$\begin{aligned} P_{i+1} &= P_i - P_{Li+1} \\ Q_{i+1} &= Q_i - Q_{Li+1} \\ U_{i+1}^2 &= U_i^2 - 2\left(r_i P_i + x_i Q_i\right) \end{aligned} \qquad (13)$$

Since the LVG are radial by nature, the solution for the simplified set of Eq. (13) can be obtained easily. For the given radial grid at

Fig. 1 the solution is defined as:

$$\begin{aligned} P_{i+1} &= \sum_{k=i+2}^{n} P_{Lk} \\ Q_{i+1} &= \sum_{k=i+2}^{n} Q_{Lk} \\ U_{i+1}^2 &= U_i^2 - 2\left(r_i P_i + x_i Q_i\right) \end{aligned} \qquad (14)$$

Now, the power losses on a branch can be approximated as:

$$Grid\_loss = \sum_{i=0}^{n-1} r_i \left(P_i^2 + Q_i^2\right), \qquad (15)$$

as we used the fact that $U_i^2 \approx 1\ p.u.$

C. Constraints

The objective function given by Eq. (10) is presented by the sum of several costs: battery investment, operation and maintenance, grid losses, etc. The solution of Eq. (10) must meet all the models constraints. A set of constraints is deriving from the grid, that include real and reactive power limits of the generators ($i=1,\ldots,ng$) (PV systems) as a function of weather conditions for the considered multi-period (nt-hourly level, ns-monthly level) Eq. (16), branch current constraints as no overload branches ($k=1,\ldots,nb$) are allowed Eq. (17), node voltages $\left(U_j^2 = V_j\right), j = 1,\ldots,nj$ must be in the defined range Eq. (18), as well as voltage constraints for the main supply bus Eq. (19) and the branches ($k = f\text{-}t$) Eq. (20).

$$\begin{aligned} 0 &\le P_{i,nt,ns}^{PV_{gen}} \le P_{i_{\max},nt,ns}^{PV_{gen}} \\ 0 &\le Q_{i,nt,ns}^{PV_{gen}} \le Q_{i_{\max},nt,ns}^{PV_{gen}} \end{aligned} \qquad (16)$$

$$-I_{k_{\max},nt,ns}^{branch} \le P_{k,nt,ns}^{branch} \le I_{k_{\max},nt,ns}^{branch} . \qquad (17)$$

$$\begin{aligned} V_{j_{\min},nt,ns} &\le V_{j,nt,ns} \le V_{j_{\max},nt,ns} \\ 0{,}9^2 &\le V_{j,nt,ns} \le 1{,}1^2 \end{aligned} \qquad (18)$$

$$V_{1,nt,ns} = 1\ p.u.. \qquad (19)$$

$$V_{t,nt,ns} = V_{f,nt,ns} - 2 \cdot \left(r_k \cdot P_{k,nt,ns} + x_k \cdot Q_{k,nt,ns}\right). \qquad (20)$$

Constraints written in vector-matrix form, given by Eqs. (21) and (22), defines the real and reactive power balance per node for the considered multi-period, while by Eq. (23) we define the supply node constraints which prohibit situations when the purchased ($\mathbf{P}_{EB}$) and sold ($\mathbf{P}_{ES}$) real power are simultaneous non-zero. $\mathbf{C}g$ and $\mathbf{A}$ are connection matrices for the generator-bus and branch-bus, respectively and **bin** is vector-matrix variable with two stages 0 and 1.

$$\mathbf{P}_{EB} - \mathbf{P}_{ES} + \mathbf{C}_g \cdot \mathbf{P}_g - \mathbf{P}_{load} + \mathbf{P}_{discharge}^{battery} - \mathbf{P}_{charge}^{battery} = \mathbf{A} \cdot \mathbf{P}_{branch} . \quad (21)$$

$$\mathbf{Q}_{EB} + \mathbf{C}_g \cdot \mathbf{Q}_g - \mathbf{Q}_{load} = \mathbf{A} \cdot \mathbf{Q}_{branch} . \qquad (22)$$

$$\begin{aligned} 0 &\le \mathbf{P}_{EB} \le \mathbf{bin} \cdot 1000 \\ 0 &\le \mathbf{P}_{ES} \le \left(1 - \mathbf{bin}\right) \cdot 1000 \end{aligned} \qquad (23)$$

Other set of constraints is derived from the storage devices and their element (battery) properties for energy storage and discharge, under the assumption that no storage devices will be installed at the main supply bus ($Nbat_1 = 0$). The first constraints given by Eq. (24) derived from available power for batteries charging, i.e. if power surplus appear into the grid it will be stored and if the storage device is empty it shall not be storing energy from the main supply bus.

$$\begin{aligned} P_{leftover}^{PV} &= \sum_{i=1}^{ng} P_{i,nt,ns}^{gen} - \sum_{j=1}^{nj} P_{j,nt,ns}^{load} \\ P_{leftover}^{PV} &\left(P_{leftover}^{PV} < 0\right) = 0 \end{aligned} \qquad (24)$$

The number of installed batteries per storage device is limited, Eq. (25). Conditions in which the batteries will simultaneously charging and discharging are not allowed, Eq. (26).

$$0 \le Nbat_i \le Nbat_{\max} . \qquad (25)$$

$$\begin{aligned} 0 &\le \mathbf{P}_{charge}^{battery} \le \mathbf{bin} \cdot Nbat_{\max} \cdot \mathbf{P}_{charge,\max}^{battery} \\ \mathbf{P}_{discharge}^{battery} &\le \left(1 - \mathbf{bin}\right) \cdot Nbat_{\max} \cdot \mathbf{P}_{discharge,\max}^{battery} \end{aligned} \qquad (26)$$

We assume that the stored energy in the battery (SOE) at the end of the day, should be the same as the beginning of the day, Eq, (27) and SOE must be in the defined range, Eq. (28). The SOE for the first hour and the remaining period of the day are defined by Eqs. (29).

$$SOE_{nt} = SOE_1 . \qquad (27)$$

$$Nbat \cdot SOE_{\min} \le SOE_{nt,ns} \le Nbat \cdot Bat\_size \qquad (28)$$

$$\begin{aligned} SOE_1 &= Nbat \cdot SOE_{\min} + P_{charge,1}^{battery} \cdot \eta_{charge} - \frac{P_{discharge,1}^{battery}}{\eta_{discharge}} \\ SOE_{nt} &= SOE_{nt-1} + P_{charge,nt}^{battery} \cdot \eta_{charge} - \frac{P_{discharge,nt}^{battery}}{\eta_{discharge}} \end{aligned} \qquad (29)$$

And the last constraints derived by the maximum charge/discharge power of the batteries.

$$0 \le \sum P^{\text{battery}}_{\text{charge},nt,ns} \le P^{\text{PV}}_{\text{leftover},ns,nt}$$

$$0 \le P^{\text{battery}}_{\text{discharge},nt,ns} \le P^{\text{battery}}_{\text{discharge},\max} \cdot Nbat \tag{30}$$

III. Case Study

The proposed methodology is applied on CIGRE low voltage benchmark grid [7], that propose different approach than [4]. The method is based on the benefits of today's optimization solvers such as CPLEX [8] and algebraic modeling systems such as YALMIP [9].

Fig. 2a) and b) depicts the daily load profiles and PV production for 12 month period, respectively. The data for the load profiles are taken from OpenDSS [10], while the PV production are average of Meteonorm calculated irradiation data, [11]. The proposed method also takes into consideration the net-metering scheme for the purchased and sold energy, Fig. 2c). At Table I the input data for batteries and economic data are given.

TABLE I
BATTERY AND ECONOMIC INPUT DATA

Battery data	Bat_size, kWh	5	$SOE_{\min}$, kWh	1
	C_{battery}, $	200	$P_{\text{charge,max}}$, kW	2
	η_{charge}	0,88	$P_{\text{discharge,max}}$, kW	2
	$\eta_{\text{discharge}}$	0,88	$Nbat_{\max}$/per bus	20
Economic data	I, %	7	OM_{econom},%	3
	N, years	20	ER_{rate}, %	15
	C_{loan}, %	75	$C_{\text{loss\_cost}}$, $/kWh	Fig. 2c)

As result of the optimization [8], the sizing of the storage devices and optimal locations ($Nbat^{@Bus}$) are presented by Table II, as well as the costs.

TABLE II
RESULTS FOR THE LOCATION AND SIZE OF THE BATTERIES

$Nbat^{@Bus}$							
$5^{@8}$	$9^{@9}$	$11^{@10}$ $3^{@13}$	$11^{@14}$	$12^{@15}$	$5^{@16}$	$11^{@17}$	$11^{@18}$
Costs, $							
$C_{capital}$ = 78 000	ER = 2 925	$C_{payment}$ = 5,522·10³		C_{EB} = 2,0389·10⁴			
L = 58 500	OM = 2 340	C_{loss} = 4,8061·10³		C_{ES} = 1,9732·10⁴			
C_{annual} = 1,625·10⁴							

The voltage profiles are depicted by Fig. 2d).

IV. Conclusion

In this paper we have presented a methodology for optimal location and sizing of storage devices (batteries) in LVG. The advantage of this approach is that the problem is modeled using multi-period optimization where we take into account daily variations in energy production and use for each month of the year. The problem is solved by using the mixed-integer quadratic programming solver CPLEX, while the model writing is facilitated by the use of algebraic modeling system YALMIP. In the model we consider the variations in energy

a) Load profiles   b) PV production

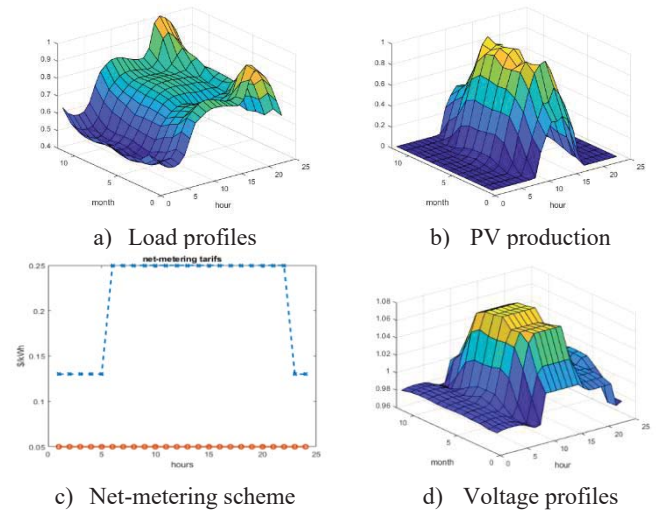c) Net-metering scheme   d) Voltage profiles
Fig. 2 Data for optimization

production and consumption by using daily variation curves for each month of the year. We show that the proper placement and sizing of storage devices enables maximum usage of renewable energy sources within the analyzed network keeping the costs at minimum even in cases with unfavorable feed-in tariffs.

References

[1] F. J. de Sisternes, J. D. Jenkins, and A. Botterud, "The value of energy storage in decarbonizing the electricity sector," *Appl. Energy*, vol. 175, pp. 368–379, 2016.

[2] R. Sims et al., Integration of Renewable Energy into Present and Future Energy Systems, ser. IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[3] K. C. Divya and J. Ostergaard, "Battery energy storage technology for power systems—An overview," *Elect. Power Syst. Res.*, vol. 79, no. 4, pp. 511–520, 2009.

[4] P. Fortenbacher, A. Ulbig, G. Anfersson, "Optimal Placement and Sizing of Distributed Battery Storage in Low Voltage Grids Using Receding Horizon Control Strategies", IEEE Transactions on Power Systems, vol.33, no. 3, May 2018.

[5] M. Baran, F. Wu, "Network Reconfiguration in Distribution Systems for Loss Reduction and Load balancing", *IEEE Transactions on Power Delivery*, Vol. 4, No. 2, April 1989

[6] M. E. Baran, and F.F. Wu, "Optimal Sizing of Capacitors Placed on a Raddial Distribution System", presented at IEEE Winter Meeting, Feb. 1-6, 1988, paper no 88WM 065-5

[7] "Benchmark systems for network integration of renewable and distributed energy resources", Cigre Task Force C6.04.02, Tech. Rep, 2014

[8] CPLEX. http://www.ilog.com/products/cplex.

[9] J. Lofberg, "YALMIP: a toolbox for modeling and optimization in MATLAB" 2004 IEEE International Conference on Robotics and Automation, New Orleans, LA, 2004, pp. 284-289

[10] OpenDSS https://www.epri.com/#/pages/sa/opendss?lang=en

[11] Meteonorm, https://meteonorm.com/en/

Bayesian estimation of the solid oxide fuel cell model

Đani Juričić[1], Gjorgji Nusev[1], Boštjan Dolenc[1], Pavle Boškoski[1]

Abstract – **Reliability and duration of solid oxide fuel cells (SOFC) systems have still to be improved for the sake of more extensive commercialisation. Accurate parameter estimates of the SOFC dynamics are thus an important prerequisite for reliable on-line assessment of their internal condition. Apart from the conventional approaches that evaluate only point estimates, we suggest capturing the full information on the estimates i.e. their probability density functions. The paper delivers assorted results of the experiments conducted on a short-stack solid oxide fuel cell system.**

Keywords – **Solid oxide fuel cell, Bayesian estimation, fault detection, Equivalent circuit model.**

I. INTRODUCTION

Solid-oxide fuel cell systems (SOFCs) are devices that perform the conversion of chemical energy of fuels into electrical energy and heat. SOFCs can operate on a broad power range from a few kW up to several hundred kW or even MW, thus covering the needs of the residential and other stationary applications. The SOFC technology exhibits over 50% electrical efficiency (in less than 60% of cases), which is an advantage for integration into low-energy buildings. It is a trend for today's and tomorrow's construction where heat demand will decrease and the need for electricity increased.

Although promising results have been achieved in the SOFC, the main problem for their commercial use remain its operational reliability and unsatisfactory life expectancy. Due to the high operating temperature (700-800 ° C), the systems are more complex and the problem is the installation of sensors to monitor the situation inside. As far as life expectancy is concerned, the Julich Development Center in December 2015 achieved a world record of malt, which lasted 70,000 hours at an average degradation rate of 0.6% / 1000 hours of operation. The test was performed under laboratory conditions with a highly instrumented short fund (SS), while long-term tests on real systems on the field do not yet exist.

A way to make SOFC technology more competitive on the market is to use techniques for early detection of injuries during the operation. Damage should be detected as soon as possible so that corrective measures can be taken in good time. The conventional approaches are electrochemical impedance spectra, relaxation time distribution (DRT), and evaluation of the equivalent circuit of the replacement model parameters (ECM). A change in the internal state of the cells, whether it is a degradation mechanism or a fault, affects the impedance

[1]All from Jožef Stefan Institute, Department of Systems and Control, Jamova 39, 1000 Ljubljana, Slovenia.

curves and the associated parameters of the ECM circuits. Since SOFCs are characterized by eigenmodes on a broad range, meaningful characterization requires excitation from mHz to tens of kHz. Classically, the system is successively probed with a sinusoidal current of the selected frequency and a small amplitude around the operating point in order to evaluate the impedance (Nyquist) curve from the amplitude and phase of current and the voltage. The problem is that low frequencies require a long time to perform. That significantly prolong the test, especially if a number of experiments have to be repeated at low frequencies. An additional problem is that during the long tests, the potential of the disturbances to spoil the results increases.

Therefore, we proposed an approach that uses broadband excitation signals, i.e. pseudo-random binary noise (PRBS). From the complex wavelength analysis (CWT) of the input and output signals, it is possible to calculate the impedance characteristics of the system with resolution defined by the sampling rate. From impedance, the parameters of the equivalent circuit (ECM) are evaluated. Because it is about optimizing criterion functions that are poorly-conditioned, the evaluation process is done in two steps, first by using the genetic algorithm the most promising solution is found, and then using the simplex method to find the optimal in its vicinity. To find the distribution of the estimated parameters, the Markov Chain Monte Carlo (MCMC) approach is used.

The process was evaluated on a short 6-cell line of solid-oxide fuel cells (SOFC) for a period of 3600 hours. The first results obtained on a short stack are presented.

II. PRELIMINARIES

A. Data acquisition

To evaluate electrochemical impedance spectra (EIS) requires voltage and current data. In the context of electrochemical energy systems, the highest frequency at which EIS is analysed is usually in the interval of 10kHz. Typically, multi-channel data acquisition systems offer interlaced sampling, which inevitably introduces error in the phase estimation. errors induced by the interlacing approach.

In order to guarantee a viable fit, the first step is to validate the correctness of the impedance data. For linear, causal and stable systems the impedance curve is a complex analytic function in the upper half-plane. For such functions, the real and imaginary part are linked through the so-called Kramers-Kronig (KK) relations [2]. Later Bode [3] was the first to successfully apply these relations to electrical impedances and since then it has become a basic tool for checking the validity of the obtained data.

The KK relations linking the real and imaginary components are:

$$\Im(Z(\omega)) = -\frac{2\omega}{\pi} \int_0^\infty \frac{\Re(Z(x)) - \Re(Z(\omega))}{x^2 - \omega^2} \, dx \tag{1}$$

The relations (1) requires integration up to ∞, which for finite length discrete measurements imposes inevitable error. A solution that circumvents this numerical calculation issue is provided by the so-called Z-HIT method [4]. The method allows calculation of the amplitude spectrum $H(\omega)$ from the phase spectrum $\phi(\omega)$ as:

$$\log|H(\omega_0)| \approx \text{const} + \frac{2}{\pi} \int_{\omega_{max}}^{\omega_0} \phi(\omega) \, d\log\omega + \gamma \frac{d\phi(\omega_0)}{d\log\omega} \tag{2}$$

where $\gamma = -\pi/6$ and $\omega_0 \in [\omega_{min}, \omega_{max}]$. Derivation $\frac{d\phi(\omega_0)}{d\log\omega}$ was calculated using the Savitzky–Golay filter. The integration constant is determined by the least squares fit. An example of the performance of the ZHIT test is shown in Fig. 1.
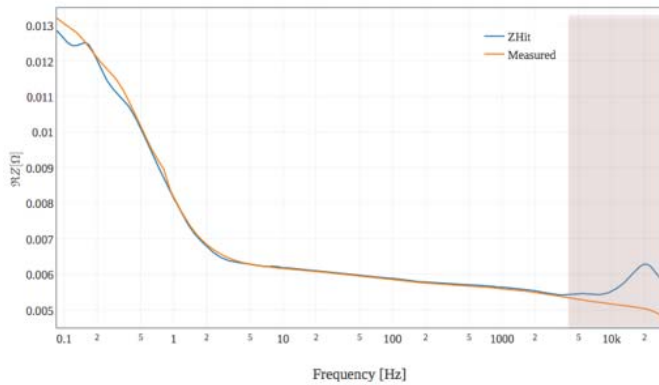
Fig. 1. Comparison between measured modulus and modulus reconstructed from phase by using the ZHIT rule. The ZHIT test starts showing inconsistencies after cca. 5kHZ.

B. Linearized SOFC dynamic model

The model of the linearized SOFC system dynamics in *complex space* can be expressed by the transfer function

$$Z(j\omega) = R_s + \sum_{i=1}^{N_D} \frac{R_i}{\tau_i(j\omega)^{\alpha_i} + 1} \tag{3}$$

where $0 < \alpha \leq 1$ stands fpr fractional derivative. Starting from measured input/output data, obtained during probing with the PRBS, the characterisation of the SOFC is performed by estimating the parameters of the model (3). This can be done either in the time domain by using the approach presented in [5] or in the frequency domain. The latter option relies on the evaluation of the transfer function (3) after evaluating the complex wavelet transform of voltage and current [1]. For the sake of simplicity, in the sequel, we will stick to the second option.

To capture the full information on the parameters it is necessary to set up the estimation problem in the probabilistic framework. By doing so, not only point estimates of the model parameters are obtained, but also their corresponding uncertainties. Uncertainties are much too often neglected in practice, however, they indirectly bear a valuable piece of information as a resultant of the quality of measurements, noise conditions, the importance of a parameter etc. For that purpose, the Bayesian approach based on MCMC approach is adopted.

III. BAYESIAN ESTIMATION OF THE FRACTIONAL ORDER SYSTEM MODEL

The idea of a Bayesian approach is to fuse *prior information* on an unknown variable $\boldsymbol{\theta}$ with the information on that variable *contained in the data* $D = \{d_1, \dots, d_n\}$.

Given a model and data D, the posterior distribution of the *unknown* parameters $\boldsymbol{\theta}$ can be evaluated via the Bayes rule:

$$p(\boldsymbol{\theta}|D) = \frac{p(D|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{\int_{-\infty}^{\infty} p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{4}$$

where $p(D|\boldsymbol{\theta})$ is a likelihood, $p(\boldsymbol{\theta})$ is prior, and and $\int p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is called a marginal likelihood or model evidence.

Selecting the model structure in Bayesian inference is the most crucial part of the modelling procedure. Since the model structure of the fuel cell impedance is well defined in frequency space $\{\omega_1, \dots, \omega_N\}$, one can easily construct the likelihood function:

$$p(D|\boldsymbol{\theta}) = \prod_i p(\omega_i|\boldsymbol{\theta})$$

where we assume that measurement points i on the Nyquist curves are independent and that $p(\omega_i|\boldsymbol{\theta})$ s defined in the following way

$$p(\omega_i|\boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(y_i - \Re\{Z(j\omega_i|\boldsymbol{\theta})\})^2}{2\sigma^2}\right)$$

The parameter vector $\boldsymbol{\theta}$ contains the resistances R_k, time constants τ_k and rational exponents α_k for each of the k^{the} element in the ECM. In the above equation, y_i denotes the real part of the measured point on the Nyquist curve at a frequency ω_i, i.e. $y_i = \mathcal{R}e(Z(j\omega_i))$.

Note that, in order to estimate the parameters $\boldsymbol{\theta}$ in (3), only one component of the complex impedance is required, hence $\Re\{\bullet\}$. Due to the Kramers-Kronig relation, it is enough to consider only the information either in real or imaginary parts of the complex value.

The prior probability of the parameter $p(\boldsymbol{\theta})$ is generally used to incorporate any prior knowledge about the modelled system. For the purposes of condition monitoring, it can incorporate knowledge about the model parameters from measurements that were done at some previous time instance e.g. at

commissioning of the system. In this work, without any prior knowledge, an un-informative truncated normal prior $p(\boldsymbol{\theta})$ was used for the model parameters R_i and τ_i, since these parameters always take positive values. On the other hand, uniform distribution was used as the prior parameters.

Having specified the prior and the likelihood of the data, the posterior distribution of the model parameters can be inferred by employing (4). However, with complex models, the analytic solution for $p(\boldsymbol{\theta}|D)$ is rarely possible due to the integral in the denominator. Readily, there are few options that resolve this issue by employing numerical methods. The integral can be solved by using grid approximation. However, this becomes computationally demanding with higher dimensions of the parameter space. On the other hand, the Markov Chain Monte Carlo (MCMC) simulations can be used to sample directly from the posterior distribution without the need to evaluate $p(\boldsymbol{\theta}|D)$. In such a case, only the proportional part of the equation (3) is required to be numerically tractable:

$$p(\boldsymbol{\theta}|D) \propto p(D|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

More on MCMC methods and Bayesian inference can be found in [6]. The demonstration is available online to fully visualise the approach at the link https://chi-feng.github.io/mcmc-demo/

IV. EXPERIMENTAL RESULTS

C. The experiment in brief

A SOLIDpower SOFC short stack operating at 750°C was used in the experiment. The stack consisted of six planar anode-supported cells which were installed in an insulated ceramic housing. The active area of a single cell was 80 cm2. The SOFC short stack was fed with a mixture of hydrogen and nitrogen with a flow rate H_2/N_2=0.216/0.144 Nl h^{-1}1cm^{-2} whereas the air flow rate was 4 Nl h^{-1}1cm^{-2}. Stack was operated at a nominal current of 32 A (0.4 A cm^{-2}) with fuel utilization FU=77.5 %. The experiment took $\approx$3600 hours. During this period, characterization was automatically performed on a regular basis by employing both conventional sinusoidal excitation and PRBS waveforms excitation every 6 hours. Hence a dataset with voltage and current recordings acquired at 600 measurement sessions is obtained.

The stack was first run at nominal conditions for 240 h (40 measurement sessions). After worming up, the first fuel starvation session was performed by gradually raising the fuel utilization (FU) stepwise each 24 h starting from 77.5% up to 92.5% (Figure 2 event E2). This was done by decreasing H_2 flow rate while keeping the current density constant. The second fuel starvation protocol was performed also at the same FU steps as in the first situation, but by keeping H_2 flow rate constant while increasing the current density (Figure 5 event E3). The current was increased according to the following sequence 32 A, 34.06 A, 36.15 A, 38.2 A for each of the FU.
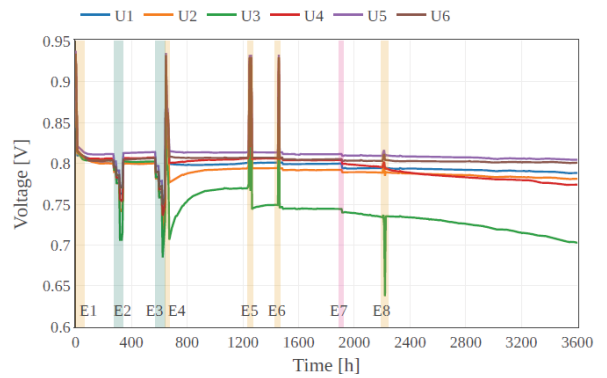
Figure 1: Cell voltages during the experiment.

D. Results obtained on data from one measurement session on a single cell

We will first demonstrate the application of Bayesian inference to the FOS parameter estimation on a single cell. Figure 3 shows the marginal posterior densities of model parameters. The first row shows three-time constants τ_i, the second row shows resistances R_i, and the third row shows the fractional orders α_i. Note that the last row displays only serial resistance. One can notice that the distributions of some of the parameters are far from normal. That is particularly the case for α_1, α_2, R, and R_2.

E. Time evolution of the estimated parameters of cell #3

We will now see how the parameters of the model of cell #3 evolve over time. The time index refers to the measurement session. Yellow strips mark the intervals of too high fuel utilization.

The increase of fuel utilization (events E2) causes the resistances R_3 (related to the slowest mode $\tau_3 \sim 1.2s$) and R_s to blow up (Fig.4). That can be seen also as blow up in the low-frequency part of the Nyquist curve (not shown here). The estimates take unusual values also in cases of incident events like hydrogen supply shutdown (HSS1, HSS2) and power supply shutdown (PSS). The model reacts also to the migration of the equipment into another laboratory.

The estimates of the parameters over time are not smooth but fluctuates (c.f. Figure 4). For instance, the estimates of R_s are relatively smooth while the estimates of R_3 and τ_3 are rather "noisy". Moreover, note that the uncertainty region of the estimates ($\pm\sigma$) is relatively narrow, meaning that the parameter estimation algorithm ends up with rather highly reliable estimates. The explanation for such results should be sought in two limited quality of current sensor as well as fluctuations in fuel flows.

V. CONCLUSION

We presented a Bayesian approach to the parameter estimation of the linearized model of solid oxide fuel cell dynamics. From the change in the marginal pdf's from their reference forms, it will be possible to detect changes in the internal condition of the system.
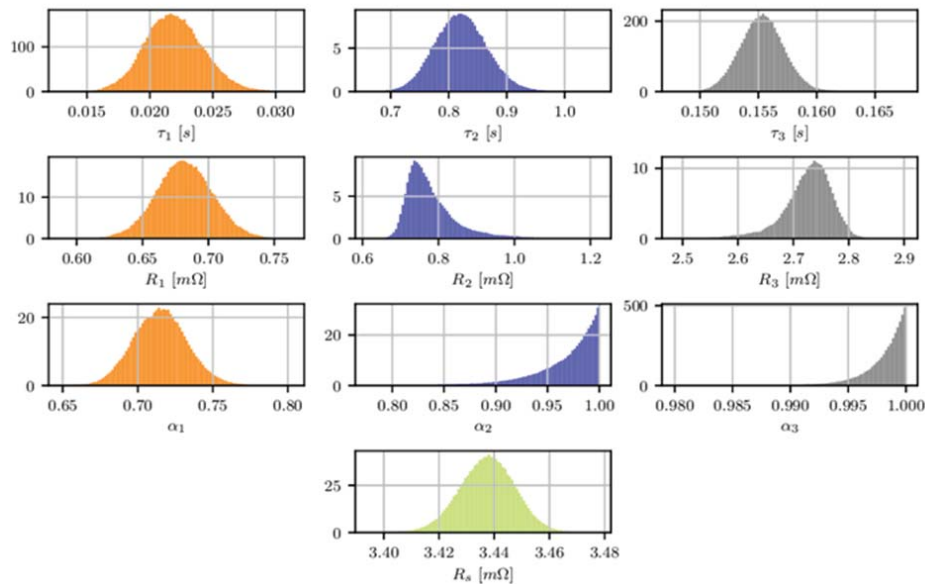
Figure 3: Results of the MCMC approach when applied to the data acquired during one measurement session on cell #3. The plots present the posterior distributions of model parameters from a model with 3 RQ elements.

ACKNOWLEDGEMENT

REFERENCES

[1] P. Boškoski, A. Debenjak, B. M. Boshkoska, Fast Electrochemical Impedance Spectroscopy, Springer, 2017.

[2] P. Agarwal, M. E. Orazem, L. H. Garcia-Rubio, Application of measurement models to impedance spectroscopy III. Evaluation of consistency with the Kramers-Kronig relations, Journal of the Electrochemical Society 142 (12) (1995) 4159–4168.

[3] H. W. Bode, et al., Network analysis and feedback amplifier design, van Nostrand, New York, 1945.

[4] W.O.Strunz, S.Feihl, The Z-HIT Algorithm, URL http://zahner.de/support/application-notes/-the-z-hit-algorithm.html, 2019

[5] B. Dolenc, Gj. Nusev, V. Subotić, Ch. Hochenauer, N. Gehring, Ð. Juričić, P. Boškoski, "Time-domain identification of fractional order systems for health assessment of SOFC", IEEE Transactions on Industrial Electronics (submitted).

[6] S. Brooks, A. Gelman, G. L. Jones and X. –L. Meng (Eds.). Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC, 2011.
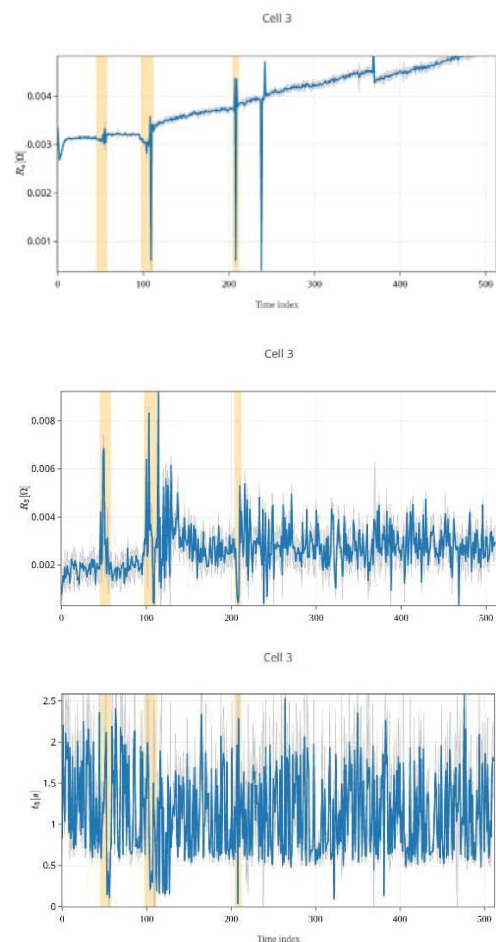
Figure 4: An excerpt from evolution of the model parameters with focus on R_s and R_3 and τ_3.

Comparison of ACSR, ACCC and AAAC Conductors for Overhead Transmission Lines

Metodija Atanasovski[1], Blagoja Arapinoski[1], Mitko Kostov[1], Ljupco Trpezanovski[1]

Abstract – **The goal of this paper is to perform comparison of three types of conductors for transmission overhead lines: Aluminium Conductor Steel Reinforced-ACSR, All Aluminium Alloy Conductors-AAAC and Aluminium conductor Composite Core-ACCC. The paper at beginning summarizes the constructive and technological characteristics of the conductors. Their physical, mechanical and electrical specified parameters are compared in table. Applicative part of the paper presents comparison of ACSR, AAAC and ACCC conductors on one tension field of 110 kV overhead transmission line. Three design solutions of tension field are presented, with usage of standard design calculations for overhead transmission line.**

Keywords – **Conductors, overhead transmission line.**

I. INTRODUCTION

The replacement of classical ACSR phase conductors with other ones from the same type, but with greater cross section (greater transmission capacity), leads to replacement or mechanical reinforcement of support constructions for conductors (towers, consoles, insulator chains etc.) [1]. This type of reconstructions on existing overhead transmission lines (OHTL) is very expensive and inefficient. Namely, if existing phase conductors are replaced with new types of conductors which will have greater ratting current, lower sag on higher working temperatures, smaller mass, and lower electrical resistance, then on very effective and more economical way transmission capacity of the OHTL will be increased [2]. If reconstruction of an existing OHTL results in double transmission capacity, the price of reconstruction is much lower than construction of new OHTL with the same transmission capacity [3]. Construction and application of new types of conductors with better characteristics than existing ones, has become challenge for many companies in the field of manufacturing conductors for transmission of electricity.

The paper presents two types of conductors as an alternative of ACSR conductors. The first one is ACCC (Aluminum Conductor with Composite Core) conductor which for the first time was introduced on the world market in 2005 [7]. This conductor is constructed from aluminium and has circle or trapezoidal shape. It has core produced from polymer composite materials.

There are several companies mainly in USA which

commercially produce ACCC conductors and they are becoming serious competition of standard ACSR conductors.

The second conductor is AAAC (All Aluminium Alloy Conductors). This conductor is produced with heat-treatable magnesium silicon type aluminum alloys. The alloys referred to have higher strength but lower conductivity than pure aluminium. Being lighter, alloy conductors sometimes have advantage compared to the more conventional ACSR, having lower breaking loads than the later. Also their use becomes particularly favorable when ice and wind loadings are low.

The paper is consisted of five sections. Section II summarizes the constructive and technological characteristics of the conductors. Their physical, mechanical and electrical specified parameters are compared in table. Section III presents the applicative part of the paper. Comparison of ACSR, AAAC and ACCC conductors is performed on one tension field of 110 kV OHTL. Three design solutions of tension field are presented, with usage of standard design calculations for OHTL. Several conclusions and further work possibilities about this matter are presented in section V.

II. COMPARISON OF ACSR, ACCC AND AAAC CONDUCTORS

Overrating load current at OHTL with conductors containing steel core, can cause elongation of conductors and spans sag increase to values which will affect the safety clearance. The company CTC Co. (USA) at 2005 produced composite core with a very high ratio of breaking load (force)/weight and small coefficient of thermal expansion. The replacement of steel core with composite one, for conductor with equal outer diameter provides almost doubling of transmission capacity, higher working temperature and significant decrease of sags in spans. Figure 1 depicts the design of conductor with composite core ACCC/TW (trapezoidal wires) compared with ACSR conductor.

Composite core is a result of a process of pultrusion. During the pultrusion process, continuous unidirectional (0^0-axis) carbon fibers form a cylindrically shaped solid core

[1]Metodija Atanasovski, Blagoja Arapinoski, Mitko Kostov and Ljupco Tpezanovski are with the Faculty of Technical Sciences-Bitola, St. Kliment Ohridski University, Macedonia, E-mail: metodija.atanasovski@uklo.edu.mk.
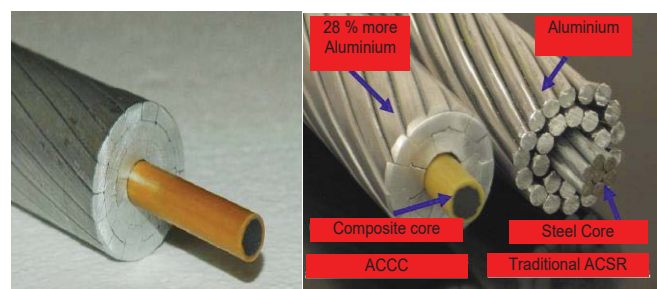
Fig. 1. ACCC/TW conductor and comparison of ACCC and ACSR

while a layer of similarly oriented E-glass fibers, is placed as a shield. The bundled fibers are wet out with a high-temperature toughened epoxy. The fiberglass layer serves two purposes: first, it separates the carbon from the conductive aluminum wires to prevent galvanic corrosion. Second, it counterbalances the more brittle carbon and improves the flexibility and toughness of the core. The one-piece rod composite core lead to a core with a greater cross-sectional area, but a smaller diameter which implies larger loading with reduced tensile stresses over that of the steel core subjected to the same loads.

Fully annealed (soft), H0 (O') tempered 1350 aluminum trapezoidal wires with conductivity of 63% IACS are helically wrapped in one or more layers around the composite core. The use of trapezoidal wires yields compact conductor with less void area than ACSR. The compact trapezoidal wires, coupled with a smaller composite core, result in an ACCC/TW conductor that has approximately 28% more aluminum cross-sectional area than ACSR and ACSS with round wires and same outer diameter. Thermal properties of composite core and soft aluminum wires enable high operating temperatures of ACCC/TW, continuously up to 180 ^{0}C. The greater aluminum content in ACCC/TW with high electrical conductivity, combined with the capability to work at high operating temperatures, can double the ampacity of an existing OHTL with ACSR conductors.

ACCC/TW conductors are produces with standard dimensions with current rating capacity in range from 300 to 3500 A [3]. Although the price of ACCC/TW conductors is three times higher than ACSR conductors with same diameter, the replacement of ACSR conductors with equivalent ACCC/TW and doubling of transmission capacity is three times cheaper than construction of new line with same capacity as the existing one with ACSR conductors. [3].

AAAC conductor is high strength Aluminum-Magnesium-Silicon Alloy conductor. It was developed to replace the high strength 6/1 ACSR conductors. This alloy conductor offers excellent electrical characteristics with a conductivity of 52.5% IACS, excellent sag-tension characteristics and superior corrosion resistance to that of ACSR. Service life of AAAC is around 60 years, twice as durable as ACSR. It is superior to ACSR conductors when used in distribution lines [1]. Figure 2 depicts the design of AAAC conductor.

Comparison of physical, mechanical and electrical specified

AAAC - All Aluminium Alloy Conductor
Al-Mg-Si age hardened Al conductor and core reinforcement (6201-T81
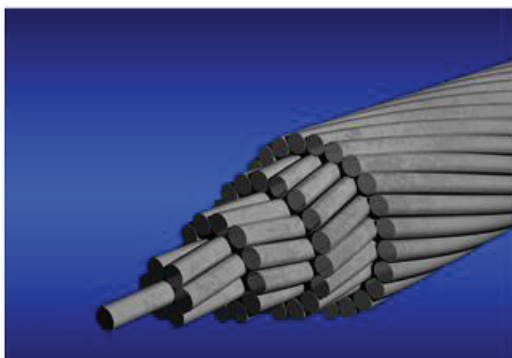
Fig. 2. AAAC conductor

parameters is performed between ACSR 240/40 conductor as traditional standard at 110 kV OHTL and compatible alternatives ACCC/TW conductor Glasgow and AAAC/TW 300 conductor. Data for ACCC conductors can be found in [4] and [7], for ACSR 240/40 conductor is taken from [5], and for AAAC in [6].

Table 1 summarizes mechanical and electrical characteristics of mentioned ACSR 240/40, ACCC Glasgow and AAAC 300 conductor types. It can be concluded that ACCC conductors, for approximately same diameter and cross section as ACSR 240/40 conductor, have: lower weight and mass, higher tensile strength of conductor (achieved with replacement of steel core with composite one), lower coefficient of linear expansion, lower modulus of elasticity, significantly higher working temperature, smaller electrical resistance and higher ratting current. Lower specific weight and smaller coefficients of linear expansion and modulus of elasticity lead to significantly smaller sags even at higher working temperatures.

TABLE I
COMPARISON OF CONDUCTORS

Mechanical and electrical characteristics	Type of conductor		
	ACCC Glasgow	AAAC/ TW 300	ACSR 240/40
Cross section Al/core (mm^2)	236.7/ 47.1	323.98	243/ 39.5
Total cross section (mm^2)	283.8	323.97	282.5
Diameter (mm)	19.53	21.7	21.9
Mass of conductor (kg/km)	731.3	893	987
Weight of conductor (daN/m)	0.717	0.876	0.968
Specific weight 10^{-3} (daN/m,mm^2)	2.526	2.704	3.426
Specific weight of normal ice load 10^{-3} (daN/m,mm^2)	2.803	2.588	2.981
Coefficient of linear expansion 10^{-6} (^{0}C^{-1})	17.1	23	18.9
Modulus of elasticity (daN/mm^2)	6700	6800	7700
Tensile strength (daN)	11500	9560	8640
Normal tension (daN/mm^2)	16.2	12	13
Tension (daN/mm^2)	30.4	22	24.5
DC resistance at 20 ^{0}C (Ω/km)	0.1184	0.1022	0.119
Rated current at 80 ^{0}C (A)	570	775	530
Rated current at 100 ^{0}C (A)	692	951	648
Rated current at 180 ^{0}C (A)	1027	-	-
Maximum working temperature (^{0}C)	180	90	80
Price (Euro/m)	10	6	4

AAAC/TW 300 conductor versus ACSR 240/40 have: lower weight and mass, higher tensile strength of conductor, smaller electrical resistance and higher ratting current. Maximum working temperature of the conductor is 90 ^{0}C.

270

III. STUDY CASE

The development of new regulations for OHTL in many countries in Europe is ongoing process. Almost all European countries have adopted standard for the design of OHTL EN 50341-1. Most of them developed national normative aspect of the standard taking into account special features characteristic for the country. Macedonia has adopted the standard EN 50341-1, but still has no national normative aspect. The design and construction of OHTL is done by Regulation of technical standard for construction of OHTL with voltage 1 kV to 400 kV [1]. For determining loads purposes, all load combinations are calculated with maximum working tension (stress) of the wire. The maximum working stress is usually calculated at -5 ^{0}C with normal ice load or at -20 ^{0}C without ice load, depending on which combination gives governed value [1].

Design replacement comparison of ACSR 240/40 conductors with proposed conductors ACCC/TW Glasgow and AAAC/TW 300 is shown on a tension field of 110 kV existing OHTL. The length of tension field is 1600 m and the terrain is relatively flat and regular (Figure 3).

Climate parameters on the terrain are: wind pressure 75 daN/m^2, coefficient of normal ice load 1.0 and coefficient of extra ice load is 2.0. Maximum working tension (stress) of conductors ACSR 240/40 is 9 daN/mm^2. Grounding wire type Fe III 50 mm^2 is used with maximum working tension 26 daN/mm^2. The OHTL is designed with steel towers type S for suspension supports (towers) and type A-150 for angle tension (strained) towers. Table II summarizes tower spotting list on the terrain (X- is tower position on the terrain on X-axis and Y is tower position on Y-axis. Ruling span of the tension field is 320.79 m.

TABLE II
TOWER SPOTTING LIST

No	Tower type	Tower height	Span (m)	X (m)	Y (m)	Difference in height (m)
1	A-150	15	280	0	130	0
2	S	17.8	310	310	138.13	9.21
3	S	16.8	300	610	145.89	6.86
4	S	16.8	330	940	153.43	7.71
5	S	17.8	335	1275	150	-2.43
6	A-150	18	325	1600	152	3.65

Using the same tower spotting solution, replacement of conductor ACSR 240/40 with ACCC/TW Glasgow and AAAC/TW 300 will be proposed. Glasgow conductor is designed with maximum working stress of 9 daN/mm^2, and AAAC/TW 300 7.8 daN/mm^2, because it has greater cross section than ACSR 240/40.

Figure 3 depicts the tower spotting and catenary curves of OHTL with all three types of conductors for maximum sag temperature +40^{0}C [1].

Several parameters important of OHTL design with each of used conductors are shown in Table III. It can be notified that all three conductors have lower critical span than the ruling span of the tension field. This fact implies that maximum

working tension will appear on temperature -5^{0}C+ normal ice load. Critical temperature for all conductors is lower than 40^0 C, what leads to maximum sag appearance on +40^0 C. It is obvious that conductors ACCC/TW Glasgow and AAAC/TW 300 have higher catenary parameter than ACSR 240/40. This means that these conductors will have lower sags than ACSR 240/40.

This is shown with sag and tension (stress) curves calculated for full range of temperatures from -20^0 C to 100 ^{0} C. Figures 4 and 5 depict sag and tension curves for ruling span of the tension field appropriately. It can be easily concluded from Figure 4 that conductor ACCC/TW Glasgow has lowest sags, significantly lower than other two conductors. Namely Glasgow has in average 2 m lower sags compared with ACSR 240/40. The situation is the same in each span of the tension field. Conductor AAAC/TW 300 has lowest tensions in the full range of temperatures, because it is designed with maximum working tension 7.8 daN/mm^2. Glasgow has lower tension than ACSR 240/40.

Table V summarizes forces table for support tower number 4 which type is S and has highest weight and wind span. The calculation is performed according to loading cases defined in articles 68 and 69 from [1]. It can be notified that mechanical loading of towers in all three axes is lowest when ACCC/TW Glasgow conductor is used. This conductor can be also calculated with maximum working tension 8 daN/mm^2 and in this case has also lower sags than other two conductors. Similar results for tables of forces are obtained for angle strain tower type A-150.

TABLE III
SPECIFIC PARAMETERS FOR OTL DESIGN WITH EACH CONDUCTOR

Parameter	ACCC Glasgow	AAAC 300	ACSR 240/40
Critical span (m)	150.5	156.8	137.1
Critical temperature (^{0}C)	36.3	19.4	23.8
Maximum working tension (daN/mm^2)	9	7.8	9
Tension on 40^{0}C (daN/mm^2)	4.19	3.62	4.54
Catenary parameter 40^0 C (m)	1660.3	1338.8	1324.3

TABLE IV
TABLE OF FORCES FOR SUPPORT TOWER 4 TYPE S

Loading case		ACCC/TW Glasgow			AAAC/TW 300		
		Vx	Vy	Vz	Vx	Vy	Vz
Art. 68_1	A	-	-	628	-	-	683
	B	487	-	318	511	-	361
	C	-	122	318	-	128	361
Art. 69_1	PP	-	1277	628	-	1253	683
	NP	-	-	628	-	-	683
	PZ	-	-	-	-	-	-
	NZ	-	-	628	-	-	683

| Loading case		ACSR 240/40		
		Vx	Vy	Vz
Art. 68_1	A	-	-	733
	B	546	-	409
	C	-	137	409
Art. 69_1	PP	-	1271	733
	NP	-	-	733
	PZ	-	-	-
	NZ	-	-	733

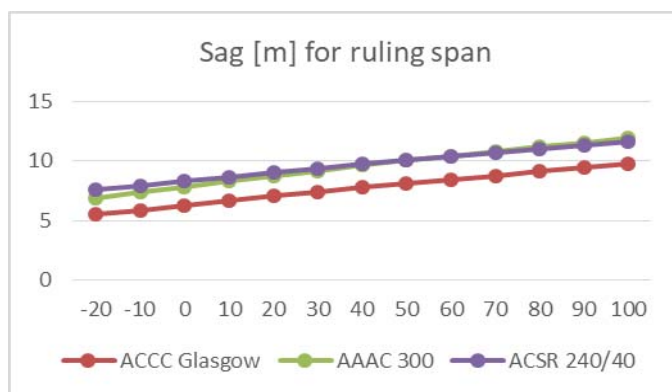Fig. 3. Tower spotting and catenary curves with each conductor

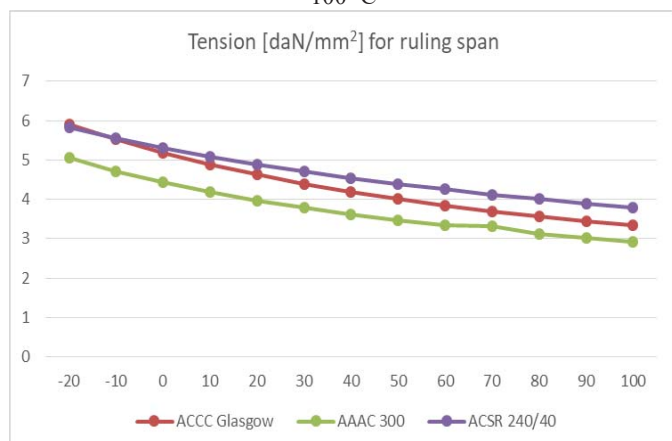Fig. 4. Sags for ruling span for full range of temperatures -20 ^{0}C to 100 ^{0}C

Fig. 5. Tensions for ruling span for full range of temperatures -20 ^{0}C to 100 ^{0}C

IV. CONCLUSION

ACCC conductors are new technology conductors for transmission of electricity, which are becoming a serious competition of mostly used ACSR conductors. ACCC conductors for approximately same diameter and cross section as ACSR 240/40 conductor, have: lower weight and mass, higher tensile strength of conductor (achieved with replacement of steel core with composite one), lower coefficient of linear expansion, lower modulus of elasticity, significantly higher working temperature, smaller electrical resistance and higher ratting current.

AAAC/TW 300 conductor compared to ACSR 240/40 has lower weight and mass, higher tensile strength of conductor, smaller electrical resistance and higher ratting current. The maximum working temperature of the conductor is 90 ^{0}C.

According to the presented results in the paper, it can be concluded that replacement of existing ACSR 240/40 conductor with new one ACCC/TW Glasgow or AAAC/TW 300 leads to significantly: lower sags and stresses in full range of temperatures, lower towers loading, higher ratting current for higher operating temperature and lower electrical resistance meaning losses decrease in transmission of electricity in the OHTL.

The price of ACCC/TW conductors is one of the main disadvantages. However with technology development of this type of conductors, it is reasonable to expect that the price will go down and ACCC will become more competitive versus ACSR. In meanwhile AAAC are good alternative for ongoing projects for conductors replacement during reconstructions of OHTL and increasing transmission capacity. MEPSO (Macedonian Transmission System Operator) is implementing AAAC/TW 300 conductor on several 110 kV OHTL in the ongoing process of reconstruction and network revitalization.

REFERENCES

[1] Technical Normative Regulation for 1 kV to 400 kV OHTL Construction in Republic of Macedonia, 1988 (2018).

[2] Lj. Tpezanovski, A. Markoski and, M. Atanasovski. "Composite Core Conductors for Transmission and Distribution of Electricity with Overhead Lines", *Horizons-Journal of St. Kliment Ohridski University*. No. 1, pp. 31-40. Bitola, December 2008.

[3] Cohen Independent Research Group, Composite Technology Corporation. Report for Investment Recommendation, San Rafael CA, November 2004.

[4] Composite Technology Corporation. *IEC Specification Sheets for ACCC Conductors.*, San Rafael CA, 26.4.2012, www.ctcglobal.com.

[5] Risto Ackoski: *Overhead and Cable Power Lines*, Faculty of Electrical Engineering and IT-Скопје, 2005.

[6] Stipe Mikulic, Davor Bajs, Metodija Atanasovski. "Study for Long Term Development Concepts in Separate Regions of Transmission Network in Republic of Macedonia", Energy Institute Hrvoje Pozar for MEPSO, Zagreb/Skopje 2017.

[7] A. Alawar, E.J Bosze, S. Nutt, "A Composite Core Conductor for Low Sag at High Temperatures", IEEE Transactions on Power Delivery, vol. 20, no. 3, pp. 2193-2199, 2005.
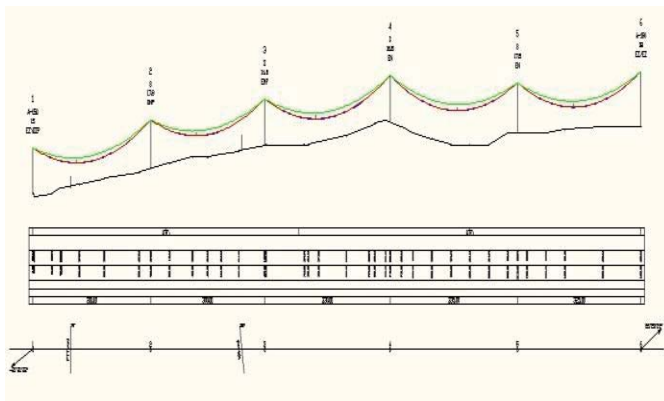
Power System Load Forecasting by using Sinuses Approximation and Wavelet Transform

Mitko Kostov[1], Metodija Atanasovski[1], Gordana Janevska[1], Blagoja Arapinoski[1]

Abstract – **Power system load predictions are important factors for planning the future electrical energy consumption. One of the important factors that influence the power load is the air temperature. This paper analyses the correlation between the power system load and the air temperature. Power load forecasting is investigated by using a combination of the best fitting approximation with sum of sinuses and wavelet transform.**

Keywords – **Power system load, Forecast, Air temperature, Correlation, Regression, Wavelet Transform.**

I. INTRODUCTION

Electricity is one of the most important and inseparable factors in social life. Each electricity company has a strategic goal to provide end users with reliable and stable power supply. Having in mind that it cannot be stored as it should be generated as soon as needed, electricity power system load forecasting is a key factor in the functioning of power systems. Moreover, renewable energy sources are increasingly being incorporated in power systems in an effort to reduce CO_2 emissions and reliance on fossil fuels. Furthermore, the use of low carbon technologies, such as electric vehicles and heat pumps, has increased in recent years. Therefore, accurate predictions are important for planning the future electrical energy consumption - they lead to significant savings in operating and maintenance costs and increased reliability of the electricity supply and delivery system.

Power load forecasting is the process of forecasting future electrical energy demand so that to cope with the growing needs. Power load forecasting would determine which power units should increase their production and which generators should be dispatched. According to the period of the load forecasting, it can be classified into three categories: short-term load forecasting (one hour to one week), mid-term load forecasting (one week to one year) and long-term load forecasting (longer than a year) [1].

Factors that play key role in forecasting of power load consumption are the air temperature, type of the day (weekday, weekend or holiday), geographical differences, people standard, demographic information, etc. This paper analyses the correlation between the power system load and the air temperature in Republic of Macedonia. In addition, forecasting of the power system load consumption is investigated. The power system load is estimated by applying a combination of best fitting function of sum of sinuses and discrete wavelet transform.

[1] Mitko Kostov, Metodija Atanasovski, Gordana Janevska and Blagoja Arapinoski are with the Faculty of Technical Sciences-Bitola, St. Kliment Ohridski University, Republic of Macedonia, E-mail: mitko.kostov@uklo.edu.mk.
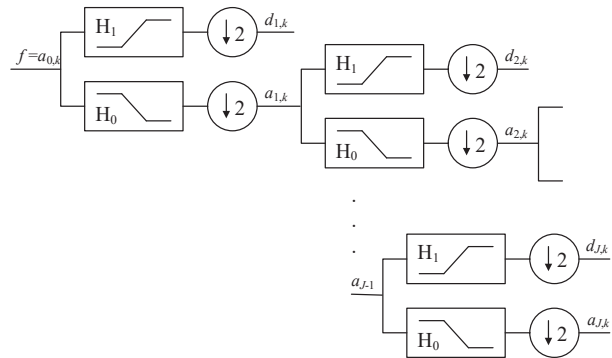
Fig. 1. Discrete wavelet transform tree.

The paper is organized as follows. After the introduction, the basic definitions of discrete wavelet transform are given in Section II. Section III gives data overview and presents regression analysis for determining the functions of variation between three typical loads and average air temperature. Section IV describes the forecasting of power system load from air temperature. The experimental results are presented in Section V. Section VI concludes the paper.

II. DISCRETE WAVELET TRANSFORMATION

Discrete wavelet transform (DWT) decomposes a signal into a set of orthogonal components describing the signal variation across the scale [2]. The orthogonal components are generated by dilations and translations of a prototype function ψ, called mother wavelet:

$$\psi_{jk}(t) = 2^{-j/2}\psi(t/2^j - k), \quad k, j \in Z. \qquad (1)$$

The above equation means that the mother function is dilated by integer j and translated by integer k. A signal f for each discrete coordinate t can be presented as a sum of an approximation plus J details at the Jth decomposed level:

$$f(t) = \sum_k a_{Jk}\phi_{Jk}(t) + \sum_{j=1}^{J}\sum_k d_{jk}\psi_{jk}(t) \qquad (2)$$

where $\Phi_{Jk}(t)$ is scaling function. The residual term corresponds to a coarse approximation of $f(t)$ at resolution J. The coefficients a_{Jk} and d_{jk} are approximation wavelet coefficients at level J and detail wavelet coefficients (or wavelet coefficients) at level j, respectively:

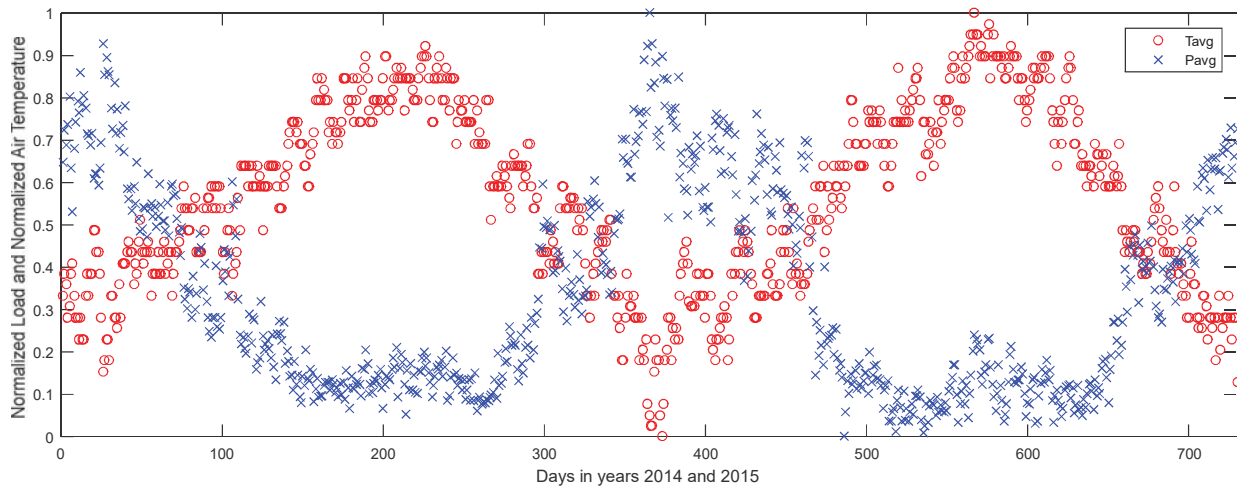$$a_{jk} = 2^{-j/2}\int f(t)\phi(2^{-j}t - k)dt, \qquad (3)$$

Fig. 2. Normalized daily loads and air temperatures for years 2014 and 2015.

$$d_{jk} = 2^{-j/2} \int f(t)\psi(2^{-j}t - k)dt . \qquad (4)$$

The estimate of d_{jk} and a_{Jk} can be achieved via iterative algorithm for decomposition using two complementary QMF filters h_0 (low-pass) and h_1 (high-pass) [3], as it is illustrated in Fig. 1. The relation between mother and scaling functions with the QMF bank is given with an efficient recursion:

$$\psi(t) = 2\sum_n h_1(k)\phi(2t - k) , \qquad (5)$$

$$\phi(t) = 2\sum_n h_0(k)\phi(2t - k) . \qquad (6)$$

III. Data Overview and Basic Statistical Analysis

Electric power load forecasting means to calculate the expected energy requirements of a system, which is important for making decisions including decisions on purchasing and generating electric power, load switching and infrastructure development. The power load forecasting relies on historical data to determine how much power customers may need. Forecasting model inputs can include day of the week, holiday calendars, weather conditions and forecasts, geographical differences, demographic information, etc. Accurate models for electric power load forecasting are essential to the operation and planning of a utility company [1].

In this paper, the case study dataset consists of hourly power system load data for Skopje for the calendar years 2014 and 2015 [4] and the corresponding meteorological information about minimal, average and maximal air temperatures obtained from the internet [5]. An analysis of the power load data shows that it depends on the time of the year, day of the week and hour of the day.

Fig. 2 presents normalized daily average power load diagram and normalized daily air temperatures diagram for Skopje across the years of 2014 and 2015, both normalized in the interval [0 1]. It can be noticed that there is a strong negative correlation between the power load and the air temperature. There are peaks at the beginning of the graphic (January 2014), the middle of the graphic (December 2014-January 2015) and at the end of the graphic (December 2015). The highest peak for the power load for the year of 2014 corresponds to the values Pavg = 1289MW and Tavg = −7C and it was registered on 31 Dec 2014. The highest peak for the power load for the year of 2015 corresponds to the values Pavg = 1247MW and Tavg = −8C registered on 2 Feb 2015. Fig. 2 confirms what is characteristic for this region that during the winter period the electrical energy consumption is bigger. The lowest average temperature in the graphic is Tavg = −9C registered on 8 Jan 2015, when the registered average power load was Pavg = 1228MW.

A regression analysis is performed over the dataset illustrated in Fig. 2 in order to estimate dependence curves of the three typical loads (minimal power system load Pmin, average power system load Pavg, maximal power system load Pmax) from the independent variable – the average temperature Tavg for the two years 2014 and 2015. The used approximation function is the sum of sinuses function of order $n = 4$ given by:

$$f(x) = \sum_{i=1}^{n} a_i \sin(b_i x + c_i). \qquad (7)$$

Fig. 3 shows the estimation of dependence curves of Pavg, Pmax and Pmin from Tavg, respectively, for the years 2014 and 2015. Table I summarizes the regression analyses presenting the equations coefficients, determination coefficients (R^2) and correlation coefficients. The determination coefficient shows the proportion of the variance in the dependent variable that is predictable from the independent variable (it ranges from 0 to 1, the coefficient 0 means the dependent variable cannot be predicted from the independent variable, while 1 means the dependent variable can be predicted without error from the independent variable) [6]. According to the Table I, the determination coefficients of the maximum, average and minimum daily load due to the
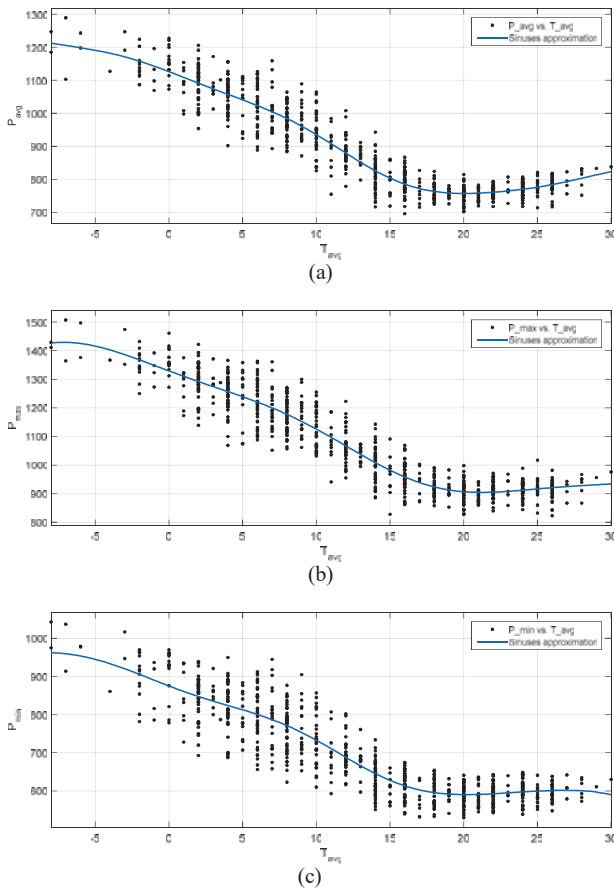
Fig. 3. Estimated approximation of power load from air temperatures by using the sum of sinuses function (7) of order 4: (a) averaged load, (b) maximal load, (c) minimal load.

TABLE I

SUMMARY OF THE APPROXIMATION FUNCTION (7) OF ORDER $n = 4$, PRESENTING THE FUNCTION COEFFICIENTS a_i, b_i, c_i, DETERMINATION COEFFICIENTS (R^2) AND CORRELATION COEFFICIENTS.

Averaged load:

Function coefficients:

$a_1 = 2235$	$b_1 = 0.05558$	$c_1 = 1.043$
$a_2 = 1355$	$b_2 = 0.08021$	$c_2 = 3.778$
$a_3 = 50.28$	$b_3 = 0.2274$	$c_3 = 0.01103$
$a_4 = 7.186$	$b_4 = 0.5079$	$c_4 = -3.128$

Goodness of fit:

R^2: 0.8961,   Correlation coefficient: -0.9466.

Maximal load:

Function coefficients:

$a_1 = 3193$	$b_1 = 0.0472$	$c_1 = 0.6216$
$a_2 = 2278$	$b_2 = 0.06366$	$c_2 = 3.367$
$a_3 = 19.29$	$b_3 = 0.3514$	$c_3 = -1.542$
$a_4 = -1.558$	$b_4 = 0.4358$	$c_4 = 0.557$

Goodness of fit:

R^2: 0.8874,   Correlation coefficient: -0.9420.

Minimal load:

Function coefficients:

$a_1 = 2719$	$b_1 = 0.06857$	$c_1 = 0.9005$
$a_2 = 2023$	$b_2 = 0.08394$	$c_2 = 3.797$
$a_3 = 30.43$	$b_3 = 0.2781$	$c_3 = -0.7307$
$a_4 = 5.246$	$b_4 = 0.4954$	$c_4 = -2.71$

Goodness of fit:

R^2: 0.8223,   Correlation coefficient: -0.9068.

average daily temperature are 88.74%, 89.61% and 82.23%, respectively. The regression analysis shows high prediction degree of the daily typical loads from the air temperature.

The correlation coefficient is a statistical measure that calculates the strength of the relationship between two variables [6]. Coefficients of correlation can have values in range from −1 (negative relation) to 1 (positive relation). There is not a significant relation if the correlation coefficient is less than 0.3. The correlation is with practical importance when the correlation coefficient is between 0.5 and 0.7. Correlation coefficient between 0.7 and 0.9 shows close correlation, while correlation coefficient greater than 0.9 shows very close correlation. According to the results presented in Table I, the correlation coefficients have values in a range from −0.90 to −0.95 what implies very close negative relation between all the combinations of typical daily loads and air temperatures.

The determination and correlation coefficients presented in Table 1 are higher than corresponding coefficients obtained when polynomial functions are used [8].

IV. POWER SYSTEM LOAD FORECASTING

The main idea is to forecast electrical power system load by analysing and processing the dataset of temperatures and corresponding power loads from Section III by applying a combination of the best fitting approximation in the least squares sense (Section IV) and the discrete wavelet transform (Section II). First, from the given dataset, the average air temperatures and the average power loads are used to estimate the best fitting function (7). This function is applied over a set of forecasted air temperatures in order to obtain the corresponding power loads. However, it should be kept in mind that the power system load of a specific day depends on not only the current air temperature, but also the weather conditions in the previous days affect the current power load. Therefore, the output (power load) from the best fitting function (7) should be related to the previous values of the air temperatures/power loads. For this reason, the wavelet transform at level (j) is applied over the output values from the function and as a result, approximation coefficients at level (j) (low-pass version) and detail coefficients at levels (1), (2), …, (j) (high-pass versions) are obtained. Finally, the inverse wavelet transform is applied over the approximation coefficients at level (j). The result is predicted power load. The procedure is illustrated with the block diagram in Fig. 4.
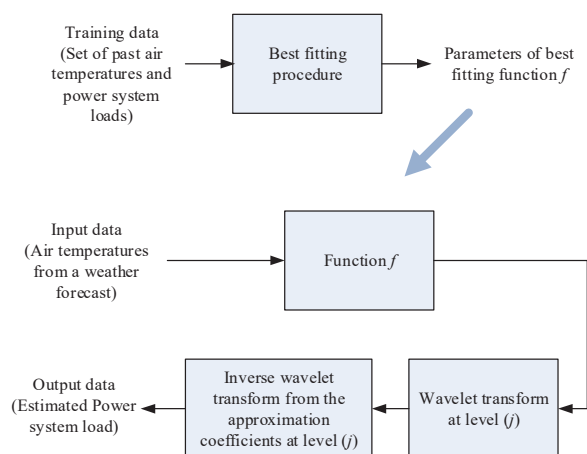
Figure 4. Block diagram of estimation of power system load.

TABLE II
WEATHER FORECAST FOR THE PERIOD 01-21 MAR. 2019.

1 Mar	2 Mar	3 Mar	4 Mar	5 Mar	6 Mar	7 Mar
3^0C	4^0C	11^0C	10^0C	16^0C	18^0C	14^0C
8 Mar	9 Mar	10 Mar	11 Mar	12 Mar	13 Mar	14 Mar
12^0C	17^0C	18^0C	20^0C	22^0C	19^0C	22^0C
15 Mar	16 Mar	17 Mar	18 Mar	19 Mar	20 Mar	21 Mar
22^0C	20^0C	6^0C	12^0C	8^0C	14^0C	19^0C

TABLE III
ESTIMATED POWER LOAD IN [MW] FOR THE PERIOD 01-21 MAR. 2019.

1 Mar	2 Mar	3 Mar	4 Mar	5 Mar	6 Mar	7 Mar
1061.7	1060.9	955.6	878.3	829.0	772.2	803.2
8 Mar	9 Mar	10 Mar	11 Mar	12 Mar	13 Mar	14 Mar
810.7	794.7	784.9	770.9	758.0	746.3	734.2
15 Mar	16 Mar	17 Mar	18 Mar	19 Mar	20 Mar	21 Mar
816.7	873.9	905.7	944.4	854.3	798.6	777.5

Fig. 5. (a) Weather forecast (air temperatures); (b) Forecasted average power load for period Mar 01-21 2019.

V. EXPERIMENTAL RESULTS

This Section presents the results obtained by experiments performed over the average power load hourly values for Skopje for years 2014 and 2015 that are used as training data for the model parameters optimization. The air temperatures for the period March 01–21 2019 (Table II, Fig. 5) from the weather forecast [7] serve as test data set for model performance assessment. A number of experiments for approximating the power load are made with best fitting polynomial functions of order 3, 4, 5 [8] and functions defined as sum of sinuses (7) of order 3, 4, 5 over the data of air temperatures and power loads for the years of 2014 and 2015. The results of these experiments show that the approximations with the sums of sinuses outperform the polynomial approximations in sense of higher determination coefficients R^2 and higher correlation coefficients. In addition, the experiments show that the order $n = 4$ for the function (7) gives satisfactory results (Fig. 3, Table I). The result of applying the proposed algorithm over the air temperatures from Table II is illustrated in Fig. 5. The squares denote the power loads as output from the function (7), while the solid line depicts the forecasted loads after applying the wavelet transform in two levels with db2 function used (Table III).

VI. CONCLUSION

Power load forecasting is a key factor in the functioning of power systems. It would determine which power units should increase their production and which generators should be dispatched. This paper analyses the relation between the power load and the air temperature in Republic of Macedonia. The correlation coefficients imply very strong negative relation between all combinations of typical daily loads and air temperatures. Regression analysis shows high prediction degree of daily typical loads from air temperature.

Power load forecasting is proposed as a combination of best fitting approximation and wavelet transform. A better load prediction can be achieved if more years are incorporated in estimating the approximation function.

REFERENCES

[1] E.A Feinberg, D Genethliou, "Chapter 12 Load forecasting", Applied Mathematics for Power Systems, pp.269-282.
[2] G.Strang, T.Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambrige Press, 1996.
[3] P.P.Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice – Hall (1992).
[4] D.Bajs, M.Atanasovski, *Longterm Forecast Study of Electrical Energy and Power Balance and Adequacy Analysis of Transmission Network of Republic of Macedonia*, Zagreb/Skopje EIHP, 2016.
[5] https://www.wunderground.com/history
[6] S.Vukadinovic, *Elements of Probability Theory and Mathematical Statistics,* Belgrade, 1973.
[7] https://weather.com.
[8] M.Atanasovski, M.Kostov, B.Arapinoski, I.Andreevski "Correlation between Power System Load and Air Temperature in R. Macedonia", *Int. Scientific Conf. on Information, Communication and Energy Systems and Technologies*, Sozopol, Bulgaria, Jun. 2018.

Impact of the Power System Stabilizer on Transient Stability of the Power System

Blagoja Stevanoski[1] and Natasa Mojsoska[2]

Abstract – **Stable operation of the power system is one of the key factors for reliable and quality power supply. Major disruptions in the system operation mode (short circuits, outages of large generating units, transmission lines, etc.) cause significant changes in the parameters of state of the electricity system. The response of the system in such cases is an emergence of electromechanical oscillations in synchronous generators reflected in the fluctuation of the regime parameters (rotational speed, active and reactive power, voltage, power output, etc.). Oscillations can reach an amount that can compromise the stable operation of the synchronous generator and the power system in general. One way to suppress oscillations is to use a stabilizer of the power system as an integral part of the excitation systems of generators. The task of the power oscillation stabilizer is to produce a torque damping component of the electromagnetic torque through the excitation systems. This paper presents the theoretical basics of contemporary power system stabilizers, and on concrete example analyzes its impact on transient stability in a case of a close short-circuit near TPP Bitola.**

Keywords – **Synchronous generator, excitation system, power system stabilizer, transient stability.**

I. INTRODUCTION

The power system is a complex non-linear system that is constantly exposed to various types of disruptions (load changes, outputs of production units, short-circuit, changes in a topology of the network, etc.). Stability of a power system is its ability to return to a running equilibrium state after an occurrence of a disturbance, while the regime parameters of the system remain within the limits that provide a complete integrity of the power system. Important variables at power system equilibrium are rotor (power) angle, nodal voltages and frequency. Rotor angle stability is the ability of interconnected synchronous machines to remain in synchronism in case of transient disturbance [1].

The stability problem involves the study of electromechanical oscillations inherent in power systems. Electromechanical oscillations are a consequence of the physical nature of a synchronous generator connected to a power system. Such a system contains multiple energy accumulators (rotating masses, inductive excitation coils, etc.) which react with electromechanical oscillations to the smallest deviation from the equilibrium state. Oscillations are superposed on stationary variables and in the worst case can endanger the stability of the synchronous generator.

Electromechanical oscillations can be divided into local and systemic oscillations. The connection of new generating units to the power system through relatively long transmission lines creates favorable conditions for the occurrence of local electromechanical oscillations of the generators in relation to the power system. When connecting and unifying smaller power systems at the regional level in a single state system or connecting multiple systems, the occurrence of electromechanical oscillations is much more complicated and falls within the category of systemic oscillations. Such oscillations are a consequence of various factors: relatively long transmission lines between the power system, the current state and system configuration (character and location of consumers and production units, occurrence of various interruptions and outages of generators, inclusion and exclusion of large consumers, poor synchronization with network connectivity, etc.).

Due to competitive energy market, in order to transmit as much power as possible, the energy system is used up to its limits of stability. This leads sometimes to stability problems like power oscillations. In the synchronous generator, the damping field and damper windings provide to the rotor oscillations is weakened due to excitation control system action. The reason for this is the appearance of additional currents in the rotor circuits induced by the voltage regulation which oppose the currents induced by the rotor speed deviations [2]. Therefore, an additional stabilizing signal is needed and the device used for this purpose is known as the power system stabilizer (PSS). This is one of the most cost-effective methods of enhancing power system stability. PSS uses auxiliary stabilizing signals to control the excitation system. Based on the operation of the automatic voltage regulator (AVR) commonly used input signals to the power system stabilizer are shaft speed, power or terminal frequency. The stabilizer introduces an additional electric torque corresponding to the deviation of the speed by increasing the damping of system oscillations and improving the stability of the power system.

II. EXCITATION SYSTEMS AND POWER SYSTEM STABILIZER

Excitation systems are one of the most important parts of the synchronous generators that are intended to provide direct current to the generator field winding. Additionally, excitation systems are also responsible for control and protection functions of the power system. Its dynamic performance has a direct impact on generator stability and reliability [1], [2].

[1]Blagoja Stevanoski is with the Faculty of Technical Sciences, Makedonska falanga 33, Bitola 7000, Macedonia, E-mail: blagoj.stevanovski@uklo.edu.mk.

[2]Natasa Mojsoska is with the Faculty of Technical Sciences, Makedonska falanga 33, Bitola 7000, Macedonia, E-mail: natasa.mojsoska@uklo.edu.mk.

Fig. 2.1 shows the general structure of voltage regulator for synchronous generator, which includes a circle for regulating the excitation current and a predetermined circle for regulating the voltage of the generator. Reactive power regulator and power system stabilizer is used according to the request of the power system. Terminal voltage controller is proportional-integral (PI) type and is superior to excitation current controller which is proportional (P) type. Output of voltage regulator is a reference value of excitation current. Based on measured values of terminal voltage and currents, active and reactive power of synchronous generator are determined [3],[4].

A power system stabilizer is a device which provides additional supplementary control loops to the automatic voltage regulator system of a generator unit. Task of the power system stabilizer is to generate a stabilizing signal which creates a damping component of electromagnetic torque during transient process, acting through the excitation system. The output signal from the stabilizer is introduced into the summator before the voltage regulator. The stabilizer has to perform phase compensation between an input of excitation system and electromagnetic torque. The usual input signal used in classical stabilizers is the active power of the generator or the frequency at the place of the connection of the generator to the power system. The damp component of the torque is in phase with the change in a rotational speed of the generator. On that way, it can be used as an additional stabilization signal [1].
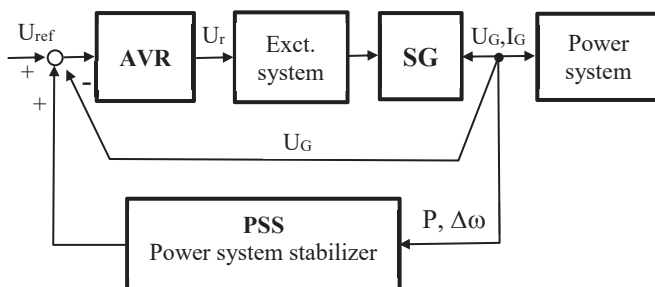
Fig. 2.1 Conventional structure of the system for managing the excitement of the synchronous generator

Power Stabilizer (PSS) is a feedback link controller that is part of the synchronous generator control system. Its main function is to damp the oscillations of the rotor of the generator in an interval of about $0,1 \div 2.5$ Hz as if electromechanical oscillations. For this purpose, PSS should produce such a component of an electromagnetic torque that is opposite to the mechanical torque. Such a component should be in phase with the deviation of the rotor speed.

The input signal in the PSS is measured at the generator connection and determines the type of controller structure. Different types of stabilizing circuits are designed. In Fig. 2.1 shows the structure of the dual input power stabilizer known as PSS2B (according to the IEEE standard 421.5.2005). PSS2B uses the deviation of the speed and the active power to determine the stabilizing signal.

In Fig. 2.2 shows the structure of the PSS1A type stabilizer (according to IEEE). A washout filter with a time constant T_f allows crossing the signal of the active power deviation of the generator without changes. In this way, a change of amount of

active power does not affect the voltage of the generator. This block allows the stabilizer to operate only in case of transient changes in the active power signal. The value of the appropriate time constant should be large enough so that the signal of a deviation can be crossing without changing the frequency range of interest. In the case of a local oscillation of the generator in a frequency range of $1 \div 3$ Hz, the time constant of the derivative block T_f in the amount of 1s may be satisfactory, while in the case of interconnection in a frequency range of $0,1 \div 0,6$ Hz. The desired value of the time constant T_f is about 10s.

The amplification K_{stab} determines the size of the damping that is introduced by acting on the stabilizer. Damping of the oscillation is increased by increasing the amplification to a certain limit when the increase in amplification causes a decrease in the oscillation of the generator [5]. The phase compensation block compensates the phase lag between the input in the excitation system and the electromagnetic torque. To achieve the appropriate phase compensation, it is possible to use two or three blocks of first-order or a second-order block with complex solutions. The output signal from the stabilizer has a positive (V_{smax}) and negative (V_{smin}) limit. The positive limit can be set (according to the IEEE) to the amount of 0,2 p.u. in order to ensure the efficiency of the stabilizer during large scales, and the negative limit set to the amount of 0,1 p.u. is considered satisfactory. This ensures sufficient bandwidth of the stabilizer.

Fig. 2.2 Structure of the stabilizer type PSS1A (according to IEEE)
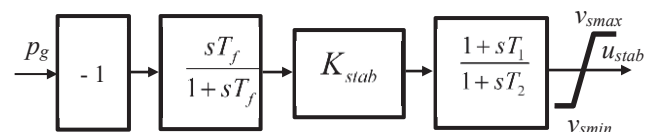
From the stabilizer is required damping of oscillations that are of a local character and an inter-area character. Depending on a type of oscillation, the stabilizer performs phase compensation in a certain frequency range. The phase characteristic to be compensated, changes with the change in the state of the system, so the setting of the stabilizer for one condition does not mean a satisfactory setup for another state of the system.

III. CASE STUDY

The current characteristic of Macedonian electric power system is domination of thermal power plants (TPP), which produce about 80% of total electricity demand. TPP Bitola is the biggest and most essential in the country. It has three units with installed capacity of 233 MW per unit or total 700 MW. This potential allows the plant to participate with more than 70% in total electricity production in the Macedonian power system. TPP Bitola is linked to the power system through substation 400/110 kV Bitola 2 that is the most important in Macedonia.

Impact of a power system stabilizer is examined on synchronous generator installed in unit "Bitola 1". Generator's main parameters under observation are excitation voltage (E_{fd}), active power (P) and terminal voltage (U). During the test three

phase short circuit and transmission line outage are simulated. The generator and the network are modelled in Neplan 5.5.3 software.

A. Power system stabilizer modelling

Static excitation system DIREMK - Koncar is installed at the generators in TPP Bitola [6]. Control of the excitation system is based on microprocessor twin-channel voltage regulator. Fig. 3.1 is a block diagram of the stabilizer of electromechanical oscillation which is an integral part of the voltage digital controller. The stabilizer corresponds to the type PSS2B (according to the IEEE Std. 421.5 classification). A combined stabilizer is used in two quantities: active power and frequency. The stabilizer has the same standard transmission function for both channels, but with different parameters:

$$F(p) = K_{ss} \frac{pT_D}{(1 + pT_F)(1 + pT_D)} \qquad (1)$$

where:

K_{ss} – gain the stabilizer per channel,
T_F – time constant of the stabilizer (filter)
T_D - time constant of the stabilizer (derivative)
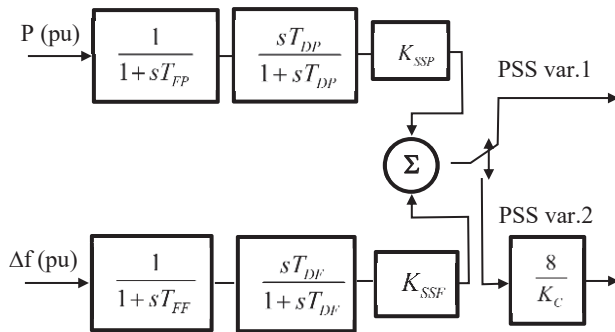
Fig. 3.1 Block scheme of the power system stabilizer (Koncar)

The input signals (active power and frequency) are filtered with given time constants. The filtered signals are further multiplied according to the amplification of the individual channels depending on whether the signals are entered in the summator before the PI - regulator (index 1) or after it (index 2). The output of the stabilizer is an output as PSS1 in the case when the stabilizer signal is kept in a summator before the voltage regulator and as PSS2 when the stabilizer signal is added to the output signal from the voltage regulator. With the help of logical parameters, the active power channel, or the channel in frequency, can be separately activated.

B. Excitation System Response to a Short Circuit with and without PSS

This test was performed in order to show excitation system reaction on large network disturbance. Three phase short circuit at 110 kV bus bar of SS Bitola 2 is simulated. Short circuit appears on 0,1s, duration of the test is 5s and short circuit time is chosen 0,2s as typical short circuit length in power system. Fig. 3.2 shows excitation voltage changes, terminal voltage and

active power fluctuation without PSS and Fig. 3.3 shows the same parameters with PSS as a part of the excitation system.
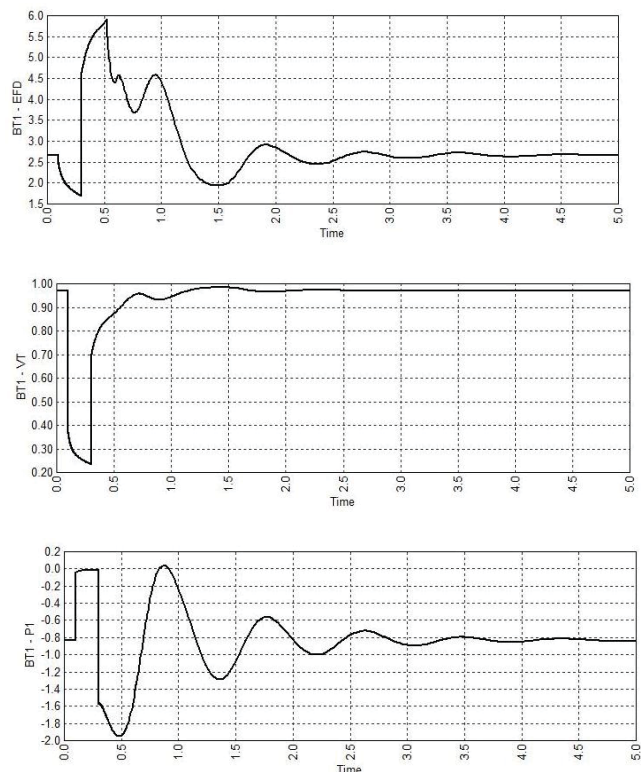
Fig. 3.2 Excitation voltage, terminal voltage and active power (PSS - off)
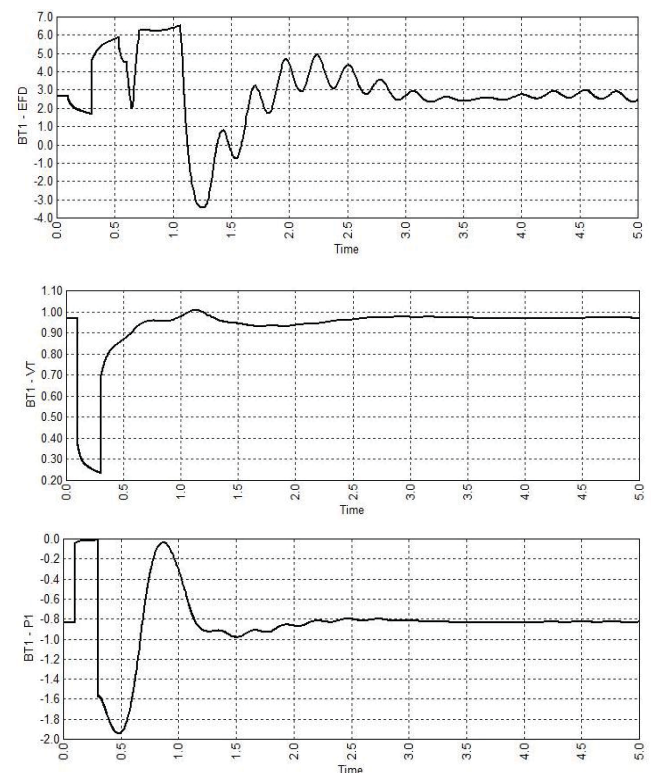
Fig. 3. 3 Excitation voltage, terminal voltage and active power (PSS - on)

C. Excitation System Response to a Transmission Line Outage with and without PSS

This test examines excitation system response during transmission 400 kV line outage. The line connects TPP Bitola with the main consumption center, the capital city Skopje. Time duration of line outage is 0,3s simulating auto reclosing of transmission line. Fig. 3.4 and 3.5 show system parameters changes without and with PSS.
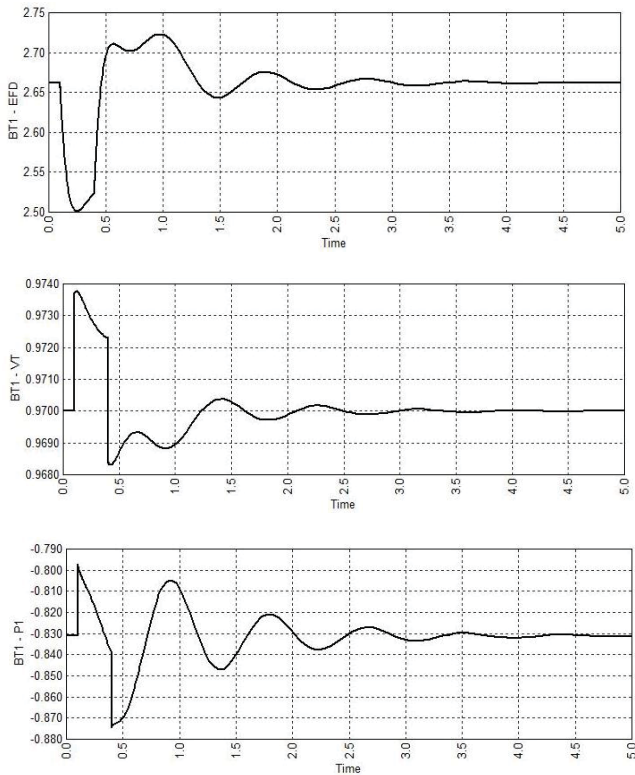
Fig. 3.4 Excitation voltage, terminal voltage and active power
(PSS - off)
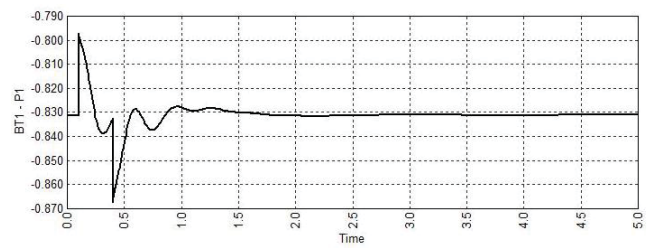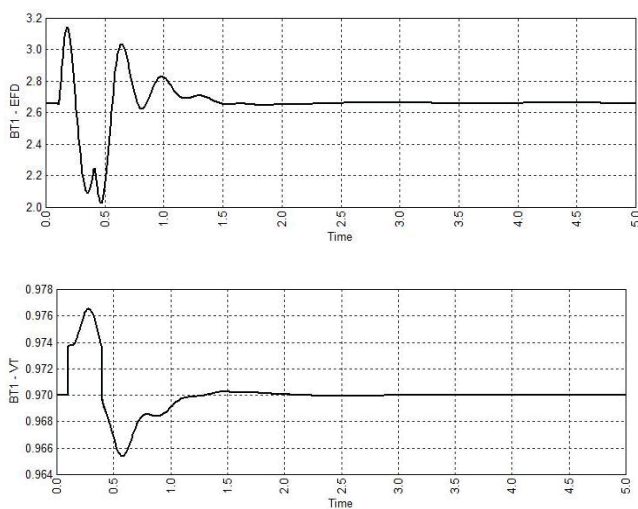
Fig. 3.5 Excitation voltage, terminal voltage and active power
(PSS - on)

IV. CONCLUSION

The aim of the work was to investigate behavior of synchronous generator excitation system at TPP Bitola in different conditions and impact of its power system stabilizer on excitation system response. The test was performed with real system parameters of Macedonian power system modelled by Neplan 5.5.3 software. Based on the performed analysis, and the simulations results it is possible to conclude that a stabilizer of electromechanical oscillations can reduce the oscillation, especially active power fluctuation, in given operation range. To provide effective damping and ensure the stability, the PSS should be properly chosen and carefully tuned.

REFERENCES

[1] P. Kundur "*Power System Stability and Control*", McGraw-Hill, 1993.
[2] A. A. Zea "Power System Stabilizers for the Synchronous Generator", MSc thesis, Chalmers University of Technology, Sweden, 2013.
[3] D. Sumina, N. Bulić, S. Skok *"*Stabilization of the Electromechanical Oscillations of Synchronous Generator", Strojarstvo no. 53 (3). pp. 209 – 219, 2011.
[4] M.Miletić "Prigušenje elektromehaničkih nihanja i regulacija napona sinkronog generatora", kvalifikaciski doktorski ispit, Fakultet elektrotehnike i računarstva, Zagreb, 2008.
[5] M. Dušak, J. Ritonja "Delovanje stabilizatorjev elektro-mehanskih nihanj na agregatih savskih elektrarn", Cigre, Portorož, 2015.
[6] Končar, "Digitalni regulator napona tip DRN", Funkciski opis LA 8636, Zagreb, 2010.
[7] V. Slenduhhov, V. Kilter "Modeling and Analysis of the Synchronous Generators Excitation System", 13th International Symposium, Tallinn, Estonia, 2013.

Electric Energy Study Applied to Six Industrial SMES Companies of the City of Bogota

Omar López Delgado[1]

Abstract – **An energy efficiency study was made to six SMEs as a sample of different industries in Bogotá. The goal of this research is to implement the benefits of the energetical efficiency. The failures must be identified to propose the ideal solutions in the operation and production to minimize the energy consumption and reducing the CO_2 emissions to the atmosphere, if the electric energy is produced by a thermoelectric way. With this diagnosis is expected to present the technical failures found in the analysis to recommend general improvements, it will allow to begin the optimization in the use of the energy at the SMEs. This study can be translated to other enterprises in the same industrial sectors, contributing to the strategic plan of the national energy and mining investigation program.**

Keywords – **energy efficiency , energy study, research, industrial SMEs,**

I. Introduction

The new regulatory framework in Colombia for the development of industrial competitiveness, emphasis is placed on rational use and energy efficiency, as an instrument to encourage the optimization of energy consumption on industrial production facilities, so in their Operating and production processes reduce consumption and energy losses without reducing the quality of the service provided, thus reducing energy costs and the environmental impact generated by these processes.

Goals of the energy efficiency:

- **Energy savings**, to mitigate the increase in the demand and to reduce the use if resources like water used to in the hydraulic generation or fossil fuel which are CO_2 producers, in the thermic processes to generate electric energy.
- To implement improvements in the industrial operation and production processes, using the equipment upgrade to the new technologies.
- Continuity and quality in the energy supply. Which avoids economic losses in the industrial and commercial and industrial sectors.

Stages of the energy efficiency:

Measurement of the consumption to make an energetic efficiency diagnosis and find opportunities to improvement. Perform energy leak adjustment, upgrading or replacement the machinery a new technology. To optimize the consumption proposing improvements in the energy quality or automating processes in isolated equipment, engines and lighting, devices that allow to manage the use and consumption. Monitor consumption permanently.

[1] Escuela Tecnológica Instituto Técnico Central – ETITC, Facultad de Ingeniería Electromecánica, Bogotá D.C. - Colombia, E-mail: lopezdelgadoomar@yahoo.es

II. Developed procedure.

To obtain and record the energy diagnosis in the 6 sample companies, visits were made to each of the companies to record the information on cards incorporating administrative data, the range and sector of the production to which it belongs, production data , general consumptions of KWH (invoice of energy consumption), flow chart of the process, photographic records with the identification and general description of the equipment and survey of the single-line diagram.

As part of the energy characteristic in the plant, of each one of the industrial companies evaluated, the installation of the measurement equipment that allowed the validation of the information in relation to the electric power consumption was carried out. For this, a network analyzer (HIOKI 3196) was installed in the electrical distribution network, mainly in the totalizers of the distribution boards and in the equipment that was taken into account in the energy consumption in the plant.

The network diagnosis allows:
- Evaluate, Detect and prevent excess consumption (KWH)
- Visualize the maximum energy demand.
- Verify reactive power consumption to perform capacitor bank calculation.
- Observe voltage variations (overcurrent, overvoltage), current and frequency fluctuations.
- Detect harmonics in the network.
- Find issues in the electrical network such as disturbances, resonances, imbalances in the network, among others.

Thermographic records were made with a thermography camera (Camera FLIR T440), which allowed to determine temperature increases in different equipment and in the conductors of the main connections. For analysis purposes, it is taken as the electrical faults and criteria defined in the ANSI / NETA ATS / 2009 standard, this reference contains criteria that allow classifying, through levels, the actions to be taken based on the temperatures found in the inspected equipment.

The data obtained from the images of the thermography camera and the measurements of electrical parameters with the network analyzer, were studied to determine the electric power consumption and the points where there are issues of electrical type and wasting of electric energy are located, this allowed to quantify the waste of energy.

Based on the results obtained, comparisons were made and applied existing energy management models, to determine and implement the use of new technologies, and to record the results of electric energy savings and the optimization rate of this service in each one of the companies studied.

III. Registered data of companies.

The information related to the production sector, voltage levels, location and 'maximum and minimum consumption' were related to each company in the following tables. During the study, the time in which the equipment worked was also taken into account.

Table. 1 Companies data

Companies	Activity	Sector	Location
1	Plastic and rubber extrusion	Plastic	Puente Aranda
2	Clean and disinfection of metals	Chemical	Mártires
3	Production of leather articles	Leather	Usaquén
4	Food selection and marketing	Food	Suba
5	lodging and food	hotel	Chapinero
6	manufacture of medicines	Pharmaceutical	Puente Aranda

The cost of energy was approximated to millions of Colombian pesos, since this is the local currency. The analysis was performed on the network analyzer installed on the general board of the company's facilities and the data was compared with the energy meter installed by the network operator, the thermographic records were made on the same board and on the equipment with the highest energy consumption.

Table. 2 Companies consumption.
Tension level 1:(220/128-208/120)
Tension level 2:(11.4KV)

Companies	Tension level	Average daily Consumption (KW/day)	Maximum daily consumption	Average monthly energy	Monthly energy value (COP)	Schedule of activities. (D-day)
1	2	32.7	81.2 / 59.7	784.65	9.226M	D
2	2	23	23 / 27	298	1.844M	D
3	2	35	79 / 45.3	861.35	11.300M	D
4	2	4.72	12.71/ 12.82	32.76	0.13 M	N
5	1	28.92	47.38/75.47	176.5	0.706M	D-N
6	1	11.2	14.8/ 8.9	440	5.65 M	D

IV. Record of electrical failures and waste of energy.

Company 1:

- Presents current imbalances between phases +/- 10% in angle and magnitude.
- Lack of maintenance in the distribution boards and adjustments in terminal connections presenting high temperatures and electrical risk due to the presence of arcs and short circuits that lead to fires.
- Low power factors recorded when working at night time = 0.33.
- The electrical resistance used in the thermosealing and press injection equipment presents high temperatures.

- High currents circulating on the neutral line of the electrical connection of the equipment.

Company 2:

-The electrical system presents voltage variations outside the limit plus 110%, presenting consequences for the operation of existing electrical equipment, shortening the useful life.

- There are current imbalances above 5% in each of the phases its magnitude and angle.

- It does not have a capacitor bank, it does not invoice reactants to the network operator.

- Presents harmonic distortion of tensions, Phase A Max 5-89% minimum 0-95%, Phase B Max 6-25% minimum 1-15%, and Phase C Max 6 68% minimum 0-12%.

- It presents voltage drop in the feeders of the motors, generating increases in current, overheating in the conductors and reducing the efficiency of the motor.

Company 3:

- It presents low tension in the boards of feeding equipment.

- It presents current imbalance in the phases, which produces increase of current through the neutral, faults in the system and equipment, overheating of conductors and lower performance in the system.

Company 4:
-It presents high temperatures in semiautomatic thermoforming machines of the order of 32 to 62 ºC, especially in electrical resistances.

- There is current unbalance in the phases between 12% and 52%, currents greater +/- 10% in angle and magnitude.

- The power factor is found at 0.82.

- THDV harmonic distortion values exist: Phase A 5.23% max, 1.94% min, Phase B 5.26% max, 2.79% min, and Phase C 5.83% max 2.76% min.

Company 5:

The records were made to systems that are connected to a general distribution board fed from a 225 KVA substation.
   Air Conditioning System
- It has low power factors between 0.86 and 0.79, even though it has a capacitor bank.

- Voltage drops are recorded between 94.99 V and 20.84% with a duration of one minute.

- There are voltage lags of 4.24%.
   Food Refrigeration and Freezing System
- Presents low power factors between 0.8 and 0.76, this results in higher current consumption, losses in the form of heat in the conductors, higher consumption therefore higher cost of energy payment and low efficiency of the system.

- It has current phase shifts of 7% and 15%, which causes high current to flow through neutral.

- Voltage harmonic distortions (THDV) are present in Phase A of 5.6% max, and 1.58% min.

Company 6:

This company is connected in medium voltage, Level 2, the power measurement is in low voltage without reactive meter, its power is made from a 225 KVA transformer installed in a structure in type H poles and shares the service with three companies more than sector.

- The maximum current recorded in 40 minutes was 289.4 A of 624.53 A given by the transformer, which is equivalent to 46% of the total load of the transformer.

- It presents a total distortion of the demand in current (TDD) of 12% out of the norm.

- The power factor is 0.75 on average, without a capacitor bank.

- The equipment used in their processes are 43% deficient evaluated in the thermodynamic analysis, 38.6 ºC in the main connection, 29.2 ºC in the main protection (3 * 100 A), and 33ºC in the connection conductors, to give a example. Which suggests electrical problems from the main connection.

- The motors of the machines and tools of the shop do not have a motor guard or starter, which in its startup produce high currents and transients of voltage and intensity, overloading the respective connections.

V. Corrective and Suggested Improvements.

Company 1:

- Balance the loads of the equipment and the system in general and the phase currents, to minimize the current flowing through the neutral.

- Change the work routine to night hours to save energy.

- Change voltage level from 1 to 2 to reduce energy tariff.

- Check and adjust the existing capacitor bank, to correct the power factor.

- Change the electrical resistances of injection and thermo-formed equipment due to obsolescence by current technologies.

- Carry out modernization and / or conversion of equipment such as presses, injectors and heat sealers for obsolescence.

- Promote the sequential automation of some processes.

- Replace fluorescent lighting with LED technology, take advantage of natural light.

Company 2:

- Inform the operator of the existence of harmonic distortion, or otherwise install electronic equipment (Install harmonic filters) to correct these distortions.

- Install a capacitor bank to correct the power factor.

Company 3:

- Inform the network operator and / or correct the low voltage on the equipment power boards.

- Balance the current in the phases from the main board.

Company 4:

-Change by current technologies the electrical resistance of the thermoforming machines due to obsolescence.

- Carry out system load balancing.

- Install a capacitor bank.

- Install harmonic filters which allows to improve the presence of harmonics in the system and increase the useful life of the equipment, efficiency and improve its performance.

- Change the electrical resistances of the sealer and thermosealer by resistance of current technology.

- Install temperature controllers in the equipment of the sealer and thermosealer.

- Perform thermal insulation in the sealer and heat seal reducing the electricity consumption.

- Install a soft starter or variable speed drive as the case to avoid excessive consumption.

- Check the air circuit and correct leaks in the compressed air duct and include control valves at the terminals of the water pipe.

- Ensure that the air intake to the compressor is outdoors, avoiding the heating of these equipment subjected to high temperatures and little free space.

- Report to the network operator the voltage variations and the harmonic distortion in the electrical system so that they take corrective measures.

- Install an AVR voltage regulator to the boards that present voltage variations.

- Carry out preventive maintenance, adjust connections and terminals of the protections of the motor and starter boards.

- Carry out measurements of the earthing system and visual verification to verify connections of outlets to the ground.

Company 5:

- Adjust the existing capacitor bank in the general board or install a bank of capacitors in the boards of each of the air conditioning and refrigeration and freezing systems respectively.

- Report to the network operator the voltage variations and the harmonic distortion in the electrical system so that they take corrective and / or install harmonic filters which allows to improve the presence of harmonics in the system and increase the useful life of the equipment, effectiveness and improve their performance.

- Install an AVR voltage regulator to the boards that present voltage variations.

- Perform equipment modernization and / or conversion due to obsolescence.

- Perform sequential automation of processes in air conditioning systems and refrigeration systems.

Company 6:

- It is necessary to correct F.P with the installation in the general board, of a capacitor bank of 808.1 Pf.

- Make the change of the mechanized T3 distribution board.

- Change the connections of the T2, T3 mechanized boards, lighting board and workshop outlets, the lighting board and office outlets.

- It is suggested to change the voltage level from low to medium voltage, Level 1 to level 2, which would require the installation of a 75 KVA transformer.

- Install saves motors and soft starters to the equipment of machines and tools of the workshop.

VI. Conclusion.

Thermographic records made by the six companies show that the equipment used in the production and operation processes have high temperatures due to lack of maintenance, especially due to the adjustment of the electrical connections, which leads to electrical faults that are removed from operation the equipment and interrupt the processes and in other of these companies the lack of adequate isolation in thermal systems, carries with it the energy losses in the form of heat transfer and greater electrical consumption in the equipment. If an adequate electrical maintenance is carried out and the use of thermal insulation in the equipment that requires heat production in its processes, the consumption of energy will be significantly reduced.

To each of the 6 companies evaluated, the reports of the studies carried out on energy efficiency were sent, each of them is expected to implement and execute the suggested proposals to optimize the use of electric power and that these results also serve as a basis for the other SMEs implement energy management aimed at the efficient use of energy and its saving so that they improve and increase the operation and production processes. The benefits of energy efficiency applying as an additional model to industrial production, were determined a significant reduction in production costs. Moreover, a better control of tension, current and temperature levels, a better functioning can be seen in production machinery. It may reduce the maintenance costs, and possible losses due to machinery replacement. Finally, must be mentioned the environmental benefits of the energetical efficiency. Since a lower consumption means a better use for the produced energy, thus , if this model generalizes, can be possible a big city supplied with less quantity of energy, giving a place to clean energy production away from fossil fuels.

REFERENCES

[1] Ministerio de Minas y Energía, *Programa de uso racional y eficiente de energía y fuentes no convencionales*, *MinMinas*, [Online] may 2010, [aforementioned, December 2014].

[2] Álvarez Carlos, Serna Francisco, Normatividad sobre eficiencia energética y edificaciones verdes, Cidet, [Online] nov 2012.

[3] Kenneth Gillingham,[1] Richard G. Newell, Energy Efficiency Economics and Policy, annual reviews, [Online] 2009.

[4] Lorna A. Greening, David L. Greene, Energy Efficiency and Consumption, Energy policy, Elsevier [Online] 2000.

[5] Adam B. Jaffe, Robert N. Stavins, The energy-efficiency gap What does it mean? Energy policy, Elsevier,[Online] october 1994.

[6] Abdeen Mustafa Omer, Energy, environment and sustainable development, Renewable and Sustainable Energy Reviews, Elsevier [Online] December 2008.

Double Fed Induction Generator turbine in the Power System of Macedonia

Mile Joncevski[1], Mile Spirovski[2]

Abstract – This paper deals with dynamic characteristics of the first DFIG wind farm which will be connected to the transmission power system of Republic of Macedonia. MATLAB/SIMULINK software is used to present the dynamic characteristics of the wind farm. The DFIG wind turbine and the power system are modeled with real characteristics, like in the initial project. Results from tests short circuit on 110 kV voltage level buses and effects of immediate wind speed change will be simulated in order to define how the DFIG will react under these circumstances and which actions should be undertaken in the power system to improve power system stability and power quality.

Keywords- DFIG, wind turbine, dynamic characteristics, Macedonian Power System

I.   INTRODUCTION

The Wind power has widely proved to be one of the most competitive and efficient renewable energy sources with a most favorable technical and economic prospects. This is due to the existence of non-exploited wind resources and to the fact it is a clean energy with a reduced cost of installation and maintenance. Wind turbines, which typically are centralized in wind farms, are constantly planned and commissioned. The produced electrical power from wind is steadily increasing. As a consequence, wind power has reached significant influence on the power production and penetration levels imposing new challenges to the Transmission System Operators (TSO). The process of high wind energy penetration requires the impact analysis of this new technology in power systems. Impact of wind energy on power systems is related to security, stability, power quality and operation of power system.

Among the several wind energy technologies, a generation of variable speed wind turbines present many advantages compare to the fixed speed wind turbines. These wind energy conversion systems are connected to the grid through voltage source converters to make variable-speed operation possible. The doubly fed induction generator (DFIG) wind turbines are nowadays more widely used especially in large wind farms. The main reason for the popularity of the DFIG is their competitve cost and performance and ability to supply power at constant voltage and frequency while rotor speed varies. Following the fact that most country intend to use renewable energy sources in their power systems, our case study came as

[1]Mile Jonchevski, ESM 11. October 9, 1000 Skopje
[2]Mile Spirovski, Faculty of Technical sciences at University of Bitola, Makedonska Falanga 33, Bitola 7000, Macedonia.

result of the initial project of implementing wind power energy into the electrical power system of Macedonia [11]. In the initial project, authorities have planned to implement DFIG because of the fact that it will drastically decrease the costs for invertors, it will enable control of the torque and as last, it will increase the efficiency of wind extraction. An electromagnetic transient model is developed on the widely used simulation platform MATLAB/SIMULINK utilizing its SimPowerSystem toolbox. The model is applied for a study of a grid-connected wind farm interacting with varying wind speed and grid fault. Because the turbine is a set of many sub models that are working as one, proper explanation for every sub model will be given. Thanks to the unlimited possibilities of MATLAB, modeling the wind turbine was easy.

II. WIND TURBINE MODEL DESCRIPTION

Several generator types are in use for wind power applications today. The main distinction can be made between fixed speed and variable speed wind generator types. The most widely used variable speed wind generator concepts are double fed induction generator.

The DFIG is a wound rotor induction generator with a voltage source converter connected to the slip rings of the rotor [9]. The stator winding is coupled directly to the grid and the rotor winding is connected to the grid via AC/DC/AC converter. Wind turbine is coupled to the induction generator through a mechanical shaft system, which consists of a low-speed and high-speed shafts and gearbox between. In order to produce electrical power at constant voltage and frequency to the utility grid over a wide operation range, the power flow between the rotor winding and the grid must be controlled both in magnitude and in direction. Therefore, there are two IGBT PWM converters: a rotor side converter (RSC) and a grid side converter (GSC) connected back-to-back by a dc link capacitor. The crowbar circuit is used to short circuit the rotor side convertor in order to protect it from over current in the rotor during transient disturbances.

The operation of the DFIG wind turbine is regulated by a control system, which generally consists of two parts: the electrical control of the DFIG and the mechanical control of the wind turbine blade pitch angle. Control of the DFIG is achieved by controlling the RSC and GSC, as shown in Fig. 1.
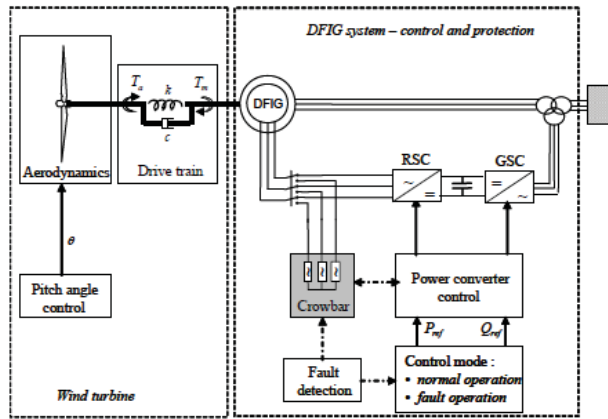
Fig. 1. DFIG wind turbine configuration and control

III. MODEL OF THE GENERATOR

It would be of interest to describe the basics of the DFIG according to electrical machines theory. The generator is modeled according the dq component system [6]. At start, a brief mathematical explanation is provided. The mathematical model needs mathematical constants according equivalent circuit presented on Fig. 2.
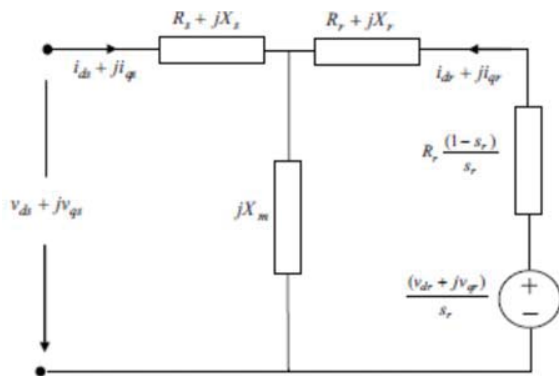
Fig. 2 Equivalent circuit of DFIG

$$E_d^{'} = -\frac{\omega_s L_m}{L_{rr}}\psi_{qr} \qquad (1)$$

$$E_q^{'} = \frac{\omega_s L_m}{L_{rr}}\psi_{dr} \qquad (2)$$

$$X_s = \omega_s L_{ss} \qquad (3)$$

$$X_s^{'} = \omega_s(L_{ss} - \frac{L_m^2}{L_{rr}}), \quad T_0^{'} = \frac{L_{rr}}{R_r} \qquad (4)$$

in order to explain the model properly.

$$\frac{X_s^{'}}{\omega_s}\frac{di_{qs}}{dt} = v_{qs} - [R_s + \frac{1}{\omega_s T_0^{'}}(X_s - X_s^{'})]i_{qs} - (1-s_r)E_q^{'}$$
$$-\frac{L_m}{L_{rr}}v_{qr} - \frac{1}{\omega_s T_0^{'}}E_d^{'} - X_s^{'}i_{ds} \qquad (5)$$

$$\frac{X_s^{'}}{\omega_s}\frac{di_{ds}}{dt} = v_{ds} - [R_s + \frac{1}{\omega_s T_0^{'}}(X_s - X_s^{'})]i_{ds} - (1-s_r)E_d^{'}$$
$$-\frac{L_m}{L_{rr}}v_{dr} - \frac{1}{\omega_s T_0^{'}}E_q^{'} - X_s^{'}i_{qs} \qquad (6)$$

$$\frac{dE_d^{'}}{dt} = s_r\omega_s E_q^{'} - \omega_s\frac{L_m}{L_{rr}}v_{qr} - \frac{1}{T_0^{'}}[E_d^{'} + (X_s - X_s^{'})i_{qs}] \qquad (7)$$

$$\frac{dE_q^{'}}{dt} = -s_r\omega_s E_d^{'} - \omega_s\frac{L_m}{L_{rr}}v_{dr} - \frac{1}{T_0^{'}}[E_q^{'} + (X_s - X_s^{'})i_{ds}] \qquad (8)$$

IV. MODEL OF THE DRIVE TRAIN

The drive train is the mechanical part of the system that consists of gearbox, shafts, turbine and other important mechanical parts. In this study, the two mass drive train model is used [7] and is presented on Fig. 3. The two mass drive train model is used because of the fact that the turbine model is represented as one mass, while the generator is the other mass. Both masses are connected with a mechanical shaft that possesses certain damping and stiffness value. In the expression, T represents the torque, Ω is the angular speed and J is the inertia moment. In the expressions, subscripts t and g represent the generator and turbine side respectively.
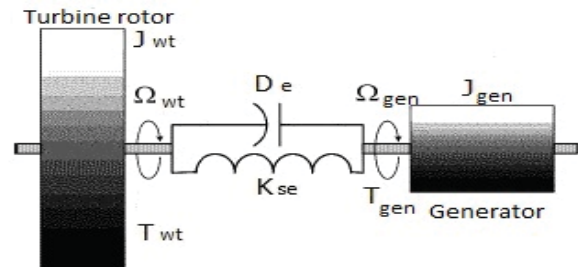
Fig. 3 Two mass drive train model

$$T_{wt} = J_{wt}\frac{d\Omega_{wt}}{dt} + D_e(\Omega_{wt} - \Omega_{gen}) + k_{se}(\Theta_{wt} - \Theta_{gen}) \qquad (9)$$

$$\frac{d\Theta_{wt}}{dt} = \Omega_{wt}$$

$$-T_{gen} = J_{gen}\frac{d\Omega_{gen}}{dt} + D_e(\Omega_{gen} - \Omega_{wt}) + k_{se}(\Theta_{gen} - \Theta_{et}) \qquad (10)$$

$$\frac{d\Theta_{gen}}{dt} = \Omega_{gen}$$

where equivalent stiffness and equivalent moment of inertia for the rotor is given by:

$$\frac{1}{k_{se}} = \frac{1}{\frac{k_{wt}}{k_{gearbox}^2}} + \frac{1}{k_{gen}} \qquad (11)$$

$$J_{wt} = \frac{1}{k_{gearbox}^2} \cdot J_{wt}' \qquad (12)$$

V. ROTOR SIDE CONTROLLER

The main purpose of using the rotor side controller is to measure and control the voltage and energy on the output of the turbine. The logic that is used is PI controllers. By using PI controllers, the overall energy is added to the overall loss of energy, and that sum is compared to the reference value from the characteristic. When the controller works in voltage regulation mode, the PI control is performed through the UI characteristic. The PI controller is actually used to decrease the error of the power to zero. The simplified block diagram of the rotor side converter is presented on Fig. 4.
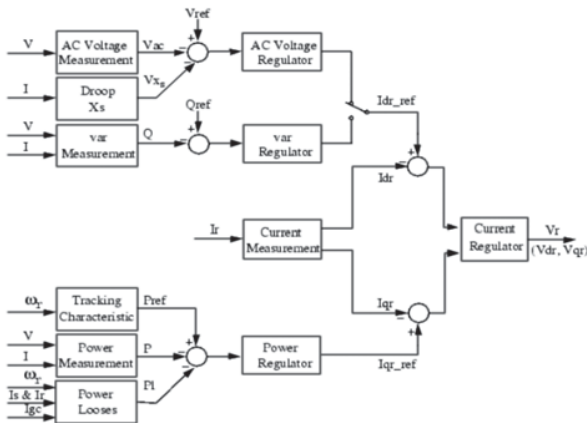
Fig. 4 Rotor side controller

a. Wind turbine model

For the wind turbine model that is used here, first we must start from the basic concept of wind turbines. Every wind turbine uses the kinetic energy that is present in the wind and transforms it into mechanical energy. The power that is developed by the turbine is represented with the following equation:

$$P = \frac{1}{2}C_p \rho A V^3 \qquad (13)$$

where C_p denotes the power coefficient, ρ is the air density, V is the wind velocity and A is the area of the turbine blades. If we insert another variable, called tip speed ratio, which is a ratio between turbine blade linear speed and the wind speed. The mathematical expression is

$$\lambda = \frac{R\omega}{V} \qquad (14)$$

If in the expression (13) we substitute the expression (14), we get the simplified expression for the power from the wind turbine

$$P = \frac{1}{2}C_p(\lambda)\rho A(\frac{R}{\lambda})^3 \omega^3 \qquad (15)$$

In this study case we have used the mechanical characteristic that is presented on Fig. 5.
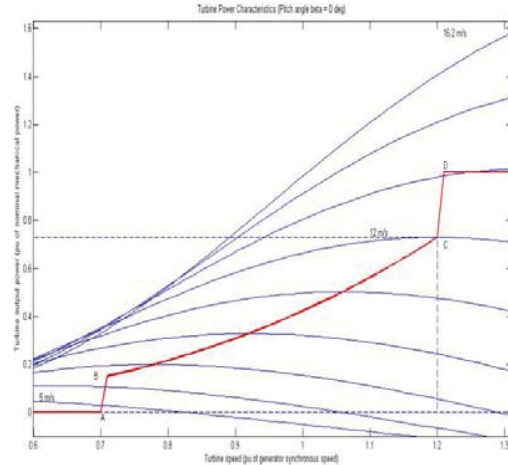
Fig. 5 Mechanical characteristic of the used wind turbine

On Fig. 5 on the horizontal axis we have assigned the parameters for the turbine speed and on the vertical axis we have presented the output power of the turbine.

VI. PITCH ANGLE CONTROLLER

In this case study we have used PI logic for the pitch angle controller. The PI controller in the model is used to limit the output electric power to the nominal mechanical power that id defined. When the measured electrical power level is under the defined nominal value, the pitch angle is kept at zero degrees. In case when the electrical power increases and achieves greater values, the PI controller takes action and increases the pitch angle in order to bring back the nominal value of the measured power. The block diagram of the PI controller that is used as pitch angle controller is presented on Fig. 6.
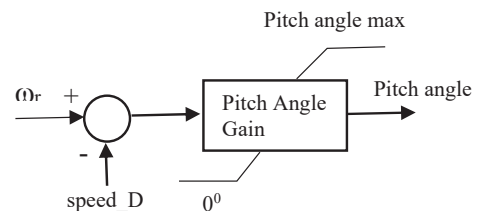
Fig. 6 Pitch angle controller

The pitch angle controller tracks the speed that is defined in the speed tracking characteristic.

VI. SIMULATION

We investigate dynamic model of wind farm which will located in southeastern town of Bogdanci [11]. The wind farm is planned to be connected through transformers followed by a single set of transmission lines to a grid. Power generation in the grid is dominated by synchronous generators, so the grid is treated as an infinite bus. In this study case we have made a simplified MATLAB Simulink model of the transmission power system of Macedonia. In that power system there is a wind farm of 50 MW that has to be installed. The number of used wind turbines is 22, each one of them with installed power of 2,3 MW. All turbines are DFIG. In the initial MATLAB model of the study case, we implemented all 22 turbines in one system, and thus we received one system with 50 MW of power. A block diagram of the model used in this study case is presented on Fig. 7.
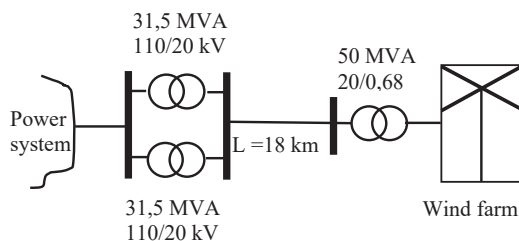
Fig. 7 Wind farm connection on power system

Two tests have been performed. The first test was made under short circuit on the 110 kV bus near the wind farm, or to be more specific, on the place where the wind farm is connected to the transmission power system. At the second test we assumed that on the wind farm there was a sudden change on the wind speed. The turbine works in two modes. The first mode is VAR Regulation while the other mode is the Voltage regulation. It is in everyone's interest to perform simulations under these two working modes. At first we will start by assuming that we have single phase short circuit on the 110 kV bus.

As first test we decided that we will use the case where the wind farm works in Voltage regulation mode. The single phase short circuit starts at 5 sec and lasts until 5.18 sec. For this purpose we present the following screenshots for the Network parameters and Turbine parameters, accordingly. As first diagram we will present the Network parameters on Fig. 8, while the Turbine parameters are presented on Fig. 9.

As it is presented in Fig. 8, on the 5th second there is an active short circuit on the 110 kV buses. During the short circuit period we see that the voltage on one phase on the 110 kV bus drops to 0 and lasts during the short circuit period. It is in our interest to track all the parameters in the wind turbine. We can see from Fig. 9 that the voltage at the wind turbine terminals during the short circuit drops to 0.97. The turbine protection system is set up to react in cases when the voltage drops under 0.95 pu for time greater than 0.1 seconds. The wind farm continues to work and gives full power in the

transmission system. In conclusion we can state that the short circuit didn't affect the work of the turbines. In the second working mode which is called VAR regulation mode, the wind turbine trips and stops working from the 5th second.
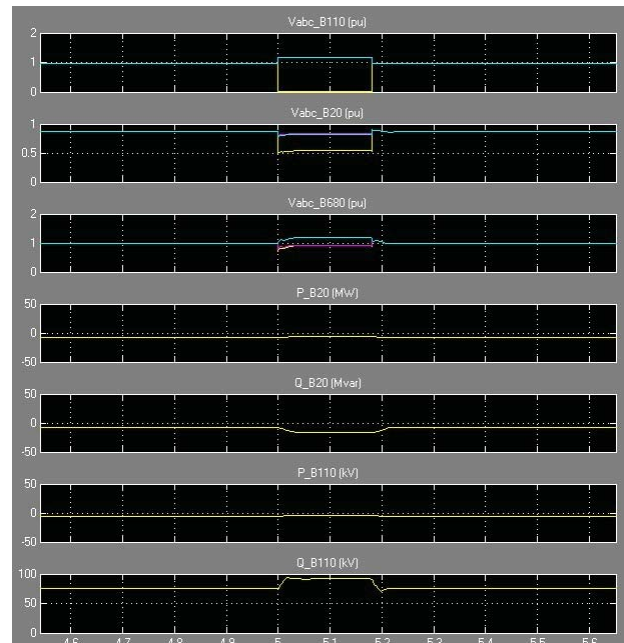
Fig. 8 Network parameters (VAR regulation mode)
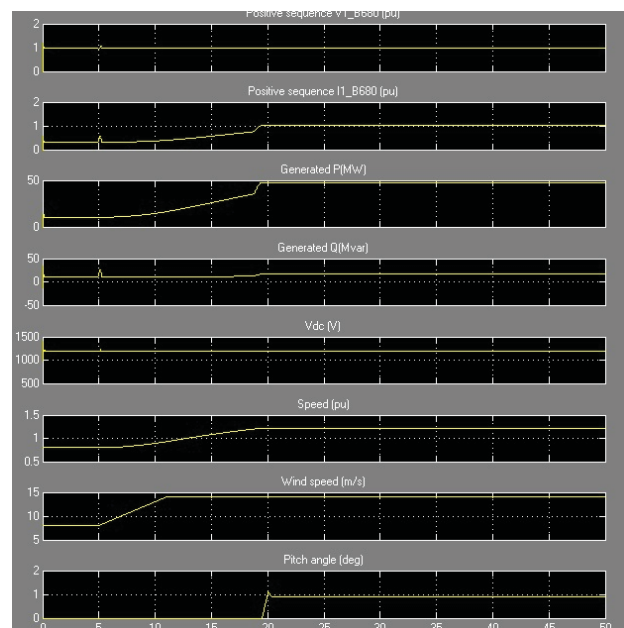
Fig. 9 Turbine parameters (VAR regulation mode)

In the second experiment we assume that the wind turbine worked in Voltage regulation mode. The turbine works during the whole period that is given in the simulation parameters. On Fig. 10 we present the Network parameters and the results from the simulation.
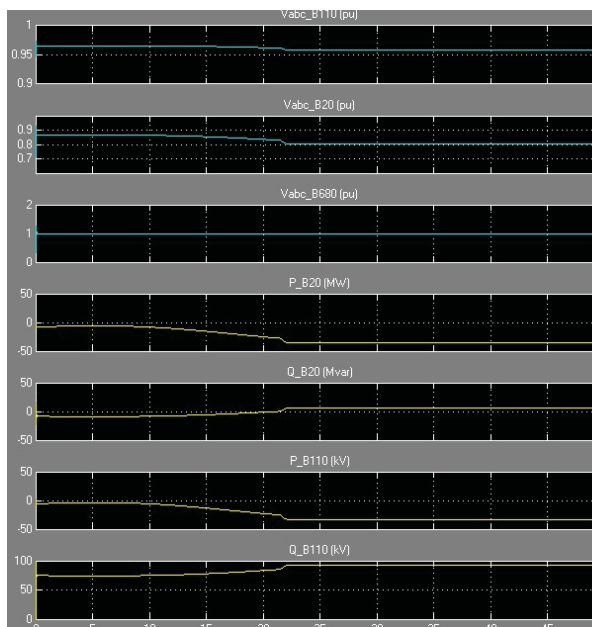
Fig. 10 Network parameters (Voltage regulation mode)
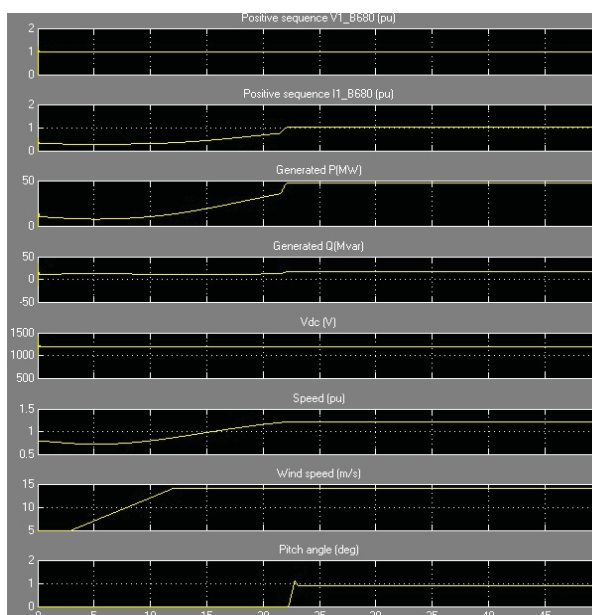
Fig. 11 Turbine parameters (Voltage regulation mode)

As it can be seen from the simulation results in Fig. 10 and Fig. 11, in this mode, the system continues to work and generates power into the transmission power system.

From Fig. 11 that represents the turbine parameters it can be seen that the wind speed increases to 14 m/s starting from the 3rd second of the simulation period. The pitch angle increases up to 1.2 degrees in order to control the mechanical power of the turbine. It is a protection algorithm which is used in the protection of the wind turbine. In the VAR regulation mode, the wind turbine trips and doesn't insert energy into the power

system, because of the fact that the protection which is set up to disconnect the turbine from the network in case the voltage on the output of the turbine drops under 0.95 pu, reacts and disconnects the turbine from the network.

VII. CONCLUSIONS

From the simulation results shown in the previous section of the article a very useful conclusion can be drawn. Because Macedonia is in constant need of electrical energy, our proposal would be to implement the wind turbines into the system, and they should work in Voltage regulation mode. The place where the turbines will be installed is windy, and if the VAR regulation mode is used, then the system won't have great benefit of them. The Voltage regulation mode is more tolerant unlike the other mode, where the turbine trips and disconnects from the network.

REFERENCES

[1] V. Akhmatov, "Analysis of dinamic behaviour of electric power systems with large amount of wind power", Ph.D. Thesis, TU of Denmark, 2003.

[2] G. L. Johnson, "Wind energy systems", 2006.

[3] A. Perdana, "Dynamic models of Wind Turbines", PhD dissertation, Chalmers University of Technology, Götegorg, Sweden, 2008.

[4] T. Sun, "Power quality of Grid-Connected Wind Turbines with DFIG and Their Interaction with the Grid", PhD dissertation, Aalborg University, Denmark, 2004.

[5] J. T. G. Pierik, Y. Zhou and P. Bauer, "Wind farm as Power Plant", TU Delft, 2008.

[6] B. Babypriya and R. Anita, "Modelling, Simulation and Analysis of Doubly Fed Induction Generator for Wind Turbines", Journal of Electrical Engineering, Vol. 60, No2, pp. 79-85, 2009.

[7] C. Rahmann, H.J. Haubrich, L. Vargas and M.B.C.Salles, "Investigation of DFIG with Fault Ride-Through Capability in Weak Power Systems", IPST2009, Kyoto, Japan, 2009.

[8] F. Iov, A.D. Hansen, P. Sorensen and F. Blaabjerg, " Wind Turbine Blockset in Matlab/Simulink", RISO&Aalborg University, Denmark, 2004.

[9] B.Chitti Babu and K.B. Mohanty, "Doubly-Fed Induction Generator for Variable Speed Wind Energy Conversion Systems-Modeling& Simulation", International Journal of Computer and Electrical Engineering, Vol. 2, No.1, pp. 141-147, 2010.

[10] A.D. Hansen and G. Michalke, "Voltage grid support of DFIG wind turbines during grid faults", EWEA, pp. 93-97, 2007.

[11] ***, "Bogdanci Wind Park Project", Main Design, Siemens Energy Sector, Wind Power Division, 2012.

Smart System for Domestic Power Consumption Measurement

Vasil Shterev[1], Hristiyan Kanchev[2] and Eltimir Stoimenov[3]

Abstract – **Smart-meters are devices that record the consumption of electrical energy with a certain time step and are able to analyze and communicate this data to the user or grid operator. This paper presents a conception how to add similar functionalities to any power-meter available on the market equipped with pulse output for external counter connection.**

Keywords – **Power, Data-logger, Wireless, M2M, Smart-meter.**

I. INTRODUCTION

A multitude of devices for data acquisition and logging exist on the market for metering the consumption of natural gas, water, electricity and other communal services essential for the modern society. Smart meters of electrical energy exist for metering of single or three phase power supply. Having in mind the popularity of single-phase power supply of domestic consumers, the study is focused on this type of smart meters, although the functionality of the device described in this paper is not dependent on the number of phases. The one thing necessary for its correct operation is an electrical or optical signal output for external metering of the already installed power meter.

For example, the South Korean company "Sanxing electric" [1] offers multiple types of smart meters, metering boxes and other customer-side solutions. Some smart meters have option for metering of prepaid services, according to the tariffs and the consumption forecast. A short survey of existing smart metering devices on the internet shows multiple producers, installers and operators of such devices. By the introduction of options for communication between the devices and the grid operator, as well as algorithms for clustered operation of smart metering devices, one of the aspects of the Smart Grid is being implemented. The Smart grid includes automated data acquisition and treatment of collected data for improvement of the efficiency and reliability of power supply from the distribution grid operator. The device described in this paper complies with the above stated, which can also be regarded as one of the aspects of Internet of Things which encompasses

bidirectional communication between devices. For this to be implemented every device should have its own IPV4 or IPV6 address in the internet or in the local network behind the consumers' router. In search for originality many producers give alternative names to their devices including advance, intelligent, smart, think, etc., but the main functionalities are mostly the same: machine to machine bidirectional communication, data logging and statistical calculations of the resource consumption (electricity, water, gas etc.).

The annual expenses related to renewal of the metering devices and installation of smart meters of several distribution system operators are presented in table 1. The increasing trend is clearly pronounced. A considerable leap was observed in 2009 and 2010 which is a consequence of the restoration after the global economic crisis in 2008. It should be noted that these annual costs include the prices for replacement of the obsolete metering devices and annual servicing and inspection of the smart meters.

TABLE I
ANNUAL METER CHARGE

Annual meter charge increases with smart meter costs in 2010 and projections to 2017 ($)

Distributor	2005	2006	2007	2008	2009	2010	2011	2012	2013	2015	2016 2017
SP AusNet	17.49	17.49	17.49	17.49	17.49	86.1	93.83	101.02	108.75	117.08	126.04
United Energy Distribution	6.60	6.60	6.60	6.60	6.60	69.21	89.18	99.57	107.62	116.33	125.73
Jemena Electricity Networks	12.87	12.87	12.87	12.87	12.87	134.63	136.7	155.84	159.86	162.34	164.88
Citipower	15.20	15.20	15.20	15.20	15.20	104.79	108.4	93.38	95.26	97.17	99.13
Powercor	17.20	17.20	17.20	17.20	17.20	96.67	105.35	92.72	93.91	95.12	96.34

The above presented is just one of the aspects. Although the technological progress and the more and more informed society, there are still numerous non-governmental organizations that are concerned by the risks (sometimes due to misinformation or other reasons) related to high-frequency communications, radiation, risks of fire or electrical shock, as well as the risks related to privacy and theft of personal data from these smart devices, etc. [2] [3] [4].

As a response to these arguments, for example the British government announced that installation of smart meters at every consumer will result in economies exceeding 6 billion of British pounds [5]. A brief synthesis of the risks is presented in table 2 [6].

TABLE I
RISKS RELATED TO PRIVACY AND THEFT OF PERSONAL DATA [6]

1. Identity Theft
2. Determine Personal Behavior Patterns
3. Determine Specific Appliances Used
4. Perform Real-Time Surveillance
5. Reveal Activities Through Residual Data
6. Targeted Home Invasions (latch key children, elderly, etc.)
7. Provide Accidental Invasions
8. Activity Censorship
9. Decisions and Actions Based Upon Inaccurate Data
10. Profiling
11. Unwanted Publicity and Embarrassment
12. Tracking Behavior Of Renters/Leasers
13. Behavior Tracking (possible combination with Personal Behavior Patterns)
14. Public Aggregated Searches Revealing Individual Behavior

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

[2]Hrystian Kanchev is with the Faculty of Electronic engineering and Technologies at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: hkanchev@tu-sofia.bg

[3]Eltimir Stoimenov is with the Faculty of Electronic engineering and Technologies at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: Bulgaria, E-mail: e_stoimenov@tu-sofia.bg.

The following figure presents a typical household consumption data for a 24 hour period: it comprises a morning peak between 08:00 and 09:30 AM, a few short peaks during the day (optional, for appliances with timers for automatic start and shutdown, like boilers, washing machines etc.) and two evening peaks – one in the early evening and one later (fig.2).
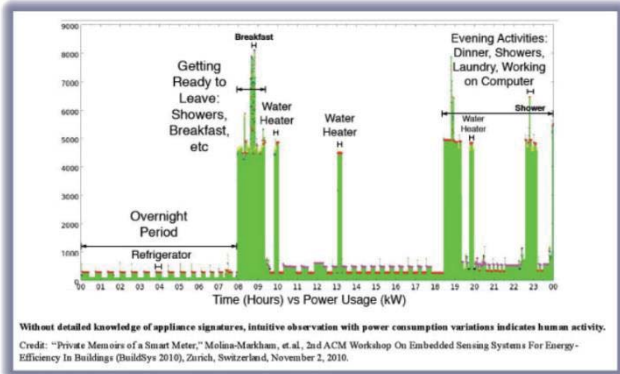

Fig. 1. Typical profile of daily power use [6]

This paper presents a smart metering device with multiple functions, which can be installed without interventions in the existing electrical installation at the consumer's side. In this way the possible dangers and risks are reduced to a minimum if not excluded at all. Mounting and operating with the device are simplified, so that even persons without qualification are able to install the device and operate with it. Furthermore, the proposed system is flexible and modular for easy implementation, cooperation with other components of a smart home energy management system and the possibility for cooperation with other components and posterior upgrades of the home system.

II. IMPLEMENTATION

The proposed system architecture is composed of a master controller and several slave-devices (called nodes) connected to the power meter, ensuring a safe communication channel with the concentrator (hub) device. Every node communicates data from the actual metering interface to the concentrator. As a concentrator can serve either a dedicated Raspberry PI-based computer or every smartphone featuring Bluetooth LE communication.

The Raspberry PI system itself presents an integrated microcomputer with ARM processor with 4 cores, 1 GB of random access memory with high throughput and multiple input/output interfaces for digital and analog communication. The raspberry PI platform includes its own open source linux-based operating system – the Rasperian that facilitates development of user interfaces and applications.

As mentioned above, in this paper the communication channel between the nodes and the concentrator is implemented by Bluetooth Low-Energy (BLE) or the Bluetooth Smart technology. It uses a short range radio with a minimal amount of power to operate for a longer time (even for years) compared to its previous versions. Its range coverage (about 100 meter) is ten times that of the classic Bluetooth while its latency is 15 times shorter [8]. This choice is logical, bearing in mind the paradigm for maximal node lifetime when using batteries as a power source and the option for interconnection with SCADA system or external database. The so exposed concept is completely compliant and ready for integration in the Internet of Things (IoT) and the future Smart Grid [1, 7].

A. Block Diagram

In this chapter are detailed the functions of the system blocks, presented at fig. 2 and their interfaces. The argumentation for the chosen electronic components for device implementation will also be presented in this chapter.
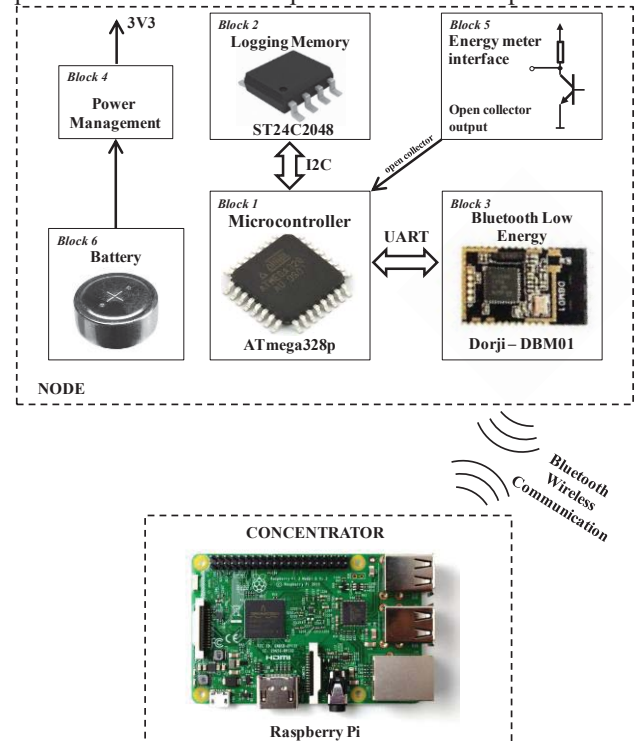
Fig. 2. Block diagram of one node and the concentrator

Block 1 "Microcontroller" is implemented by an ATmega328P microcontroller of the Atmel Company. The choice of this microcontroller is based mainly on its compatibility with the Arduino system developer environment and the power supply of 3,3V. Furthermore, this microcontroller features all the peripheral modules required for implementation of the system presented in this paper.

The Block 2 - "Logging memory" is an external memory module EEPROM, type 24C1014 of the ST Microelectronics Company. The memory size is 1 Mb, which is completely enough for storage of the electrical consumption data for more than one year. The connection between this block and the "Microcontroller" block is realized by an I2C digital interface.

Block 3 of the system diagram is responsible for the wireless communication between the node and the concentrator. The block is implemented by an integrated module DBM01 of the "Dorji" company. The module is compliant with the Bluetooth 4.0 specification and communication is performed by standard AT commands on the microcontroller UART interface.

The Block 4 - "Power Management" ensures the necessary power supply with a voltage 3,3V corresponding to the requirements of the device components. The block is actually a buck/boost DC-DC converter. This is necessary because the voltage of a fully charger lithium-ion battery exceeds 3,6V and a step-down conversion is required. At contrary, by a partially discharged Li-ion battery the voltage may drop below 3,3V and the converter has to work in "boost" mode (corresponding to the actual duty ratio of the electronic switch) to supply the necessary voltage to the device.

Block 5 of the presented block diagram is the interface to the power metering device. Most if not all of the modern electronic power meters have an "open collector/drain" type pulse output for connection of external counter. The number of pulses per second at this output is proportional to the power consumption (for example 1800 or 3600 pulses per kWh). Connection of the external metering device only requires an external pull up resistor and a port of the microcontroller featuring an interrupt request (IRQ).

The battery block consists of the battery supplying voltage to the DC-DC converter. A Lithium-ion battery is chosen for their high energy density and low self-discharge. The device is designed to be able to work at least one year on battery and also the metered domestic power consumption doesn't affect the battery lifetime.

B. Block Algorithm

The block diagram of the device operation algorithm is presented on fig. 3. The algorithm is composed of three major parts: 1) The main program (main function), 2) the subprogram for operation by a GPIO interrupt request and 3) subprogram for operation by an interrupt from the timer subsystem of the microcontroller. The main program executes first a system initialization, including initialization of the microcontroller peripheral modules, the Bluetooth module and the external EEPROM memory.

The subprogram for operation by a GPIO interrupt is triggered by an active front at the microcontroller input pin connected to the metering circuit, in other words when power consumption is registered by the power meter. This program is developed on an algorithm with the goal of minimal data

recorded on the EEPROM memory. This is achieved by a comparison in the timestamp values of two consecutive pulses (recorded by the timeStamp variable). If the difference of the two timestamps is longer than a certain critical value (t_{cr}), in the memory is recorded a timestamp for every next pulse. But if the difference between the two first timestamps is less than t_{cr}, in the memory is recorded a number of pulses for a given time period. In this way, when power consumption is greater than a certain value, the unnecessary recording of a timestamp for every pulse at every 1/10000 of the second is not performed, but instead a timestamp for several hundred pulses is recorded. This reduces the memory requirement. The data format will be detailed in the next chapter of this paper.

The internal timer of the microcontroller calls an interrupt over a defined period of time. Every interruption coming from the internal timer increments the *timeCnt* variable, which is later responsible for storing data in the EEPROM memory.
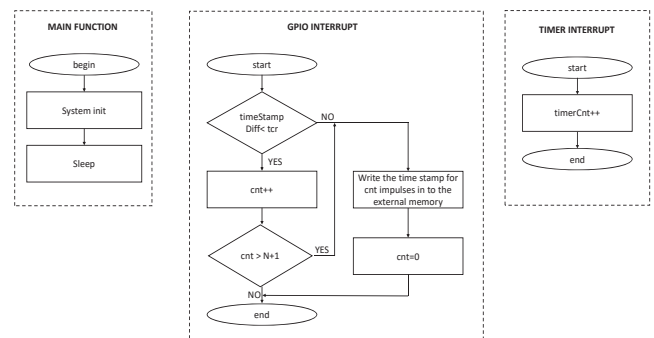
Fig. 3. Block algorithm

C. Data Format

Fig. 4 presents the data array format, recorded in the node's external memory. By request this data is sent to the concentrator by the Bluetooth communication channel. The data array itself is composed by smaller 6-byte blocks of data that contain information about the number of recorded pulses, as well as timestamps defining the period over which the pulses are recorded.
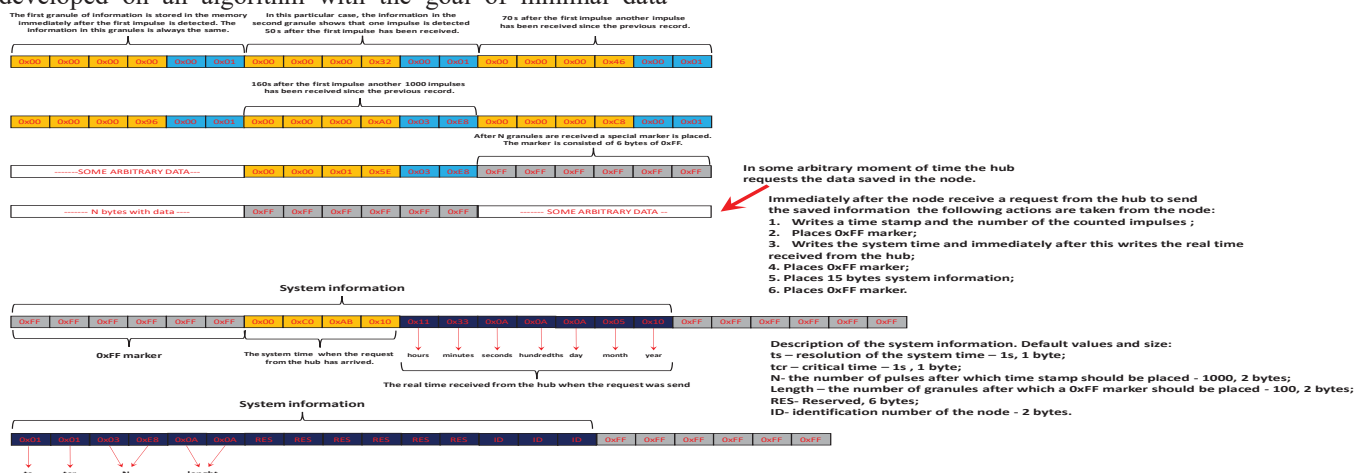
Fig. 4. Data format

At the end of the data array is recorded certain system information like the current time and the system parameters and constants, etc.

The data blocks (grains) are the smallest unit in the data array structure. The composition of a single data block is presented on fig. 5.
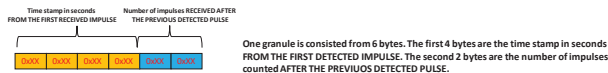
Fig. 5. One grain consists of 6 bytes

It contains a total of 6 bytes of information. The first 4 bytes are a timestamp indicating the time at which data was recorded. It should be mentioned, that time is relative to the first recorded pulse. The last 2 bytes of information contain the number of pulses recorded after the first one.

An example of data array recorded in the node memory is presented on fig. 6. The first block of the array is recorded when the first pulse from the power meter output is registered. This is the reason why the system time in the block is always 0x00 0x00 0x00 and the number of pulses is 0x00 0x01. This content is always the same and can be used as a marking for the beginning of the data array. For every N blocks of the data array, a synchronization marker is added in the array. The synchronization marker is composed by a data block with values of 0xFF on all the 6 bytes (for this reason it will be called "FF marker" below). The data array continues until all the available memory is allocated or until a data transfer request is received from the concentrator. Upon a data transfer request, the following actions are performed before sending the data:

1) A final data block is saved containing information of the actual time and the total number of pulses counted by the meter (which is in fact the total energy consumption for the recorded time period;

2) Saves an "FF marker" at the end of the data array;

3) Records 4 bytes of data with the time when the data transfer request is registered followed by the actual time and date the request is executed. Saving of the real time is important, because up to this moment all timestamps are in relative time (counted from the first pulse registration) and in this way they can be converted in calendar time for statistical consumption data or other calculations performed by the system;

4) Records another "FF marker";

5) Records 15 bytes of system (service) information: their meaning and default values are as follows:

1 – the resolution of time counting – 1s, 1 byte;

2 – the critical time *tcr* used for data loging – 1s, 1 byte;

3 – the maximal number of pulses each data block can contain – 1023 pulses , 2 bytes;

5 – the number of data blocks between the synchronization "FF markers" - 100 blocks, 2 bytes;

7 – Bytes reserved for implementation of future options and features – 6 bytes;

13 – device identification number.

III. CONCLUSION AND FUTURE WORK

A smart system for monitoring, recording and graphical representation of information about the consumption of electrical energy. The main objectives for this device implementation are: maximum duration of device operation without the need for battery replacement/charging, intuitive user interface featuring the option for visualization of multiple values related to the instantaneous power, average power over a given time period (including historical data). A protocol was developed (including the format of the recorded data) for exchanging data between the nodes and the concentrator. Last, but not least the developed software can give advices for energy saving without influencing the consumers comfort. Future developments of this system include the options for user control over a distant node (by the use of an internet connection), remote data acquisition and logging to a remote server.

ACKNOWLEDGEMENT

REFERENCES

[1] http://www.sanxingelectric.com/
[2] https://en.wikipedia.org/wiki/Smart_meter
[3] http://www.electrosmogprevention.org/stop-ca-smart-meter-news/when-smart-meters-cause-fires-and-kill/
[4] http://www.naturalnews.com/050743_Smart_Meters_explosions_house_fires.html
[5] http://www.telegraph.co.uk/finance/personalfinance/energy-bills/11617413/Smart-meters-will-you-pay-more-for-peak-electricity.html
[6] https://smartgridawareness.org/privacy-and-data-security/how-smart-meters-invade-individual-privacy/
[7] Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications Ala Al-Fuqaha, Senior Member, IEEE, Mohsen Guizani, Fellow, IEEE, Mehdi Mohammadi, Student Member, IEEE Mohammed Aledhari, Student Member, IEEE, and Moussa Ayyash, Senior Member, IEEE
[8] R. Frank, W. Bronzi, G. Castignani, and T. Engel, "Bluetooth low energy: An alternative technology for VANET applications," in Proc. 11th Annu. Conf. WONS , 2014, pp. 104–107.

Mining Ring Diagnosis using Artificial Neural Networks

Dejan Stevanović[1], MionaAndrejević Stošović[2], Marko Dimitrijević[2]

Abstract – **In this paper we will present one aspect of energy consumption caused by equipment for cryptocurrency mining. We will first address problems concerning increscent number of nonlinear loads leading to the fact that the active power no longer represents the main part of total power delivered to customer. We will stress in this paper losses produced by power supply unit in the mining ring, and we will engage an artificial neural network to diagnose how many cards are working at the moment.**

Keywords – **Artificial neural network, Power Meters, Utility Losses, Cryptocurrency Mining.**

I. INTRODUCTION

Our life has become more comfortable during last few decades due to plentiful of smart electronic appliances. Simultaneously, the electronic control systems became inevitable parts of equipment for industrial production. Most of electronic gadgets and apparatus require DC supply. Therefore, AC to DC converters have become the most numerous loads at power grid. Unfortunately, their nonlinear nature generates harmonics in the power network causing numerous unwanted problems [1], [2], [3].

The permanent growth of the number and types of nonlinear loads aggravates the problems caused by harmonics. That enforced almost every country to introduce its own standard that restricts the allowed amount of each harmonic. Two widely known standards in this area are the IEEE 519-1992 and IEC 61000 series [1], [3]. The standard IEC/EN61000-3-2 entered into force in the European Union. It specifies the limits for the allowed nonlinear distortion of the input current up to the fortieth harmonic. The standard is applied to the distortion produced by electronic and electrical appliances in households. This includes loads up to 16A per phase supplied with voltage up to 415V. Both standards regulate limits for the harmonics pollution but do not specify what happens if a customer exceeds them. There are two possibilities: the first suggests that the utility could disconnect that customer but that is stressful and not profitable solution. The better way and the most effective tactic is to charge the harmonics producers a penalty tax if they exceed limits of harmonics pollution. The penalty tax should be proportional to the pollution levels. But this can be possible only in case when we have precise method for identification of

the harmonics' producers. The overview of these solutions can be found in [4].

Lately, one of the greatest nonlinear consumers is equipment for cryptocurrency mining. Blockchain technology and its most popular cryptocurrency, Bitcoin, have been called of one of the most intriguing issues of nowadays, having almost equal importance as the internet. There are presumptions and hopes that cryptocurrencies will change the world, for the better, of course. But there is a dark side of this story. According to the Bitcoin energy consumption index, the digital currency already consumes 0.15% of the world's energy, and far exceeds the electricity consumption of Ireland or of most African nations [5]. Or, to get better insight, it costs 29 times as much energy to produce Bitcoins last year as it did to power all the Tesla cars driving today [6].

The reason Bitcoin mining consumes so much energy is because in order to produce each new Bitcoin, solving a complex mathematical puzzle is required, and it takes cryptographic process performed by high-powered computers. The mining computations serve to verify Bitcoin transactions on a digital ledger known as the blockchain, and the greatest advantage is because it ensures security. But, again, this process is extremely energy intensive, and in order to put the energy consumed by the Bitcoin network into perspective we can compare it to another payment system like VISA for example. Considering the numbers [7], we can conclude that Bitcoin is extremely more energy intensive per transaction than VISA, because Bitcoin transaction requires several thousands of times more energy. These problems do not refer only to energy price, it should be about the environment, too.

In this paper we will give one aspect of energy consumption caused by equipment for cryptocurrency mining. Namely, we will refer to power consumption of power supply unit in the mining ring. We will measure quantities of active, reactive and distortion power, and using these values we will diagnose the number of the processing (GPU) cards operating. Artificial neural networks will be used for diagnosis.

II. THE FUNDAMENTAL QUANTITIES

Traditional power system characterization quantities such as RMS values of current and voltage, power (active, reactive, apparent) are defined for ideal sinusoidal conditions. However, in the presence of nonlinear loads, these definitions need correction. The instantaneous values of a quantity rich with harmonics (voltage or current) can be expressed as:

$$x(t) = \sum_{h=1}^{M} X_h \sin\left(\omega_h t + \alpha_h\right) , \qquad (1)$$

[1]Dejan Stevanović iis with Innovation Centre of Advanced Technology (ICNT), Niš, Serbia, dejan.stevanovic@icnt.rs.
[2]MionaAndrejević Stošović, Marko Dimitrijević are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, Niš, Serbia, {miona.andrejevic, marko.dimitrijevic}@elfak.ni.ac.rs

where h is the number of the harmonic, M denotes the highest harmonic, while X_h, ω_h and α_h represent amplitude, frequency and phase angle of the h-th harmonic. The RMS value of the signal expressed by (1) is defined as:

$$X_{RMS} = \sqrt{\sum_{h=1}^{M} X_{RMSh}^2} , \qquad (2)$$

where X_{RMSh} is the RMS values of the h-th harmonic.

Product of the voltage and current having the same harmonic frequency gives the harmonic power. Total active power is defined as:

$$P = \sum_{h=1}^{M} V_{RMS_h} I_{RMS_h} \cos(\theta_h) = P_1 + P_H , \qquad (3)$$

where θ_h denotes phase angle between voltage and current, P_1 denotes power of the fundamental component ($h = 1$). Therefore, it is known as *fundamental active power* component while P_H comprises sum of all higher components ($h = 2,\ldots,M$) and is referred to as *harmonic active power*.

According to Budeanu [3], [8], [9] reactive power is defined as:

$$Q = \sum_{h=1}^{M} V_{RMS_h} I_{RMS_h} \sin(\theta_h) = Q_1 + Q_H , \qquad (4)$$

where, similarly to (3), Q_1 and Q_H denote *fundamental reactive power* and *harmonic reactive power*, respectively.

Many scientists claim that the Budeanu's definition is not correct and cannot be used for calculating reactive power. According to one of the authors of IEEE1459-2010 standard, professor Emanuel [10], [11], "even today this definition occupies a significant number of pages on *The IEEE Standard Dictionary*". Its past acceptance and popularity among engineers and top scientists is hard to dispute. Modern textbooks written by highly respected researchers are presenting Budeanu's resolution of apparent power as the right canonical expression". More about calculating reactive power can be found in [10].

It is well known that the apparent power is a product of RMS values of voltage and current. In presence of harmonics, the apparent power is calculated as:

$$S = I_{RMS} \cdot V_{RMS} = \sqrt{\sum_{h=1}^{M} V_{RMSh}^2} \cdot \sqrt{\sum_{h=1}^{M} I_{RMSh}^2} \qquad (5)$$

The obtained value for apparent power obtained by using previous equation is greater than the value that can be obtained as $S = \sqrt{P^2 + Q^2}$. This was noticed for by Budeanu for the first time in 1927. Therefore, he introduced the term *distortion power* and revised the equation for apparent power:

$$S^2 = P^2 + Q^2 + D^2 \qquad (6)$$

Consequently, distortion power can be calculated by using next equation:

$$D = \sqrt{S^2 - P^2 - Q^2} . \qquad (7)$$

III. Measured results

We obtained measured results that will be used in this paper for a neural network training by using standard power meter

produced by EWG [12]. It is based on standard integrated circuit 71M6533 [13]. The power meter completely fulfils IEC 62053-22 standard [14]. The only additional effort was to gather data provided by the meter and to acquire them using a PC to calculate distortion power according to (7).

Figure 1 illustrates the implemented set-up. It consists of the meter, the load and PC. The meter sends the measured data through its optical port. PC receives them on RS232 port. Dedicated software processes data and forwards them to MATLAB script that calculates the distortion power.
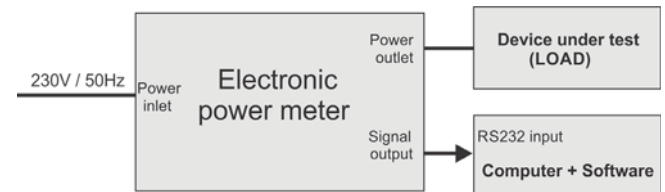
Fig. 1. Set-up circuit for distortion power measurement

The two main components of a mining ring are graphics card with graphic processor (GPU) and power supply unit (PSU). According to mining experts the most profitable GPUs in 2018 are RX580 from AMD and GTX1070 from NVIDIA. The second important thing in mining ring is power supply unit. This unit must be highly efficient because it needs to decrease power consumption, and thus, the costs of mining. So, the most of mining rings have a PSU unit with efficiency more than 80% (bronze, silver, gold, platinum and titan design). The PSUs with the same design have different efficiency when they are loaded differently. For example, 80 PLUS Gold PSU supplied by 230V has 88% efficiency when load rate is 20%, but it has 92% efficiency when load rate is 50% [15].

In this paper we present measured powers' results (active, reactive, distortion and apparent power) for two PSUs.

The mining ring that was used as a device-under-test consists of: PSU Cooler Master 750W bronze design, Sesonic Focus PSU 850W gold design, 6 Gigabyte GTX1070 graphics cards (3 cards are with one single fan, and other 3 are with three fans). The Sesonic Focus PSU supplies graphics cards with three fans, while Cooler Master PSU supplies the other three graphics cards, motherboard and solid-state disc (SSD) [16].

Table I presents obtained measured results for all mining rings. At the beginning of the measurement process, we turned three cards on, one by one. After that we turned off two cards and then started to turn on card by card. In situation when all six cards were working the active power was around 810W, reactive power was around 80VAR, while distortion power was about 125VAR. At the end we turned off 3 cards at the same time, and so only 3 cards continued to work.

We need to stress that each of these values presented in the Table I is averaged out of approximately 50 measurements. For example, value of 225.49W measured for active power when one card was working is obtained as average value of about 50 measurements in the period of 6 hours. Few measurements were not taken into account when averaging, because they were left for testing, what will be given later in the paper.

TABLE I
MEASURED RESULTS GIVEN IN FIG. 2

Meas. No.	No. of cards	P (W)	Q(VAR)	S(VA)	D(VAR)
1	1	225.49	106.85	251.85	34.07
2	2	333.26	120.57	360.75	67.17
3	3	447.19	134.10	476.38	94.38
4	1	217.62	102.03	241.01	16.47
5	2	309.43	113.91	333.50	49.43
6	3	429.93	128.43	456.21	81.63
7	4	556.81	158.62	585.93	89.93
8	5	686.23	172.41	715.87	108.56
9	6	812.45	184.84	842.57	125.09
10	3	434.60	130.37	461.10	82.03

These measurement results are given in the Figure 2, where it is more obvious that values of all the powers are not negligible, what is especially important for the values of distortion power, that is not registered on the network [4].

Also, the values of powers when the same number of cards was working are not always the same. For example, when 3 cards work, we obtain average values of 447.19W, 429.93W and 434.60W for active power. For distortion power, values differ even more (94.38VAR, 81.36VAR, 82.03VAR). This was the reason why we got to an idea to involve a neural network in order to resolve this situation. In fact, we will train an artificial neural network (ANN) that will diagnose how many graphics cards are working at the moment.
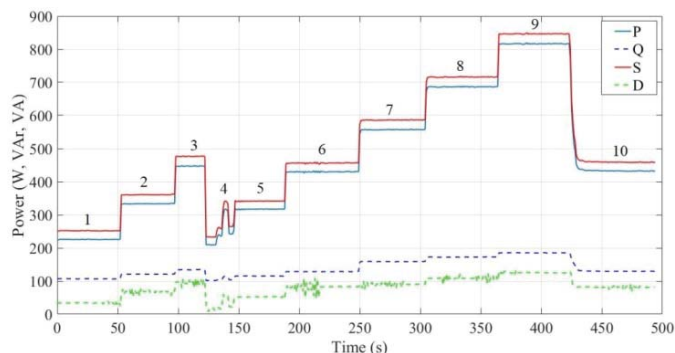
Fig. 2. Measurement results for the whole mining ring

This ANN has four inputs (active power, reactive power, apparent power, distortion power) and one output that should give information about how many cards are operating at the moment. This information is very useful because we can be informed if some card is damaged or it cannot work properly.

After we trained an ANN, we obtained 4 neurons in the hidden layer. The validity of this ANN is first checked when we used excitations that were employed in the training process. Results are given in the Table II, and we can see that we obtained very good results, and expected results match with the ANN response. Error is negligible.

TABLE II
ANN RESPONSE TO EXCITATION GIVEN IN TABLE I

P (W)	Q(VAR)	S(VA)	D(VAR)	Expected response	ANN response
225.49	106.85	251.85	34.07	1	1.0001
333.26	120.57	360.75	67.17	2	1.9997
447.19	134.10	476.38	94.38	3	2.99909
217.62	102.03	241.01	16.47	1	1.00001
309.43	113.91	333.50	49.43	2	1.99977
429.93	128.43	456.21	81.63	3	2.98698
556.81	158.62	585.93	89.93	4	3.99789
686.23	172.41	715.87	108.56	5	5.0011
812.45	184.84	842.57	125.09	6	5.99977
434.60	130.37	461.10	82.03	3	3.01556

The next step in ANN validation is to check if this ANN gives proper response to unknown excitations. So, as unknown excitations we will use some of the measured combinations that were not used in the training process, i.e. they were not used in getting average values presented in the Table I.

TABLE III
ANN RESPONSE TO UNKNOWN EXCITATIONS

P (W)	Q(VAR)	S(VA)	D(VAR)	Expected response	ANN response
226.14	106.76	251.26	24.384	1	1.25892
333.8	121.08	361.97	70.276	2	1.9373
446.84	134.14	475.6	92.39	3	3.01691
236.19	103.1	258.27	16.927	1	1.38708
317.24	114.94	341.14	50.261	2	2.09345
430.19	129.23	456.72	82.64	3	2.96514
558.06	158.95	587.26	90.447	4	4.00486
687.49	172.98	715.95	100.106	5	5.0791
817.11	185.34	847.02	124.224	6	6.05286
432.48	129.49	458.76	81.607	3	3.00226

TABLE IV
ANN ERROR

Expected response	ANN response	Error (%)
1	1.25892	**25.8**
2	1.9373	3.13
3	3.01691	0.56
1	1.38708	**38.7**
2	2.09345	4.67
3	2.96514	2.01
4	4.00486	0.12
5	5.0791	1.58
6	6.05286	0.88
3	3.00226	0.075

Obtained results are shown in the Table III. We chose 10 combinations, each one as a representative of a different group. The ANN response is considered to be correct (i.e. acceptable) when its value was in the range [(m-0.5), (m+0.5)]. So, we can conclude that all the responses are correct, but additionally, we analyzed obtained errors, given in the Table IV.

We can notice from Table IV that significant error occurs when only one card is working. In all other cases, the error is very small. This is expected, when only one graphics card is in operation, the contribution of all other parts of the PC in power consumption is significant, as it accounts for almost half of total power consumption. When we turn on more cards, overall power consumption increase is caused only by the cards, consumption of other components remains approximately the same, so its influence is less important.

IV. CONCLUSION

From the measured results presented in this paper we can conclude that PSUs used in mining rings are very efficient and generate small number of harmonics. We used these measured quantities as inputs to a neural network whose output diagnoses number of graphics cards operating at the moment. This diagnosis was very successful. We should notice that these quantities are measured for branded PSUs, so in our future work we will focus on PSUs produced by other manufacturers.

ACKNOWLEDGEMENT

REFERENCES

[1] Singh, G.K.: Power system harmonics research: a survey. *European Tran. on Electrical Power*, 19, 151–172, (2007).
[2] Alhazmi, Y.: Allocating power quality monitors in electrical distribution systems to measure and detect harmonics pollution. PhD dissertations, University of Waterloo, Ontario, Canada, (2010).
[3] Wakileh, J. G.: *Power Systems Harmonics*, Springer (2001).
[4] Stevanović, D., Petković, P.: A Single-Point Method for Identification Sources of Harmonic Pollution Applicable to Standard Power Meters, Electrical Engineering, Volume 97, Issue 2, June 2015, 2015, pp 165-174, ISSN 0948-7921(prn), ISSN 1432-0487(online), doi:10.1007/s00202-014-0324-z.
[5] https://qz.com/1204842/bitcoin-mining-should-use-renewable-energy-if-we-want-cryptocurrencies-to-be-ethical/
[6] http://fortune.com/2018/01/11/bitcoin-mining-tesla-electricity/
[7] https://digiconomist.net/bitcoin-energy-consumption
[8] Emanuel A. E.: Summary of IEEE standard 1459: definitions for the measurement of electric power quantities under sinusoidal, nonsinusoidal, balanced, or unbalanced conditions, *IEEE Trans. on Industrial Applications*, Vol. 40, No. 3, May 2004 pp. 869–876.
[9] Webster J. G.: The measurement, instrumentation, and sensors handbook, *IEEE Press*, 1999
[10] Emanuel A. E.: Power Definitions and the physical mechanism of power, *Wiley and IEEE Press*, 2010.
[11] IEEE Std 1459-2010, IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, Or Unbalanced Conditions.
[12] EWG - multi metering solutions, www.ewg.rs
[13] http://datasheets.maximintegrated.com/en/ds/71M6533-71M6534H.pdf.
[14] IEC 62053-22 Electricity metering equipment (AC) - Particular requirements - Static meters for active energy (classes 0.2S and 0.5S).
[15] https://alldaymining.com/best-psu-mining.
[16] Stevanović, D., Andrejević Stošović, M., Dimitrijević, M., Petković, P., "Utility losses due to cryptocurrency mining", Web proceedings of the 10th International Conference ICT Innovations, Ohrid, Macedonia, September 2018, pp. 27-36, ISSN 1857-7288.

Determining of Magnetic Fields in the Area of Power Distributional Transformer ETN 630 kVA

Emilija Sarafska[1], Blagoja Arapinoski[2] and Vesna Ceselkoska[2]

Abstract: **The continuous increase of the number of appliances using electrical energy in everyday life, causes accordingly increased problems deriving from the effect of the magnetic fields in our surrounding area. Magnetic fields from the conduits, transformation stations and the high-voltage fixtures pose a significant source that effects the environment.**

This paper will depict the measuring of the magnetic fields in the area of an energetic power transformer by the Finite elements method FEM 4.2 at normal loads. With this method the distributions of magnetic inductions and the magnetic field in immediate proximity of a transformer, will also be shown.

Key words: *power distributional transformer ETN 630, Finite elements method, magnetic induction, magnetic field.*

I. INTRODUCTION

In the past twenty years there has been growing concern among the scientific, but also with the general public, about the influence of magnetic fields on electrical devices that surround us, precisely from external and internal sources. The external sources are transmission lines, transformers, underground cables, substations and earthing systems.

The energy distribution transformers are the most widely used in the distribution network of power systems. Hence it is particularly important to make different types of analyzes that would improve the basic characteristics, and thus reduce their harmful impact on the immediate environment and on the people.

This paper will show the analysis of the power transformer type ETN 630. The distribution of the electromagnetic field in all domains of the transformer will be calculated, and the magnetic induction calculation will be performed.

As a tool that made magnetic analysis possible in this paper is the applicable software package FEM 4.2 which is specialized software for analysis of this type. The algorithm which is the basis of the software, performs calculations using the method of finite elements.

II. CHARACTERISTICS OF MATERIALS OF

[1]Emilija Sarafska is with the Municipality of Bitola bul. 1 Maj 61 7000 Bitola, Macedonia,

E-mail: emijov@yahoo.com

[2]Blagoja Arapinoski, and Vesna Ceselkoska is with the Faculty of Technical Sciences, Makedonska Falanga 33 7000 Bitola, Macedonia, E-mail: b.arapinoski@gmail.com

THE POWER ENERGY TRANSFORMER AND FEM 4.2 CALCULATION

The power transformer ETN 630, which in the given paper is a subject to analysis, has initial sizes and nominal data given by the test protocol of an already produced transformer of this type, which is quite present in these areas and is produced by the EMO factory, Ohrid, R.N. Maceedonia. The nominal data of the analyzed transformer are : $S_n = 630kVA$,

$U_{n1} = 10000V$, $U_{n2} = 400V$, $I_{n1} = 36.373A$,

$I_{n2} = 909.3A$ .The analysis will be done with the software package FEM 4.2, which is a specialized program for design, modeling and analysis of electrical machines and appliances. The basis on which the program is based is the finite element method (Finite Elements Method).

III. TRANSFORMER MODEL DISPLAY IN FEM 4.2.

The complete look of the model of the energy transformer (its geometry on the magnetic circuit, the windings and the names of the materials constituting the unit blocks) are given in Figure 1. With the given boundary conditions for this model, an electromagnetic analysis with FEM 4.2 can be made.
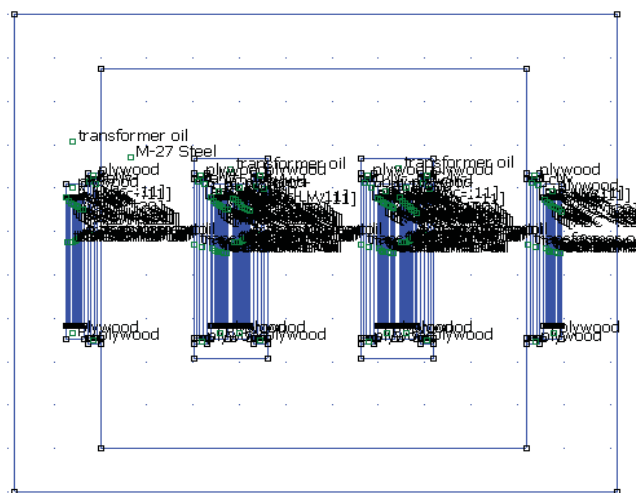
Fig. 1: Full geometry of the transformer with the input characteristics of the materials

IV. THE EQUATIONS OF A TRANSFORMER WITH A FINITE ELEMENT METHOD

After defining the geometry of the transformer and the input of the materials with their characteristics, a solver mod with FEM 4.2 is starting. This option of the program performs automatic finishing of the finite elements network through the entered parameters.

Figure 2 shows the finite element mesh over the geometry of a three-phase power transformer. It contains 34590 nodes and 68828 triangular finite elements.

By activating the "solver" fkern.exe, calculations of the equations that describe the transformer are performed. Calculations are repetitive iterative and more iterations are made of which each subsequent is faster than the previous one, since it can start with a previous solution that is close to the actual one. In the actual analysis, 12 Newton assays are carried out, for which approximately a time of about two minutes takes time with a computer with standard domestic conditions. The time for calculations depends on the number of nodes and triangular finite elements, that is, the dimension of triangular finite elements.
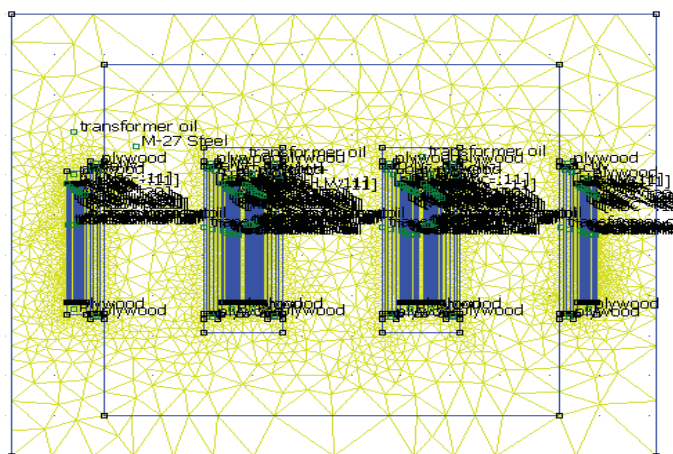
Fig. 2: Finite element mesh of power transformer ETN 630

In the electromagnetic analysis of the transformer model, in order to obtain a good representation of the solution for the magnetic flux distribution according to the finite element method, it is best to draw magnetic lines. They are lines in which the magnetic flux is established in the transformer geometry.

When the magnetic lines are close to each other, then the flux density is high. In FEM 4.2, magnetic lines are drawings of the threaded contours of the magnetic vector potential A for planar problems such as in the concrete case, or threaded contours for 2 for axisymmetric problems. For this specific case, 20 magnetic bullets are shown, as in Fig. 3, and their number can be larger and smaller.
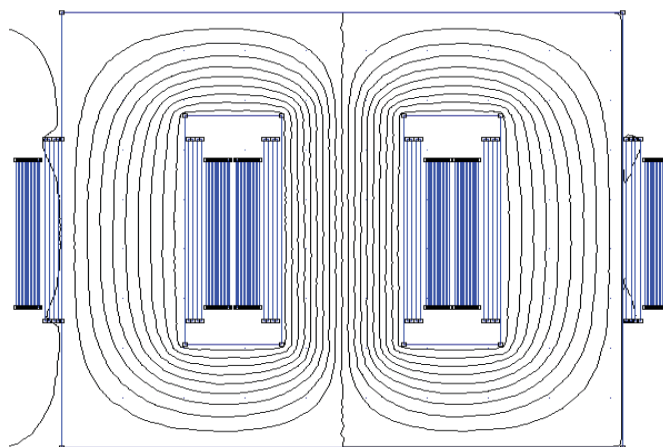
Fig. 3: Distribution of the magnetic flux in the transformer

V. CALCULATION OF THE CHARACTERISTIC MAGNETS OF THE TRANSFORMER WITH FEM 4.2

Calculations with FEM 4.2 allow us to obtain a visual overview, and in addition mathematically determine the values of the distribution of the magnetic induction and the magnetic field strength in certain regions of the analyzed distribution transformer.

In Figures 4, the distribution of magnetic induction is shown. According to the colors and values themselves, it can be noticed that at a certain point in the middle has the highest values of the magnetic induction and the magnetic field strength. The more we move away from the middle of the transformer, the lower values are obtained in relation.
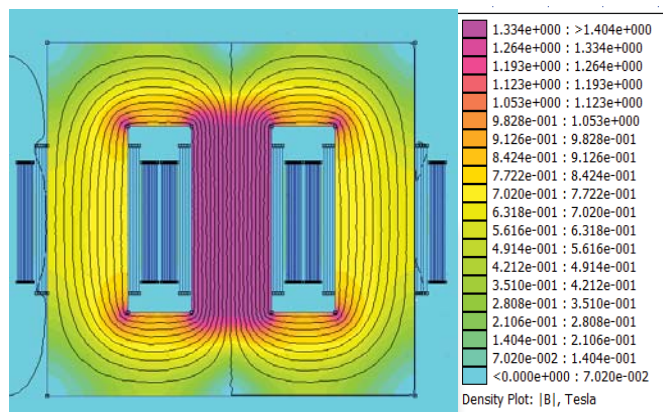
Fig. 4: Distribution of magnetic induction in different regions of the transformer

According to this image and values, it can be noticed that at the point of the middle the value of the magnetic induction module is $|B| = 1.39935$ T, and the size of the magnetic field is $|H| = 440.33$ A / m.

In order to make a comparison with the measurements, we will take a few points in which the measurement will be made. In addition to the measuring point in the middle, the values of the magnetic induction and the strength of the magnetic field

will be measured in 3 points: between the windings and two more in the outside of the windings.

In Fig. 5 the values measured in point 1 are given, between the windings, in the middle with a yellow color, where the value of the magnetic induction module is $|B| = 0.712563$ T, and the size of the magnetic field is $|H| = 82.411$ A / m. The absolute values of both parameters fall.
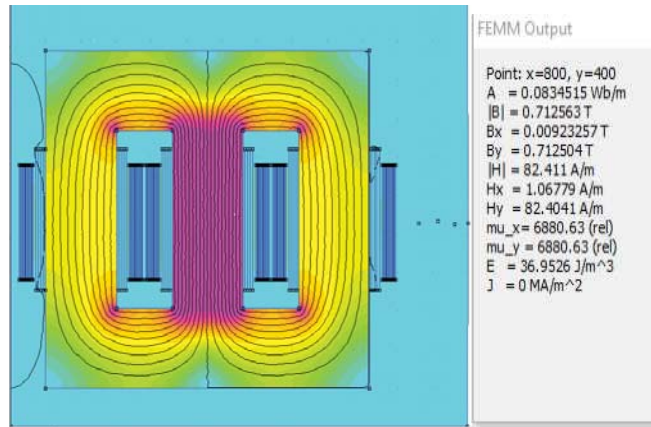
Fig.5 Measured values of $|B|$ and $|H|$ in point 1

Figure 6 will show the values of the parameters measured in point 2, which is closest to the windings, but still outside of them. The value of B is still decreasing and is $|B| = 0.00441$ T, and the absolute value of the strength of the magnetic field is $|H| = 3509.36$ A / m.
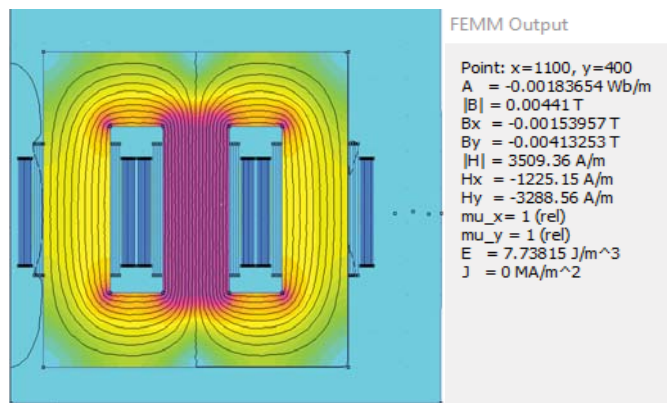
Fig.6 Measured values of $|B|$ and $|H|$ in point 2

The measurement values in points 3 are located on the edge of the transformer. The value of the magnetic induction B decreases here and its value is $|B| = 0.00108225$ T, while the absolute value of the strength of the magnetic field is $|H| = 861,227$ A / m.
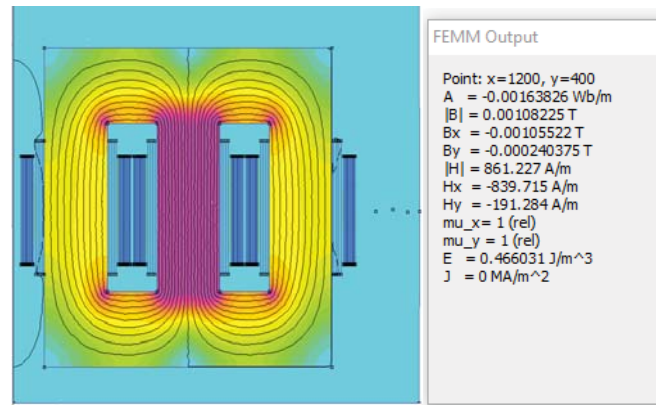
Fig.7: Measured values of $|B|$ and $|H|$ in point 3

VI. CALCULATION OF THE CHARACTERISTIC MAGNETS OUTSIDE OF THE TRANSFORMER WITH FEM 4.2

For measuring the value of the magnetic induction B and the value of the magnetic field strength H in the vicinity of the corresponding transformer, we used the 3D EMF TESTER measuring device, Model: EMF-828 shown in Fig. 8.

Fig.8 3D EMF TESTER Model: EMF-828

The measurement was performed at several measuring points on two sides of the transformer, one of which is high voltage. The measurement results are given in Table 1.

Measure points	Bx (µT)	By (µT)	Bz (µT)	$|B|$ (µT)
Side 1 (highvoltage) On the door	0,75	0.12	0,28	0,799
20 cm from the door	0,31	0,05	0,24	0,395
40 cm from the door	0,80	0,08	0,10	0.810
Side 2	0.95	2,8	0,65	3,023
10 cm from the door	0,66	1,70	0,62	1,926
20 cm from the door	0,54	1,23	0,4	1,402
30 cm from the door	0,31	1,01	0,27	1,09

Table 1: results from external measurements

However, it must be taken into account that the field measurements are carried out without knowing the thickness of the sheet metal and the internal parameters of the transformer.

The check was made with FEM 4.2 in order to determine whether the values of the magnetic induction at least roughly coincide.

Comparing the values obtained with the measurements made by FEM 4.2, it can be concluded that the calculations obtained by the program package do not deviate largely from the field analysis. It follows that the use of the FEM 4.2 software program greatly facilitates the analysis of the magnetic fields in and around the transformer.

VII. Conclusion

From this research and completed magnetic analysis of the model of the energy transformer type ETN 630, the following conclusion may be adopted:

1.With this transformer model, the magnetic values typical of this transformer roughly match those obtained by computing time using the FEM 4.2 software. by finite element method. This manner of analysis offers a quick and easy way of displaying the magnetic field and magnetic induction in all sections of the transformer as well as outside it.

2.This type of transformer analysis with FEM 4.2 offers the ability to test the magnetic impact and radiation near a particular transformer. Measurements in the future can help improve the characteristics of power transformers in order to reduce the radiation near the transformer.

References

[1] B. Arapinoski, M. Popnikolova Radevska, "*Electromagnetic and thermal analysis of power distribution transformer with FEM*" ICEST 2010, Ohrid, R.Macedonia 2010.

[2] Dragutin. M. Veličković: "Electromagnetics" - First Council, Faculty of Electronic Engineering, Čuperk Plavi, Science, Nis 1994.

[3] David Meeker: Finite Element Method – Version 4.2. Useres manuel, September 22, 2006

[4] Mateja Turkalj: Biological influence of electromagnetic radiation and its protection measures - Karlovac university of applied sciences safety and protection department, Karlovac 2018

[5] Branislav Vulević: Elektromagnetska polja u životnoj sredini - JP „Nuklearni objekti Srbije" u Vinči Belgrade, 10.2014

[6] Vesna Ceselkoska: Mutual impact of electromagnetic fields and people – MACO CIGRE. Ohrid 10.2007

[7] Vesna Ceselkoska: Contribution to research related to the influence of electromagnetic fields on the human body. Doctoral dissertation - Skopje

Case Study: Energy Audit of the High School Dorm "Mirka Ginova"- Bitola

Vladimir Mijakovski[1], Vangelce Mitrevski[2] and Tale Geramitcioski[3]

Abstract – **Measures for increasing of energy efficiency in buildings are closely related to Energy Audit. Faculty of technical sciences – Bitola (FTSB) is one out of five companies that were chosen as educators for training of energy auditors. High School dorm "Mirka Ginova" in Bitola is the only state-owned dorm in the city. Preliminary energy audit for the nearby building of the dorm was performed as a part of the training for energy auditors. The calculations were performed by using of ENSI© EAB software.**

Keywords – **energy audit, dorm, energy class, ENSI software.**

I. INTRODUCTION

Following recent adoption of EU regulative in the area of energy auditing in the country, [1,2], the first step was to train energy auditors with a purpose of obtaining licenses for energy auditing. One of the institutions licensed for training of Energy Auditors is the Faculty of Technical Sciences in Bitola. In the course of this training, the building of the nearby High School dorm "Mirka Ginova", was used as an example for energy auditing with determination of its energy class using ENSI© EAB software.

The building of the dorm is located in the south-eastern part of the city of Bitola. The object does not have attached building to it, located in averagely urbanized part of the city, next to the city park, bus station and railway station. It was built in 1960 and significant reconstruction and extension took place in 1994. Main entrance of the building is on the south-western side (Fig. 1).

Fig. 1. Location of the high school dorm "Mirka Ginova" in Bitola

II. REQUIRED DATA FOR ENERGY AUDIT

[1]Vladimir Mijakovski is with the Faculty of Technical Sciences at University "St. Kliment Ohridski", Bitola 7000, Macedonia, E-mail: vladimir.mijakovski@tfb.uklo.edu.mk

[2]Vangelce Mitrevski is with the is with the Faculty of Technical Sciences at University "St. Kliment Ohridski", Bitola 7000, Macedonia.

[3]Tale Geramitcioski is with the is with the Faculty of Technical Sciences at University "St. Kliment Ohridski", Bitola 7000, Macedonia.

Dormitory "Mirka Ginova" in Bitola is educational institution within the Ministry of Education and Science of the Republic of Macedonia, student standard department. The building is mainly divided in 2 parts: north and south part. South part consists mainly of bedrooms, while in the north part, the kitchen, dining room and administration offices are located. South part consists of basement, three floors with wooden roof construction covered with metal sheet roof, while the northern building has basement and two floors also covered with metal sheet roof. The capacity of the dorm is 270 high school students and 26 employees. In the summer months, the dorm is open to accommodate guests of different events in the city. In this period of the year, an average of 100-150 guests are staying at the dorm. Total net area of the building is 3364 m², while total net volume is approx. 9420 m³.

Last reconstruction of the building consisted of partial replacement of external windows and carpentry and took place in the year 2010.
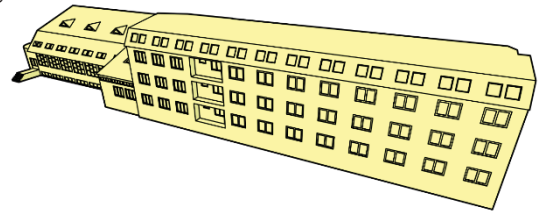
Fig. 2. Appearance of the building from the south-west

Part of the other relevant data for the energy audit, required by the legislative, are given in the following table:

TABLE I
PART OF THE DATA RELEVANT FOR ENERGY AUDIT OF DORMITORY'S BUILDING

Characteristics of the building construction	Material (concrete, brick, hollow brick)	Total thickness [cm]	Thickness of the thermal insulation layer [cm]	Area of the construction [m²]	Heat transfer coefficient U [W/m²K]
External wall NORTH	Concrete	36	5	31,14	0,914
	Brick	43		297,61	1,16
External wall SOUTH	Brick	43		252,40	1,16
	Hollow brick	27		30,22	1,6

Prior to entering of data in ENSI© EAB software, a detailed calculation of areas of all surfaces (external building envelope) as well as heat transfer coefficient for all materials was performed. In the following figures, example of calculated

areas for building's south façade and cross-section of one type of external wall and roof are shown:

At the end, we grouped external walls and windows in three groups according to building construction and heat transfer coefficients.

For the heating of building, hot water radiator heating system with forced circulation (with pump) is used. Heating installation is of a two-pipe system with lower horizontal branching. Two pumps are used for circulation of heating media (water). Boiler house consists of three hot water boiler connected in parallel, with a total of approx. 1100 kW installed heat power. Light oil is used as fuel. As part of the energy audit, a measurement of flue gases emission from one of the boilers was also taken.

Electrical equipment in use consists of more than 10 electric heaters (with total installed electric power of 54 kW) that are used prior to/after heating season (before 15.10 or after 15.04), electric appliances in the kitchen (total installed electric power of 116 kW), electric appliances in the laundry (around 53 kW), electric boilers with installed electric power of 102 kW, 15 personal computers copier machines, 11 air conditioning units (split system) etc. There are also around 180 fluorescent lightning tubes with electronic ballast installed for lamination and 52 light bulbs with total electric power of 5,2 kW.

For the purpose of energy audit preparation, detailed invoices – bills for electric energy and water consumption were collected from the accounting department.

III. CONCLUSION

Following recent adoption of EU regulative in the field of Energy Auditing of buildings, the Faculty of Technical Sciences in Bitola is one of country's five licensed training centers for energy auditors, [5]. In the scope of training, practical part, the building of dormitory "Mirka Ginova" was taken as an example and general energy audit was performed on it. Calculations were performed using ENSI© EAB software for quick energy performance calculations.

The results from the calculations categorized the building of the High school dorm "Mirka Ginova" – Bitola as class "E" building. Calculated value of energy consumption is 158,4 kWh/(m²a).
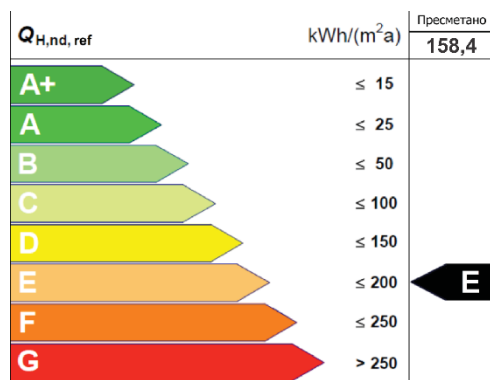
Fig. 4. Calculated energy class of the building

According to the National legislative, [2], all buildings undergoing 'substantial reconstruction' must reach at least "D" energy class.
In the example of preliminary energy audit of the dormitory "Mirka Ginova" – Bitola building, the proposed measured would include:

- Thermal insulation of all external walls in order to reach maximum allowed U-value of 0,35 W/m²K;
- Partial replacement of windows and carpentry in order to reach maximum allowed U-value of 1,7 W/m²K;
- Installation of additional thermal insulation for the roof in order to reach maximum allowed U-value of 0,25 W/m²K;
- Replacement of one of the hot water boiler running on light oil fuel with high efficiency hot water boiler running on wood pellets.
- Replacement of light bulbs and fluorescent lighting tubes with LED lights;

Implementation of these measures would 'raise' building's energy class to "C".

Return on Investment (ROI) period for implementation of these measures was also calculated and it ranges from 2 years (lights replacement) up to 5,5 years (replacement of windows and corresponding carpentry)

REFERENCES

[1] Rulebook for energy auditing, official gazette of RM, No. 94/2013.
[2] Rulebook for energy characteristics of buildings, Official gazette of RM, No. 94/2013.
[3] User guide for ENSI© EAB Software, Version 8.1, May 2009, Energy Saving International AS, Norway.

Electric Field in the Environment of 110kV Power Line and its Impact on Biological Systems

Blagoja Arapinoski[1], Mirka P. Radevska[1], Metodija Atanasovski[1] and Mitko Kostov[1]

Abstract − The paper will present the influence of the electric field with an industrial frequency that originates 110 kV power line of the man who is in his immediate environment. The human model is composed of parallelepipeds representing the individual parts of the body and is located adjacent to the power line. The distribution of the electric field and the potential in the vicinity of the 110 kV power line is presented without the presence of a man under the power line obtained by the finite element method. The work is supplemented with certain standards regarding the maximum allowed values for the strength of the field and the allowed density of streams permeated into the human body.

Keywords − electric field, potential, biological systems, induced currents.

I. INTRODUCTION

With the rapid development of human society and the more advanced technological development, there is a need to increase the amount of electricity, which in turn leads to an increase in electromagnetic fields at the level of urban and working environment. In recent years, there is increasingly the opinion that electromagnetic fields derived in different ways, from energy lines to mobile phones, cause various diseases in people, including the most severe. In relation to this issue, due to insufficient knowledge, there is still no single position. A better insight into the real situation shows that potential health hazards, from electromagnetic fields, should not be ignored.

This paper deals with the aspect of the harmful influence of the electric fields in order to systematize the knowledge for estimating the electric field in the surrounding of the three-phase conductors. Knowing the values of the electric field and its distribution under and around the three-phase power lines is important for the protection of workers working in substations and the population living near the power lines.The paper contains an analysis of the electric field that originates 110 kV power line per person located in its immediate surroundings. The human model is composed of parallelepipeds representing the individual parts of the body. The method of finite elements is used to obtain the distribution of the electric field under and in the environment of the power line without and in the presence of man.The work is supplemented with certain standards regarding the maximum allowed values for the strength of the field and the allowed density of streams permeated into the human body.

[1]Blagoja Arapinoski, Mirka Popnikolova Radevska, Metodija Atanasovski and Mitko Kostov are with the Faculty of Technical sciences at University of Bitola, st. Makedonska Falanga 33, Bitola 7000, Macedonia.

II. MATHEMATICAL MODEL

The electric field in the vicinity of the power line is caused by the electrical charge of the phase conductors and the protective ropes. Phase conductors and protective ropes are modeled as infinitely long cylindrical conductors with a cross-section with finite dimensions. Considering that the earth is at zero potential, a mirror image method should be used to calculate the electric field generated by this system. Maxwell's postulate in a differential form is:

$$div\vec{D} = \rho \tag{1}$$

The connection between the vector of the electric field strength and the electrical induction is given by the expression:

$$\vec{D} = \varepsilon\vec{E} \tag{2}$$

The vector of the electric field strength expressed through the electric potential is:

$$\vec{E} = -grad\varphi \tag{3}$$

By connecting relations (2) and (3), the following partial differential equation is obtained:

$$-divgrad\varphi = \rho \quad \text{Or:} \quad -\varepsilon\Delta^2\varphi = \rho \tag{4}$$

Equation (4) can be applied to inhomogeneous regions and provides a solution for the electric potential and the strength of the electric field, under certain boundary conditions.

III. DISTRIBUTION OF ELECTRIC FIELD IN THE ENVIRONMENT OF 110kV POWER LINE

In Fig. 1 presents the cross-section of 110 kV power line with the corresponding dimensions of the distances between the phase conductors and protective ropes [1].
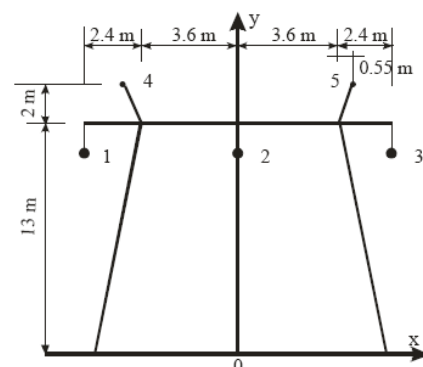
Fig. 1. Cross section of the conductors of the 110 kV power line

In Table 1, the dimensions of the 110kV transmission line are given, as well as the values of the voltage and current through the conductors.

Table 1. Coordinates and radius of phase conductors and protective ropes of 110 kV power line

	First phase	Second phase	Third phase	Rope 1	Rope 2
n	1	2	3	4	5
$x[m]$	-6	0	6	-4.1	-4.1
$y[m]$	12	12	12	15	15
U_n	$U_f e^{j0}$	$U_f e^{j\frac{2\pi}{3}}$	$U_f e^{-j\frac{2\pi}{3}}$	0	0
I_n	$I_f e^{j0}$	$I_f e^{j2\pi/3}$	$I_f e^{-j2\pi}$	0	0
a[mm]	8.74	8.74	8.74	5	5

The distribution of the electric field in the vicinity of the 110 kV power line in the absence of a person, obtained using the FEMM 4.2 program package is shown in Fig.2. of the power line.
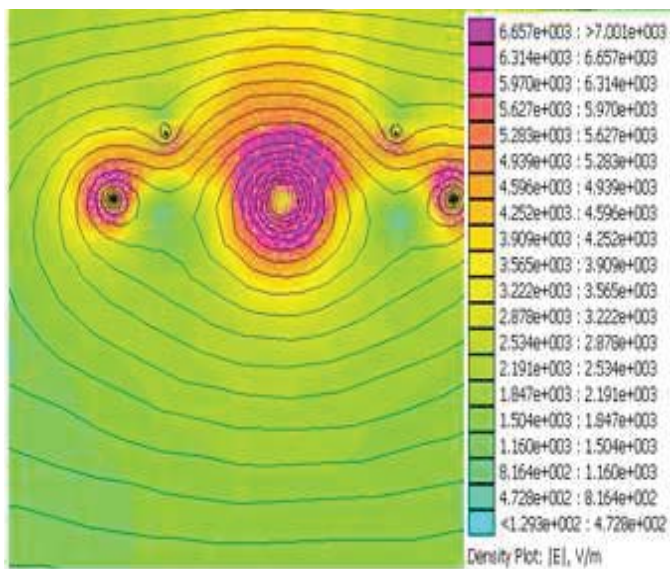
Fig. 2. Distribution of the electric field in the environment of the 110kV power line in the absence of a person

Fig.3. shows the distribution of the field per line vertical placed under the middle phase conductor perpendicular to the surface of the earth and in Fig.4. the distribution of the electric field per line parallel to the ground surface at height h = 1.5m and the length of x = -3m to x = 3m.
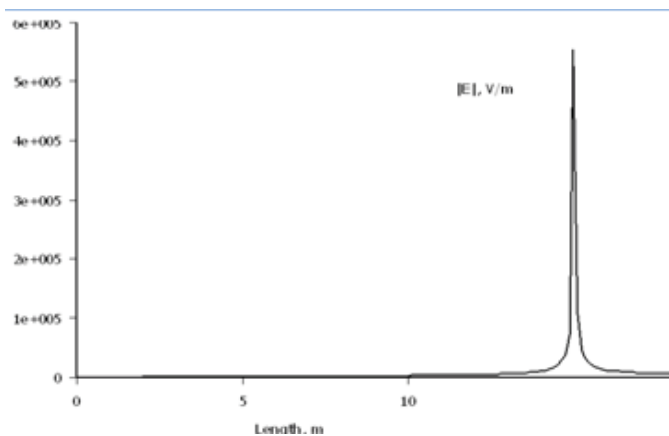
Fig. 3. Distribution of the electric field per line vertical placed under the middle-phase conductor perpendicular to the surface of the earth
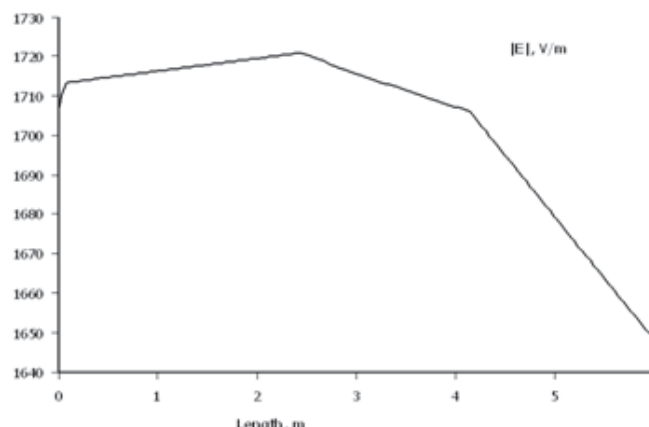
Fig.4. Distribution of the electric field per line parallel to the ground surface at height h = 1.5m and with a length of x = -3m to x = 3m

IV. DISTRIBUTION OF ELECTRICAL FIELD IN THE ENVIRONMENT OF 110KV POWER LINE IN THE PRESENCE OF MAN

Human tissue in an electromagnetic view acts as a non-homogeneous, paramagnetic semiconductor environment, whose specific conductivity σ and relative dielectric permeability ε_r very intensively with the frequency and is located in a wide range whose boundaries define the values of these quantities for fat tissue and blood. The dependence of the specific conductivity of the tissues of the frequency is such that with the increase in the frequency, the tissue conductivity grows, and the dielectric permeability decreases with the increase in the frequency. For an industrial frequency of 50Hz, for fat tissue $\varepsilon_r = 10^5$ and $\sigma = 0.02$S/m, and for the blood $\varepsilon_r = 10^6$ and $\sigma = 1$S/m.

The magnetic permeability of biological tissues is practically equal to the magnetic permeability of the free space $\mu_0 = 4\pi \cdot 10^{-7}$ H/m (the exception is the immediate proximity of microscopic magnetosomes), so that the biological tissues are non-magnetic.

When penetrating the electric field through human tissues, the intensity of the field is significantly reduced due to the electromagnetic properties of the biological tissues.

In order to simplify the analysis, the human body model has been chosen to be homogenous with average values for electromagnetic features. In the paper, using the FEMM 4.2 method, finite element method will show the shape of the potential and distribution of the electric field caused by an overhead 110 kV power line in the case when a human body model is placed under the very power line with hands lowered by the body.
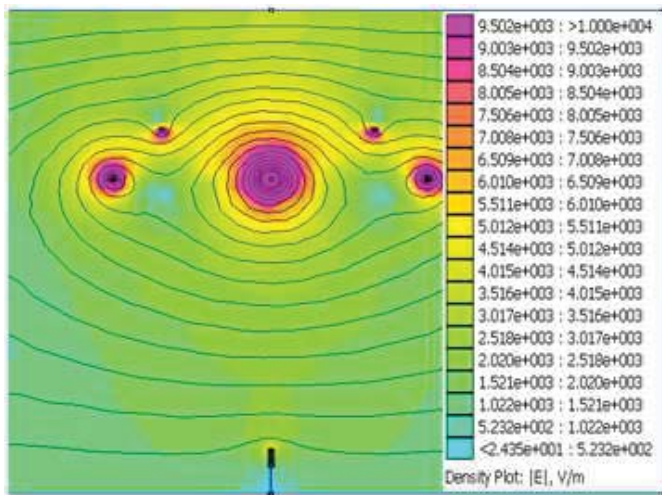
Fig.5. Distribution of the electric field below the 110 kV power line when a man with his hands is attached to the body under the very power line.

In Fig.6. is given detail from the previous image to show that the presence of a man under the power line causes deformation of the lines of the field. The intensity of the electric field around the head has a significantly increased value than that for the field without the presence of the human body, which is a potential danger.
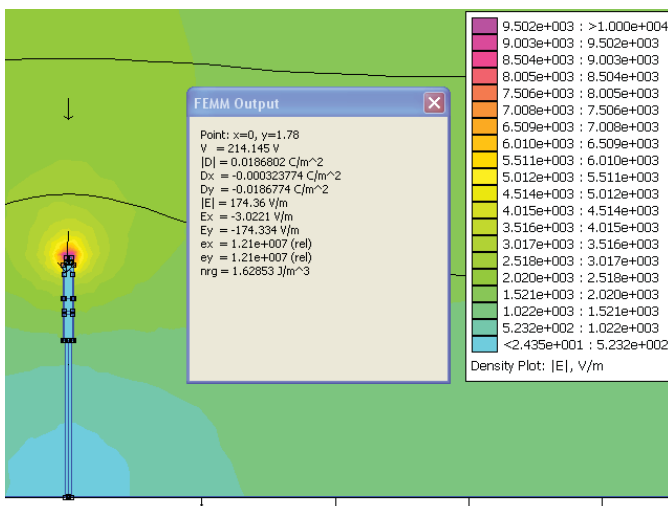
Fig.6. Distribution and intensity of the electric field in the environment of a man-model placed under the 110 kV power line in the presence of a man (detail from the previous picture)

Fig.7. Shows the distribution of the field per line vertical placed under the middle phase conductor perpendicular to the surface of the earth, at height h = 1.5m and the length of x = -3m to x = 3m when a human model is placed under the transmission line.
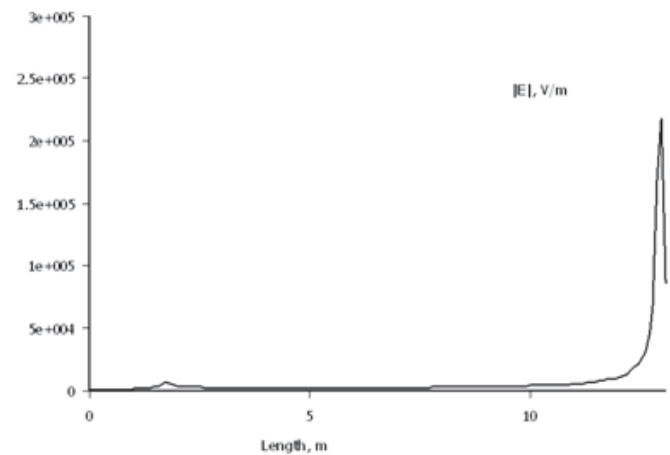
Fig.7. Distribution of field per line vertical placed under the middle phase conductor perpendicular to the earth's surface in the case when a human model is placed under the power line.

V. REGULATIONS AND STANDARDS OF THE AREA OF ELECTROMAGNETIC IMPACTS

The presence of electromagnetic fields with a wide range of frequencies in the work and environment and the inability of the human sensory organs to register them make these fields potential pollutants to the working and living environment.

With the problems of the ecology of the power lines of electricity in the high voltage circuits, a large number of institutions in the world are engaged.

The standards and recommendations of these institutions are the only criteria for harmfulness, while there are differences in permitted levels and time of exposure. [3]-[5].The status and background of each standard are related to several clear categories: there are real differences, which are dependent on the legal arrangements of individual states and factors influencing the establishment of standards. Political and social factors can play an important role as well as scientific factors. Other constraints are based on the position and structure of buildings in different countries.

International standards in this area are produced by the International Electrotechnical Commission IEC (International Electrotechnical Commission), within its Technical Committee.

Regional standars for Europa are produced by CELENEC (CLC), subclause SC111.

Recommendations for the permitted levels of exposure to electrical and magnetic fields have also been reported by the International Radiation Protection Association (IPRA) and the World Health Organization (WHO).

Limits of exposure to electromagnetic fields with industrial frequency according to the IPRA recommendations are given in Table 2.

In some countries (USA, Russia, Austria, Poland), by law or by other regulations, certain protection zones - "zones of safety" or "right-of-way zones" are located in the right-of - way directions under the power lines limited use of land for making living facilities, hospitals, schools, children's facilities, etc. Here, the upper permissible field strength of the zone or the edge of the zone is usually determined (usually from 1 [kV / m] to 10 [kV / m]) or the width of this zone is below the power lines (33 [m] to 115 [m]), depending on the voltage level and the state.

Florida was the first state to introduce magnetic induction into the ROW zone as a criterion for hazard assessment: in the vicinity of 500 [kV] power line, 20 [μT] and in the vicinity of 230 [kV] power line, 15 [μT]. Japan regulates the minimum height of the conductors below the power line of the earth, the minimum distance of the conductors from the objects and the strength of the electric field in the power line area. In many countries, there are no such regulations, but it is determined that the strength of the electrified and magnetic field under the power lines must not exceed the limits determined by the national standards.

Table 2 IPRA recommendations for exposure limits with electromagnetic fields with an industrial frequency

Characteristics of elongation	Electric field (kV/m)	Magnetic induction (mT)
In the workplace		
Full working day	10	0.5
Short-term	30 (E=80/ t)	5(max 2 hours)
Extremes	-	25
Population in general		
24 hours a day	5	0.1
Two hours a day	10	1

VI. CONCLUSION

Low-frequency electromagnetic fields are present everywhere in our environment. Their interaction with humans and other living organisms is expressed through direct action and through induced currents inside the tissues of the action of the external field. The importance of the properties of biological materials from which the tissues are made in the human body is great in determining the mutual influence. An important aspect of the interactions of NF fields that act on the human body in the air is that they do not heat the tissues more than basal metabolic processes in the body.

REFERENCES

[1] D. Petkovic, D. Krstic, V. Stankovic. "The effect of electric field on humans in the immediate vicinity of 110 kV power lines" Facta Universitatis, Vol 3, N^0 1,2006, pp 63-72.

[2] V.Cheshelkoska. Contribution to research related to the influence of electromagnetic fields on the human body. Doctoral dissertation, Skopje, 2003.

[3] D.Velickovic, S. Ilic, V. Ceselkoska, A. Milovanovic. "Calculation of the earth's atmospheric electric field in the living and working environment" ETRAN 2004. Cacak, June 6-15, 2004

[4] D. Velickovic, V. Ceselkoska. "Electric field distribution in the environment of the human in the external electric steady field" 5th -International symposium EL-TEX 2002. Lodz, 14-15.11.2002

[5] C.Polk, E.Postow. Handbook of biological effects of electromagnetic fields. CRC Press, New York, 1996.

[6] M. Stuchly. Biological Effects of Electromagnetic Fields International Journal of Bioelectromagnetism. Vol. 4, No.2, 2002

[7] D.Deno. "Current induced in the human body by high voltage transmission line electric field measurement and calculation of distribution and dose" IEEE Transactions on Power Apparatus and Systems, Vol. PAS-1996, No.5, 1997.

[8] 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.

[9] 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.

[10] B. Arapinoski, M. Popnikolova Radevska, *"Electromagnetic and thermal analysis of power distribution transformer with FEM"* ICEST 2010, Ohrid, R.Macedonia 2010.

[11] 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.

[12] B. Arapinoski, M. Popnikolova Radevska, D. Vidanovski *"FEM Computation of ANORAD Synchronous Brushless linear motor"* ELMA 2008, Sofia – Bulgaria.

[13] M. Popnikolova Radevska: "Calculation of Electromechanical Characteristics on Overband Magnetic Separator with Finite Elements", ICEST 2006, p.p. 367-370, Sofia, Bulgaria 2006.

Cyber Security Protection and Defence Measures in the Electricity Transmission Networks in South-East Europe

Aleksandra Krkoleva Mateska[1], Petar Krstevski[1], Stefan Borozan[2]

Abstract – **The paper describes the practices for protection of information, control systems and assets by transmission system operators (TSOs) in the region of South-East Europe (SEE). Based on a survey undertaken among the TSOs of the SEE region, the paper indicates state of the art practices and future developments related to critical infrastructures in electricity transmission networks.**

Keywords – **transmission systems, critical information infrastructure, cyber security.**

I. INTRODUCTION

The operation of power systems relies on both legacy systems and new technologies. The existing electricity infrastructure is combined with sophisticated control systems and intelligent components with bi-directional communication capabilities. The information and communication systems represent an overlay to the conventional electricity systems and allow real time control in the operation of generation, transmission and distribution of electricity [1]. As the implementation of information systems and technologies in power systems becomes a necessity, it also increases the risks of cyber security threats and requires implementation of adequate protection measures that should reflect the multi-actor environment of the contemporary power systems [2].

The cyber-attacks in electricity systems potentially endanger the electricity supply chain, threaten other essential services because these systems cannot be disconnected as easily as other information technology systems. Furthermore, there is a growing interdependence among sectors and systems that enables threats to become a cross-border issue not only in the EU, but in other neighbouring countries.

The general challenges to the energy sector with regards to cyber security include [3]: grid stability in a cross-border interconnected networks; protection concepts reflecting current threats and risks; handling of cyber-attacks in Europe; effects by cyber-attacks not fully considered in the design rules of an existing power grid or nuclear facility; introduction of new highly interconnected technologies and services; outsourcing of infrastructures and services; integrity of components used in energy systems; increased interdependency among market players; availability of human resources and their competences;

and constraints imposed by cyber security measures in contrast to real-time/availability requirements.

In fact, all the above-mentioned challenges are applicable to electricity systems, which is not the case for other energy sectors. This only implies that the operators of interconnected electricity systems with legacy and next generation technologies, operating in multi-actor environment, need to reassess their approach to cyber security issues, following the recommendations of relevant national authorities and international organizations.

This paper describes practices for protection of information and control systems and assets by transmission system operators (TSOs) in the region of South-East Europe (SEE). The investigation is based on a survey undertaken within the framework of the CROSSBOW project [4] and encompasses the practices related to information and control systems security of the TSOs of Bulgaria, Croatia, Greece, Romania Bosnia and Herzegovina, Montenegro, North Macedonia and Serbia. For confidentiality reasons, the eight TSOs that participated in the survey are anonymized, thus in the paper, instead of the company name, TSOx is used, where x is a number from 1 to 8.

II. INFORMATION SYSTEM SECURITY LEGISLATION AND GOVERNANCE

A. Legislation and Practices in SEE

The above described challenges have to be addressed through transposition of relevant legislation, including the Directive (EU) 2016/1148 (NIS Directive) [5] and the Directive 2008/114/EC (Critical Infrastructure Directive) [6] as well as implementing numerous measures at utility level. The observed SEE region consists of countries that are EU Member States (MSs) as well as countries that are Contracting Parties of the Energy Community (EnC) and members of the Western Balkans 6 Initiative [7]. Therefore, the countries of the region represent a uniquely varied part of Europe and for that reason, may serve as an example of extension of frameworks, rules and practices that bring European Union (EU) closer to its adjacent regions.

The NIS Directive aims to increase the overall level of security of networks across the EU and to build a systematic approach in counteracting the possible threats to networks and information systems. Its transposition should enable the establishment of Computer Security Incident Response Teams, adoption of national cyber security strategies by each EU MS, as well as designation of one or more national competent authorities and a single point of contact on the security of network and information systems [5]. The NIS Directive

[1]Aleksandra Krkoleva Mateska, Petar Krstevski, are with the Ss Cyril and Methodius University in Skopje, Faculty of Electrical Engineering and Information Technologies, Rugjer Boskovic No. 18, Skopje, Republic of North Macedonia, E-mail: {krkoleva, petark}@feit.ukim.edu.mk.

[2]Stefan Borozan is with Elektrodistribucija DOOEL, 11 Oktomvri No. 9, Skopje, Republic of North Macedonia, E-mail: s.borozan2@gmail.com.

foresees minimum common planning requirements, exchange of information and common security requirements for Operators of Essential Services (OESs) and Digital Service Providers (DSPs) [5]. From the aspect of electricity systems, the NIS Directive affects the operational activities of TSOs, distribution system operators (DSOs) and electricity suppliers. These entities can be considered as OESs and therefore, are obligated to fulfil all the requirements related to information systems security and incident notification.

The Critical Infrastructure Directive fosters EU cooperation in identification and designation process for potential European critical infrastructures and regular review of the process. It sets the requirements for operator security plan and security liaison officer for each European critical infrastructure, identification of their critical assets and security solutions [6].

The transposition of these Directives is not obligatory for the Western Balkan (WB) countries. The investigations done in [2] show that WB countries lack a strategic and cooperative approach in identification of critical infrastructures, their protection, as well as providing necessary level of protection of networks and information systems. They all face the same potential risks as EU MSs, hence, there is a substantial need for these countries to develop legislation that is compliant with the relevant EU legislation. The Energy Community Secretariat has started initial activities to overcome regulatory gaps inside the EnC, as well as towards the EU. Among the first activities within the envisaged tasks are to identify current legal framework and if some of the measures from the above discussed Directives are already implemented by the EnC Contracting Parties [8].

B. Utility Information Systems Security Governance Model

The adoption of network security related legislation is one of the pillars for achieving functional high-level security in critical information infrastructures (CIIs). The second pillar consists of set of measures and actions to be performed by utilities, including TSOs, with the aim to complement the measures implemented on national/EU level.

The information systems security governance model presented on Fig. 1 is based on the recommendations of [9] and it shows that performing risk assessment and developing an adequate security policy based on the performed risk analysis is substantial for the security of any information system, including CIIs used by TSOs. Starting with identification of the CII whose failure impedes the essential service delivery, the OES (in this case TSO) sets up an information security system. The identification process may be operator driven or state driven, but the outcome is clear identification of these infrastructures [10].

The scope of the security measures [11] are the assets exposed to threats which when breached/failing can have a negative impact on the networks or services. In the context of TSOs, these assets include the SCADA systems, hardware, software and computer databases at national dispatching centres, communication links between the significant nodes within the high voltage transmission network, software platforms with access for end users and other information infrastructure. The overall information system security policy should be based on performing risk analyses and supported by a security management system. In fact, risk analysis is the essential tool that is used to build a consistent and adequate security policy. This policy should aim to set up the security strategy of the utility and to set references to all important information system security policies including the security accreditation process, security audit, cryptography, security maintenance, incident handling [9].
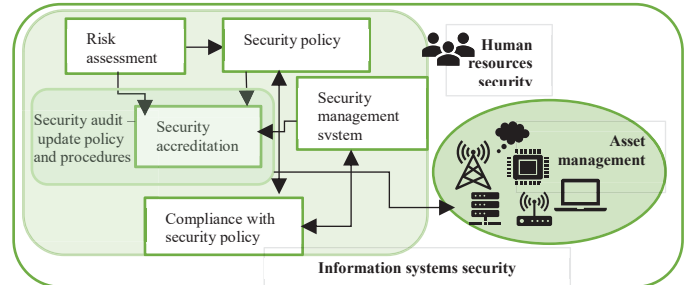
Fig. 1. Information system security, [2], [9]

The information security policy should lay out the accreditation process for the CII performed by the OES itself. The purpose of the process is to integrate the critical infrastructure in the security management system and formalize the processes. The accreditation process is used to map all the CIIs and apply adequate security measures. Within the process, a security audit to reassess the security policy and effectiveness of measures should be envisaged. The security of the system largely depends on the employees and contractors who should assume their responsibilities and on effective asset management.

III. INFORMATION SYSTEMS SECURITY PRACTICES IN THE REGION

A. Risk Assessment

Risk assessment is an essential tool for evaluating the potential cyber security threats to CIIs and assets. It is based on a pro-active approach, consisted of continuous monitoring and assessment, in a closed loop. As described in [11] the process should be performed during requirements definition, procurement, control definition and configuration, system operations, and end of system lifecycle, i.e. throughout the whole lifecycle of the infrastructure. The general risk assessment approach is to consider the possible threat, the vulnerability of the system and the consequence that the infrastructure and the system shall suffer due to the threat. The overall risk is determined using the following formula: risk = threat (probability) x vulnerability x consequence (impact). When the risk is assessed, then applying risk management techniques shall allow the utility to control, avoid, minimize or eliminate risks.

The survey on information systems security in SEE TSOs indicates that the companies are aware of the benefits of risk assessment in defining cyber security measures. However, regular performing of risk assessment is not a common

practice. Table I represents indicative summary of the results referring to risk assessment procedures. The "-" sign indicates that there are no developments related to the issues from the first column of the table, the "±" sign indicates that there are ongoing or planned activities for near future and the "+" signs show that adequate actions are implemented.

TABLE I
RISK ASSESSMENT PROCEDURES IN SEE TSOS

TSO	1	2	3	4	5	6	7	8
Perform risk assessment	±	+	+	±	-	±	+	+
Risk assessment is regular procedure	-	-	+	±	-	±	+	-
Measures implemented after risk assessment	-	+	-	±	-	±	±	+
Information security policy implemented	-	+	+	±	-	±	+	+
List of assets of critical infrastructures	+	+	+	+	-	±	+	+

As presented in Table I, half of the TSOs of the observed region perform risk assessment, but only two of them regularly. A general characteristic of the region is that the TSOs maintain lists of assets of critical infrastructures, which is essential aspect of the development of adequate security policy that refers to those assets. Although not all analysed companies implement risk assessment-based security policy, it should be noted that all of them implement security measures that are based on the assessment of the information and communication departments in the companies and common approach of TSOs in organizing their information infrastructure.

B. Protection

The set of cyber security measures consist of protective and defence measures. The protective measures encompass the information systems architecture, administration, identity and access management and maintenance. Concerning the architecture of the information systems [11], the operator should only connect equipment and/or install services and applications that are necessary for the functioning of the critical infrastructure. It should consider measures to segregate systems to avoid propagation of threats. However, when interfacing of systems is needed, the OES should apply additional measures for protection as traffic filtering (deny flows that aren't necessary for the functioning of the systems) in the CII to decrease the possibilities for a cyber-attack. A cryptography policy may be implemented to protect information confidentiality, authenticity and integrity in the infrastructure.

From the aspect of administration [11], OES should set up specific accounts for employees performing installations, configuration, management, maintenance and other system administration activities. These accounts should be used to connect to the system, after which administrator accounts should be used to perform the actual administration activities. Identity and access management include set of protective measures that minimize possibilities for unauthorized access to systems and processes related to CII. Unique accounts should be used as a part of the identification process. In addition, authentication credentials should be required for accessing systems and processes related to critical infrastructure.

Table II and III summarize the protection measures to minimize propagation of threats and of external access to the CIIs, respectively.

TABLE II
MEASURES TO MINIMIZE PROPAGATION OF THREATS IMPLEMENTED BY SEE TSOS

TSO	System architecture – propagation of threats
TSO1	Segregation, firewalls
TSO2	Implements adequate measures
TSO3	Secure protocol algorithms, advanced cryptography algorithms
TSO4	Security measures for e-mails and web traffic, sandbox solution for zero-day malware protection for network and email traffic
TSO5	Implements standard practices
TSO6	Segregation, firewalls, anti-malware, traffic filtering
TSO7	Segregation, LAN segmentation, traffic filtering, IPsec encryption
TSO8	Traffic filtering, cryptography

TABLE III
MEASURES TO MINIMIZE EXTERNAL ACCESS BY SEE TSOS

TSO	System administration – external access
TSO1	Secure VPN connections with authentication and authorization of access control
TSO2	Implements adequate measures
TSO3	VPN secure connections, personalized accounts with limited duration for specific tasks and access to dedicated isolated intranet zone
TSO4	Secure VPN connections with authentication and authorization of access control
TSO5	No reply
TSO6	Administration accounts with limited duration, complex passwords, and limited access, data backup, dedicated network for network administration
TSO7	Administration accounts with limited duration, dedicated network for network administration, authentication and encryption mechanisms
TSO8	VPN connections with authentication and authorization of access control

Table II shows that some of the protection measures are more common, as segregation, traffic filtering to control flows, firewalls and anti-malware. Similarly, as presented in Table III, the administration of critical infrastructure is usually performed using accounts that have some form of restriction (time duration limit, password protection, access restrictions). Concerning access rights, some of the end users implement (or

plan to implement) multifactor user authentication and some are using various access levels for different users.

The protection measures used for SCADA systems are summarized in Table IV.

TABLE IV
SCADA PROTECTION IN SEE TSOs

TSO	SCADA protection measures
TSO1	Segregation
TSO2	No internet access, firewalls, segregation
TSO3	Segregation, firewalls
TSO4	Dedicated firewall systems with next generation functionalities and comprehensive SCADA protocol support
TSO5	No reply
TSO6	Restricted network access for different services to one host, remote access via VPN allowed to dedicated serves only, dedicated network for system administration
TSO7	Segregation, LAN segmentation
TSO8	No internet access, closed network (segregation)

C. Defence

The defence of the CIIs includes implementation of measures for detection of threats, setting up logging systems on each CIIs to record events, as well as log correlation and analysis system. Detection of threats is the most important step in providing adequate system response. The second step is creating logs of events, especially for access, management of access rights and modifications of security policy. With appropriate log correlation and analysis system, the operator should be capable of data mining for events that have implications to security of CIIs. The summary of the results of the threat detection measures implemented by the TSOs in the observed region is available in Table V.

TABLE V
DEFENCE OF CIIs IN SEE TSOs

TSO	Threat detection measures
TSO1	Log analysis
TSO2	System control of network traffic
TSO3	Monitoring and logs
TSO4	Event or alarm logs
TSO5	No reply
TSO6	Event or alarm logs, threat detection systems
TSO7	Event or alarm logs
TSO8	System control of network traffic

It can be observed that all TSOs apply threat detection systems (mostly event logs and their analyses). However, the investigation showed that incident reporting procedures for some of the TSOs are subject of ongoing changes, which will be completed in near future. The notification procedures are under development and in two cases the ENTSO-E Wide Awareness System is implemented.

IV. CONCLUSION

The paper presents an overview of the various protective and defence measures implemented by the TSOs of the SEE region. The results show that risk assessment procedures should be implemented on regular bases and used to develop adequate security policies for the CIIs. The TSOs implement various protective measures to minimize propagation of threats or unauthorized access and to limit access to the CIIs and to protect CIIs from external parties. Stronger cooperation and regional approach to threat defence is required, especially related to notification of threats to relevant bodies in neighbouring countries. Further cooperation between TSOs and early implementation of the relevant EU legislation in the WB countries shall increase the security of CIIs in the observed region.

ACKNOWLEDGEMENT

REFERENCES

[1] M. S. Thomas, J. D. McDonald, *Power System SCADA and Smart Grids*, Boca Raton, CRC Press, 2015.

[2] V. Borozan, A. Krkoleva Mateska, P. Krstevski, R. Taleski, S. Borozan, "D3.2 Privacy and Data Protection in a Multi-Actor Environment", CROSSBOW H2020-773430, January 2019.

[3] EECSP, "Cyber Security in the Energy Sector-Recommendations for the European Commission on a European Strategic Framework and Potential Future Legislative Acts for the Energy Sector," February 2018.

[4] CROSSBOW, available [Online] http://crossbowproject.eu/

[5] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union.

[6] Directive 2008/114/EC – identification and designation of European critical infrastructures and assess-ment of the need to improve their protection.

[7] Memorandum of Understanding of Western Balkan 6 on Regional Electricity Marker Development and Establishing a Framework for Other Future Collaboration, 27 April 2017, Vienna, available [Online] https://energy-community.org/regionalinitiatives/WB6/MoU.html

[8] Energy Community, "Study on Cybersecurity in the energy sector of the Energy Community," Ljubljana, September, 2018.

[9] Cooperation Group, Reference document on security measures for Operators of Essential Services, 2018

[10] ENISA, "Methodologies for the identification of Critical Information Infrastructure assets and services," 2014.

[11] ENISA, "Smart Grids Task Force EG2 Deliverable - Proposal of a list of security measures for smart grids," 2013.

Cyclic voltammetry study on electrochromic copper(I) oxide thin films

Ratka Neshkovska[1], Mimoza Ristova[2] and Julijana Velevska[2]

Abstract – **Electrochromic copper(I) oxide thin films were deposited onto conductive glass substrates using two different methods: chemical bath and low vacuum deposition. The coloring/bleaching processes of the films were investigated by cyclic voltammetry.**

Keywords – **Copper(I) oxide, Thin films, Electrochromism, Cyclic voltammetry.**

I. INTRODUCTION

A. Electrochromism

Electrochromic materials are able to vary their optical properties (color, transparency, absorbance, etc.) in a reversible manner, upon electrically induced oxidation and reduction when they are subjected to a small electric field (1–5 V) [1]. They can be colored in a reduced state and bleached in an oxidized state, or vice versa.

The concept of electrochromism originated in the 1960s. In the early 1980s, the electrochromic materials were a subject of great interest and were extensively studied due to their possible use in displays and watches. Due to durability issues and manufacturing problems, another type of chromogenic material, liquid crystals, was chosen for displays. Furthermore, liquid crystals were able to switch from a transparent state to an opaque one faster, within fractions of a second. The interest in EC materials had a new impetus in the early 1990s, particularly in the aerospace and aviation sectors [1]. They are currently being used in the automotive industry as glare-free automotive rear view mirrors and sunroofs. The great interest of researchers during the last decades in the optical, electrical and mechanical properties of electrochromic materials, as reflected by the large number of studies developed worldwide, is due to the fact that the most promising electrochromic device use is "smart windows" - all-solid-state electrochromic devices used for modulating of incident solar radiation. They change their optical properties (transmittance, reflectance, absorbance, etc.) in their colored and bleached state, thus achieving energy efficiency and human comfort in buildings. The technology has recently been implemented in large-area glazing (windows and glass facades) in order to create buildings which combine energy efficiency with good indoor comfort [2].

A number of materials, both inorganic and organic, liquids, solids and polymers exhibit electrochromism, but the main materials with electrochromic properties are metal oxides of some transition elements, in particular WO_3, MoO_3, IrO_2, NiO, and V_2O_5.

Major techniques for analyzing electrochromic oxide films include techniques for studying their crystal structure and elemental composition, optical properties, electronic structure, ion intercalation and deintercalation. The processes of ion intercalation and deintercalation are studied by:

- Chronoamperometry
- Coulometric titration
- Cyclic voltammetry
- Impedance spectrometry
- Beam deflectometry
  - Mirage effect
  - Beam bending
- Nuclear magnetic resonance (NMR)
- Microbalance measurements [3].

B. Cyclic voltammetry

Cyclic voltammetry is a very important analytical method of characterization in the field of electrochemistry. Any process that includes electron transfer can be investigated with this characterization. Cyclic voltammetry is a potentiodynamic method of tracking the current flowing in the circuit, caused by the change in the potential of the working electrode [4, 5]. This method permits the investigation of the redox and transport properties of a given system in a solution. Most often for this purpose, three electrode systems consisting of a working, reference and counter electrode are used, in which the working electrode can be exposed to a single or multiple linear change in the potential, relative to the reference electrode. The large impedance between the reference and the working electrode forces the resulting current to flow through the counter electrode. The current flowing through the opposite electrode is measured. The potential is measured between the working electrode and the reference electrode, while the current is measured between the working electrode and the counter electrode.

The velocity variation of the potential between the electrodes depends on the type of the process and can have a value from millivolts per second, to a few volts per second. When the potential of the working electrode is positive, there is an oxidation of the material applied to it, i.e. the electrons move from the solution to the working electrode and give an anode current. Similar, when the potential is changed in the opposite direction, the potential of the working electrode becomes more negative than the reduction potential, and

[1]Ratka Neshkovska is with the Faculty of Technical Sciences at "St. Kliment Ohridski" University of Bitola, Makedonska Falanga 33, Bitola 7000, R. Macedonia, E-mail: ratka_neskovska@yahoo.com.

[2]Mimoza Ristova and Julijana Velevska are with the Faculty of Natural Sciences and Mathematics at "Ss. Kiril&Metodij" University of Skopje, Arhimedova 10, Skopje 1000, R. Macedonia.

reduction occurs (i.e. the electrons "flow" from the electrode). Thus, the cathode current is obtained. In order to transfer electrons, the molecules in the solution should be brought to the electrodes. In cyclic voltammetry, the solution is not mixed, so the mass flow occurs only because of the diffusion (due to the concentration gradient) around the surface of the electrode. The signal size is a function of the diffusion characteristic of the system. The intensity of the current is dependent on the size of the surface of the working electrode, the concentration of the electroactive components and on the speed of the potential change (scan rate). On the diagram which gives the dependence of the current density on the potential between the electrodes, a maximum (or peak) occurs, when a certain reaction on some of the electrode's surface begins to occur. The current in the peak has a different value for reversible and irreversible processes and, among other things, depends on the square of the velocity of the potential change, the molar concentration of the electroactive component in the solution and the surface of the electrode. The peak current in a reversible system at 298 K is given by the Randles-Sevcik equation:

$$I_p = \left(2.69 \cdot 10^5\right) n^{3/2} A D^{1/2} v^{1/2} C \qquad (1)$$

where n is the number of exchanged electrons during the redox process, A (cm^2) is the active surface of the working electrode, D (cm$^2 \cdot$s^{-1}) and C (mol$\cdot$cm^{-3}) are the coefficients of diffusion and the concentration of the electroactive components in the solution, respectively, and v (Vs^{-1}) is the scan rate (the velocity of the voltage change).

In electrochromic phenomena, cyclic voltammetry is used to investigate the processes of intercalation and deintercalation of ions into the electrochromic film. The voltage applied between the working electrode with electrochromic film and a counter electrode, is swept back and forth between two selected values. Often, the voltage is changed triangularly between the selected values, but it may also have a pulsating character between two specified voltage levels. With such a change in the voltage, the current flowing into and out of the film in conjunction with the ion intercalation/deintercalation is measured. The results of the cyclic voltammetry can be used to extract non-quantitative conclusions, such as: confirming that an electrochromic process takes place between the electrodes, to show the reversibility or irreversibility of the process, to determine the voltages that lead to the stable operation of the electrochromic system. Voltages outside the stability range lead to the release of gases, the covering of electrodes with metal, etc. Quantitative cyclic voltammetry is also possible, so one can measure the charge densities associated with intercalation/ deintercalation (coloring/bleaching processes). However, it is often enough to see that the electrochromic process is reversible, i.e. the input current in each cycle is equal.

The electrochromic behavior of the films is influenced by some structural parameters, such as the crystallinity and morphology of the films. Various interpretations exist for their influence. These parameters strongly depend on the method of film preparation and the conditions of preparation in each method. Therefore, the electrochromic properties of films prepared by different methods are studied.

The voltammograms (the current's dependence on the voltage) depend on the method of obtaining the film, the used electrolyte, the quantity of water in it, the voltage scan rate, the thickness of the film, the number of previous cycles of coloring/bleaching of the electrochromic film, its porosity, its age, etc. Cyclic voltammograms measured in various voltage scan rates can't be normalized to one universal curve. This also applies to the best electrochromic materials, including WO$_3$ [4, 5].

The coloring/bleaching of the copper(I) oxide thin film is related to the intercalation/deintercalation of ions and electrons in the film, according to the general reaction:

$$Cu_2O + xM^+ + xe^- \leftrightarrow M_xCu_2O,$$
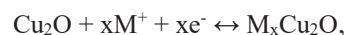
where M$^+$ is Na$^+$, Li$^+$, K$^+$ etc. cations and e$^-$ are electrons. The film is reversibly bleached with electrochemical oxidation, and is colored with reduction, in an electrolyte that contains light cations.

II. EXPERIMENTAL

It has been recently found that Cu$_2$O thin films exhibit cathode electrochromism [6-9].

Several methods for obtaining thin copper(I) oxide films have been developed, i.e. thermal oxidation, electrodeposition, sputtering, anodic and chemical oxidation, chemical oxidation, chemical deposition, sol gel technique, laser ablation, etc.

In this work electrochromic Cu$_2$O thin films were prepared by two different methods - chemical bath deposition and low vacuum deposition onto glass substrates precoated with fluorine-doped tin oxide (FTO) film. FTO film wete deposited as proposed by other autors [10].

A. Deposition of Cu$_2$O thin films by chemical deposition method

Because the optical properties of copper(I) oxide depend on the methods and parameters of deposition, the prepared FTO substrates were subjected to the electroless chemical deposition method of Cu$_2$O films, which has been proposed by other authors [12] and low vacuum deposition. For that purpose, two beakers with aqueous solution: a cold and a hot one were used. The cold solution contained 200 ml of colourless complex mix of 1 M CuSO$_4$ and 1 M Na$_2$S$_2$O$_3$. The hot solution contained 2 M NaOH at 60-80° C. The FTO coated glass substrates were stuck together with a tweezer-like holder, with their conductive layers facing outwards, in order to be alternatively immersed into the hot and the cold solution. This procedure lasted until the desired thickness was obtained. The thickness was estimated by the number of immersions, since it was previously established that for a given concentration the film thickness is proportional to the number of successive immersions. Hence, each ten immersions

yielded an additional 0.1 μm in the film thickness [12]. Finally, the thin films were rinsed under a running tap water and distilled water. Subject of our studies were the films about 150-200 nm thick (fifteen and twenty successive immersions) [7-9, 11].

B. Deposition of Cu₂O thin films by low vacuum deposition method

Cu₂O thin films were deposited onto the prepared FTO substrates using low vacuum evaporation method [11, 14].

The melting temperatures of Cu, CuO and Cu₂O are $1083^0 C$, $1326^0 C$ and $1235^0 C$ respectively. The temperature difference between elemental and oxides' melting points makes the evaporation technique attractive for depositing copper oxide thin films. Stable copper oxides can be deposited over a source temperature region of $1090 - 1350^0 C$. In general, the oxidation rate and the evaporation rate of the source material define the deposition rate of the product grown on the substrate surface. Copper oxide thin films were deposited on glass substrates pre coated with FTO by the low vacuum evaporation technique, using granulated ($0.3 \, mm$) copper with 99.8 % purity as a starting material. The deposition was performed in a Balzers Automatic BA 510 apparatus. The distance between the source (copper) and the substrates was about $50 cm$. The evaporation pressure was $0.67 \, Pa$. During evaporation the transparency of the substrates changed by 50 %. The deposition lasted several minutes, until all of the starting material from the tungsten boat was evaporated. If the vacuum in the chamber is very high, pure copper will be deposited on the glass substrates. In our case, the vacuum was not so high, so the evaporated copper reacted with the gasses in the chamber. As a result of that reaction cuprous oxide thin films were obtained on the substrates.

The deposited films were yellowish in appearance. The film thickness, measured by weight difference method was about $130 nm$ [11, 14].

Cyclic voltammograms of the deposited copper(I) oxide electrochromic films obtained by the two different methods were recorded with a computer-controlled three-electrode system, part of the Modular electrochemical instrument (AUTOLAB) apparatus. The working electrode of the cell was the electrochromic film of copper(I) oxide deposited on the FTO glass substrate, platinum wire was the counter electrode and Ag/AgCl was used as a reference (nonpolarized) electrode. In all cases, the voltage was changed from +1 V to -1 V and back to +1 V. 0.1 M solution of NaNO₃ in deionized water was used as an electrolyte The surface of the working electrode was about 1 cm².

III. RESULTS AND DISCUSSION

The prepared Cu₂O films revealed cathode electro-chromism. They were colored at a negative potential of –4V and were bleached at a positive potential of +4V in the constructed electrochromic test devices (ECTD) consisting of a glass supstrate/FTO/Cu₂O film as a working electrode, a 0.1 M NaOH aqueous solution as an electrolyte and an FTO/glass substrate as a counter electrode. The films are transparent to visible light in their oxidized state and almost black in their reduced state (Fig. 1) [7].

a)   b)   c)
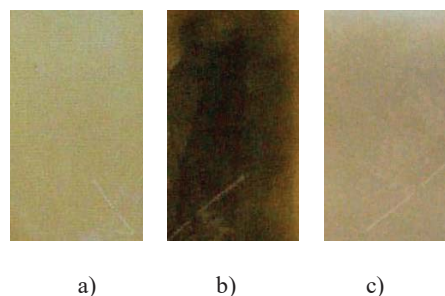
Fig. 1. Photos of electrochromic copper(I) oxide thin film: a) as prepared; b) in colored state; c) in bleached state.

The cyclic voltammogram of the chemically deposited film in 0.1 M aqueous solution of NaNO₃ is given on Fig. 2. The peak potential values of intercalation/deintercalation of ions are −500 mV, -180 mV and 600 mV.
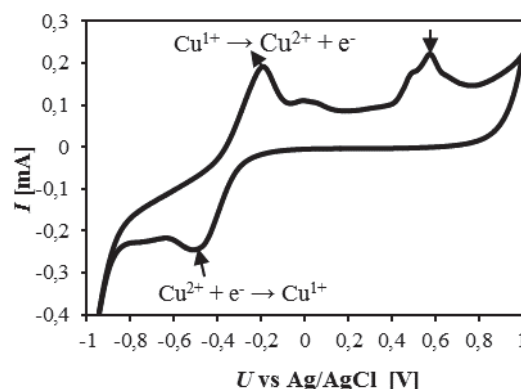
Fig. 2. Cyclic voltammogram of chemically deposited thin Cu₂O film in 0.1 M NaNO₃ at scanning rate 10 mV/s.

Fig. 3 shows five consecutive voltammograms of chemically deposited film in the same electrolyte at the same voltage scan rate [7].

Fig. 4 and Fig. 5 present one and five consecutive cyclic voltammograms of Cu₂O thin films deposited by low vacuum evaporation method, in 0.1 M NaNO₃ electrolyte. The cycling was carried out within a potential range from -1 to +1V and back to -1V vs. the counter electrode. The voltage scan rate was 25 mV/s, and the film working area was 1 cm².
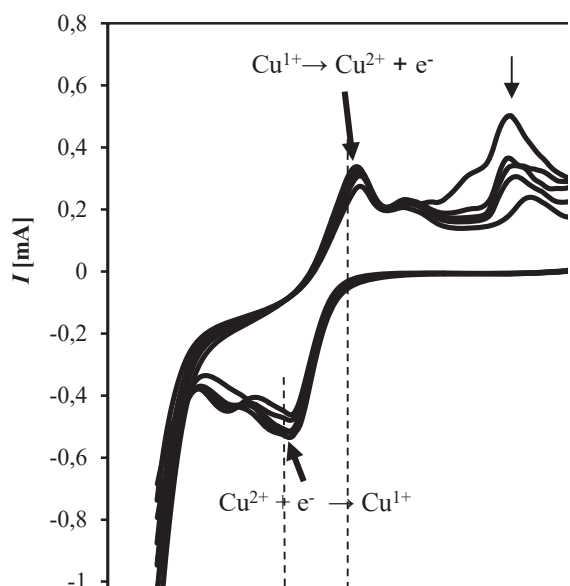
Fig. 3. Five consecutive cyclic voltammograms of chemically deposited thin Cu₂O film in 0.1 M NaNO₃ at scanning rate 10 mV/s.
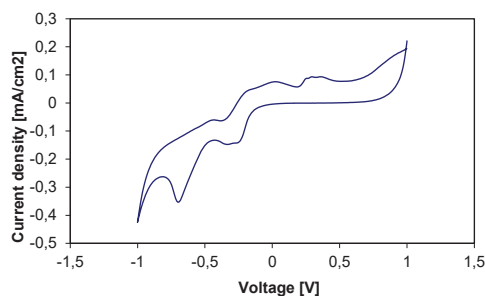
Fig. 4. Cyclic voltammogram of low vacuum deposited thin Cu₂O film in 0.1 M NaNO₃ at scanning rate 25 mV/s.

The above results confirm the electrochromic process between the electrodes in the electrochromic device in which the working electrode is made of copper(I) oxide film, deposited by the two methods. Also, they confirm the chemical reversibility of the processes of coloring and bleaching that occur in copper oxide electrochromic films deposited by the two methods. The cathode and anode peaks indicate that the coloring and the bleaching of the films are related to the reversable transition of Cu(II) in Cu(I), respectively, according to the eqations:

$$Cu^{2+} + e^- \rightarrow Cu^{1+},$$
$$Cu^{1+} \rightarrow Cu^{2+} + e^-. \qquad (2)$$
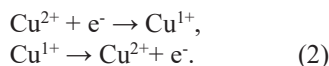
All voltammograms exhibit resolved cathode and anode peaks (corresponding to the reversible red-ox conversion of Cu₂O into CuO), the reversibility is fully established and could be provided by cycling within the voltage range between −800 mV and 500 mV.
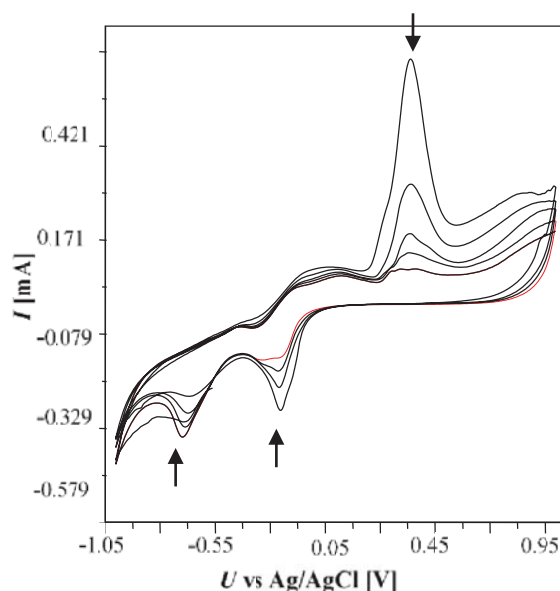
Fig. 5. Five cyclic voltammograms of low vacuum deposited thin Cu₂O film in 0.1 M NaNO₃ at scanning rate 25 mV/s.

IV. Conclusion

In this paper, the ion intercalation/deintercalation processes in thin electrochromic Cu₂O films, deposited by two different methods were studied with cyclic voltammetry. The results of the cyclic voltammetry were used to extract non-quantitative conclusions, such as: confirming that an electrochromic process takes place between the electrodes, to show the reversibility of the process, to determine the voltages that lead to the stable operation of the electrochromic system.

References

[1] M. Pittaluga, Electrochromic glazing and walls for reducing cooling building needs in Eco-efficient Materials for Mitigating Building Cooling Needs, 1st Edition, Woodhead Publishing, 2015, p. 473-497.
[2] C.A.C.Sequeira, D.M.F.Santos, Polymer Electrolytes, Fundamentals and Applications in Electronic and Optical Materials p. 3-6, Woodhead Publishing, 2010.
[3] C. G. Granqvist, Handbook of Inorganic Electrochromic Materials, Elsevier, Amsterdam, 1995.
[4] D. K. Gosser, Jr., *Cyclic Voltametry*, VCH, 1993.
[5] J-C. Brodovitch, *Chem*. 367-2, 39-46.
[6] Ozer, F. Tepehan, Sol. Energy Mater. Sol. Cells 30 (1993) 13.
[7] R. Neškovska, M. Ristova, J. Velevska, M. Ristov, Thin Solid Films 515 (2007) 4717-4721.
[8] M. Ristova, R. Neškovska, V. Mirčeski, Solar Energy Materials & Solar Cells 92 (2007) 1361-1365.
[9] M. M. Ristova, V. Mirceski, R. Neskovska, Journal of Solid State Electrochemistry, 2014, DOI 10.1007/s10008-014-2666-x.
[10] M. Ristov & co. Annuare Physique 47 p. 113 (1997).
[11] R. Neshkovska, M. Ristova, J. Velevska, ICEST 2016, 28-30 June 2016, Ohrid, R. Macedonia, pp 371-374.
[12] M. Ristov, Gj. Sinadinovski, I. Grozdanov, Thin Solid Films 123 (1985) 63.
[13] R. Neshkovska, M. Ristova, J. Velevska, ICEST 2016, 28-30 June 2016, Ohrid, R. Macedonia, pp 375-377.
[14] R. Neshkovska at all., Horizons, Series B, Volume 1 p. 25-32 (2014).

CONTROL SYSTEMS

PRACTICAL EXPERIMENTS AND ANALYSIS FOR DETECTION OF SMOKE IN THE BLUE AND INFRARED LIGHT SPECTRUM

Milen Kirov, Velimira Todorova

Abstract – At the present stage of development, there is a steady tendency to increase the fire hazard. To prevent fires, technical means are used to automatically detect, limit and eliminate the fire. Fire detectors are the main elements for fire detection. This report will describe the research on detecting smoke aerosols in the most commonly used type of Optical Smoke Detectors. The aim is to benchmark and propose a solution for improving the efficiency of fire detection by using a pair of optical couplers in the blue spectrum of smoke detector light emitted during the combustion process.

Keywords – **Optical smoke chamber, Optical smoke detector, Smoke aerosols**

I. INTRODUCTION

The study and analysis of smoke detectors and smoke detection techniques in the blue light spectrum makes it possible to carry out research and apply an approach that is new to its smoke detection method.

In case of a fire of the substances, solid and liquid particles are formed along with gaseous products, which form smoke. Smoke is a dispersing system in which the dispersed phase is small particles - resins, oxides, water, etc., and the dispersed medium is gaseous - air, gaseous products of the fire. Smoke is often called an aerosol product of fire, and smoke particles - smoke aerosols. The smelting particles, after their generation, are in the range of 0.1 μm to 1.0 μm. in the dispersion system they move with the flow of hot gases and air. As an accompanying process of fire phenomena, the formation of aerosols plays a dominant role in the creation of means of fire detection. Many studies have been done on the particle size of different types of fires as well as its different phases. The overall conclusion is that in all cases the amount of invisible particles is greater than that of the visible particles. The amount of smoke in a given zone is expressed by weight concentration, number of smoke particles in unit volume, for example [obsc./foot] and optical density [%]. Optical density expresses the ability of the smoke to alter the optical properties of the environment by reducing its transparency.

The advantages of the method of detecting smoky or diffuse blue light smoke aerosols is that the use of blue light offers considerable advantage in the absorption and reflection of smoke aerosols of dimensions smaller than the respective wavelength. Compared with previously existing solutions with infrared light (940nm wavelength), the blue light spectrum is almost twice the wavelength (470nm). Blue light detection enters almost the whole range of smoke aerosols in the process of combustion of substances of different origins.As mentioned above, some substances emit a lot of dense smoke. Other substances emit considerably less smoke, with some derivatives burning without visible particles. It is precisely for those substances that have no visible particles in the combustion that the optical pair pairing with blue light will be effective.

The idea of using blue light is made possible with advancing technology and the ability to produce semiconductor LEDs with a wavelength of 470nm and high emission, which is in the visible area of the blue light.

II. EXPLANATION

For comparative analysis and demonstration of the applicability of the method, an experimental laboratory device for measuring smoke aerosols under real conditions was constructed.

The installation consists of:
- Laboratory "tunnel" for test fires;
- Controller with input for analog signal measurement 3;
- Optical chambers where the examined optical pairs are mounted;
- Power supply;
- Oscilloscope for capturing and confirming data from photo sensors.

The laboratory plant for simulating fires is constructed of fire resistant plasterboard. It is proportional to the test fires room described in the European Standard EN-54-7.
At the top are placed the two test optical devices - the purpose of the survey, and on the bottom there is a small hearth where the materials that will burn and imitate a real fire will be put.

An optical smoke chamber is used for comparative analysis. The camera is designed and built on a 3D printer. The material used is black ABS. In the optical camera, infrared and blue light sources and photo receivers are mounted respectively. The geometric dimensions are so designed that the light emitted does not fall directly or indirectly into the photodetector. Fig. 1 and Fig.2 shows a model but the optical camera with which the tests were made.
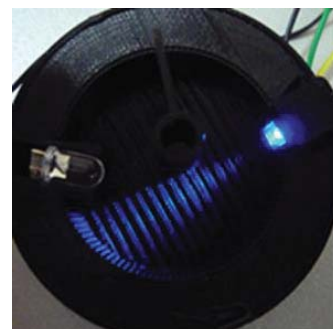
Fig. 1 Optical smoke chamber bottom

Fig. 2 Optical smoke chamber top

The experiment uses an LL-503IRT2E-2AC infrared LED with a selective characteristic.
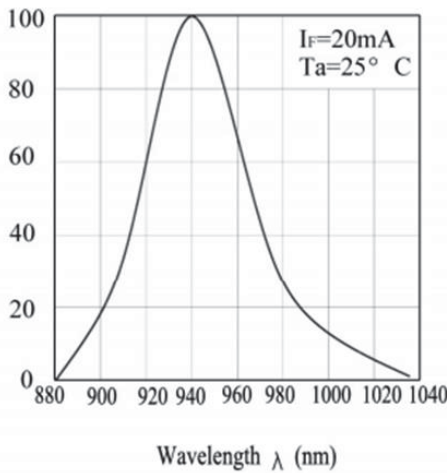
Fig.3 maximum radiation at wavelength 940nm

A blue LED for the tests is LUB50343 with a selective characteristic.
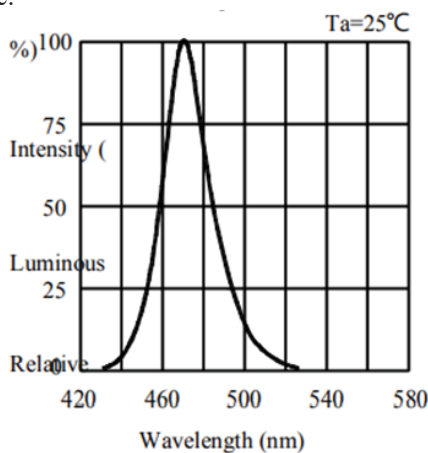
Fig.4 maximum radiation at wavelength 470nm

For both the source of the blue and infrared light using a photo detector SFH 203, with selective characteristic.
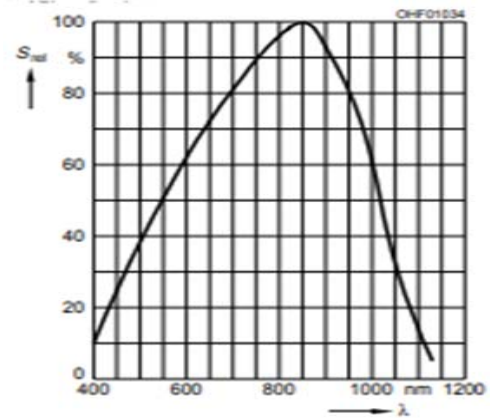
Fig.5 Wide spectrum including 470nm and 940nm

The terminals of the optical elements are connected to a circuit board that buffers input signals from a dedicated controller.
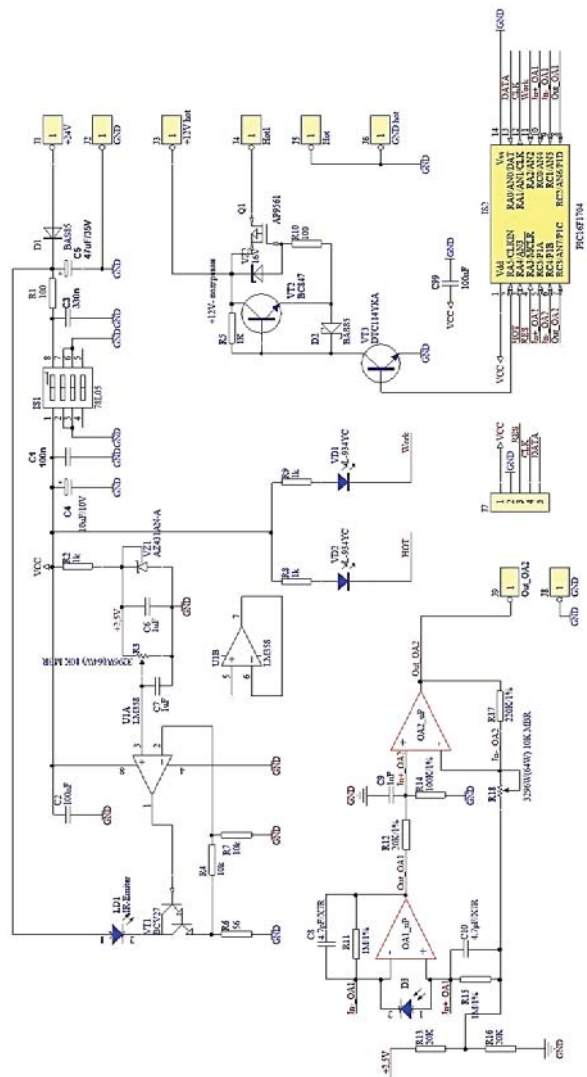
Fig.6 Schematic diagram of an optical meter

The controller produces pulses for the LEDs and reads with an analogue digital converter the values of the photo-current generated by the reflected light of the aerosol smoke particles. Data from the serial interface controller goes to a standard PC input where it can be visualized and processed in tabular form. The oscilloscope monitors the levels of the two optical elements to control the process and to visualize and confirm performance.

The idea for laboratory tests is to assess the difference in the level of registration of reflected aerosol smoke particles in both detection types at 430nm and 760nm wavelength, with different combustible agents and detection time.

For the laboratory tests the experimental set-up for three test fires according to EN54-7, TF1-burning wood was made. TF1 is characterized by the rapid burning of thin wooden sticks without smudging. The test is carried out in the following sequence: the chamber is ventilated, all laboratory instruments are run, the photodetector is measured and visualized by the oscilloscope photometer. This is the starting point of the process being measured.

Place wooden sticks on an electric heater. The heater switches on, causing smoldering and ignition of the tree. The photovoltaic current behavior of the receiver, which is a criterion for registering a fire situation, is monitored..

After increasing the smoke concentration, the oscillograms Fig. 7,8 and 9 show the increasing photoelectric current, which clearly demonstrates the process's efficiency.

The yellow graph shows the pulse output to the LED emitter, the blue graphic reflects the receiver's photoelectric current.

The sharp impulses of blue graphics at the beginning and end of the rise and fall of the front are parasitic and are not taken into account. They are due to transient processes in the scheme. The level of the useful signal is measured in the middle of the pulse.

The values of the photoelectric current are received by a computer that converts them into dB / m.
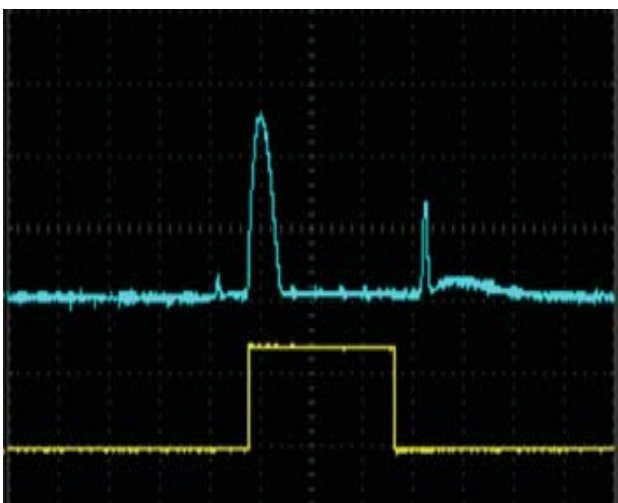
Fig.7 without smoke aerosols
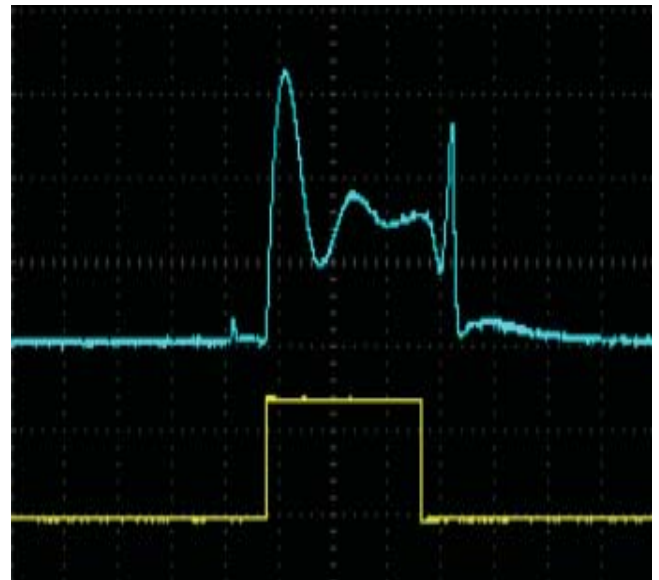
Fig.8 a low concentration of smoke aerosols

Fig.9 high concentration of smoke aerosols

For the quantitative judgment of the sensitivity of the optical detectors, the term "absorption index" is inserted as follows:

$$m = \frac{10}{d} \cdot \log_{10} \frac{P_0}{P} \qquad \text{[dB / m]}, \qquad (1)$$

where:

d [m] - optical beam length in the measuring chamber (maximum 1.1m);

P0 - the light output, measured without aerosols in the chamber;

P - the received light output, measured when there are aerosols in the chamber at the time of commissioning, of the fire detectors.

On this basis, photocurrent conversion is made in dB/m.

Test results obtained in tabular form are for the graphical picture.

From the table results, the graph is built.
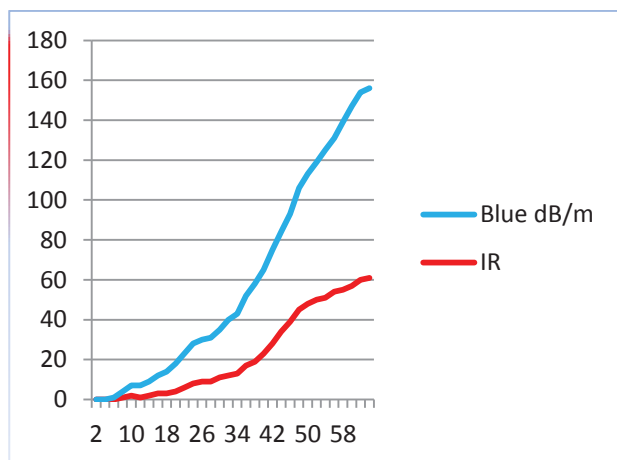
Fig.10 At the same levels of smoke aerosol concentration, the optical pair with blue light has nearly twice as high levels of the electrical signal

- The top curve of the chart is that of the blue light detector.
- The bottom curve of the chart is an infrared detector.
    After building the graph, you can see the advantages of the optical pair using blue light to detect smoke aerosols.

CONCLUSION

The conclusion is that blue light with its spectrum and wavelength 430nm reflects most of the smoke particles, respectively, resulting in greater sensitivity and signal processing capability to prevent false alarm signals, which is a drawback of so far used mass smoke detectors.

The selected optical couple: LUB50343 blue light transmitter and SFH 203 receiver have good general characteristics and can be used as base elements of an innovative smoke detector.

REFERENCES

[1] Benchev R., Fire alarm systems, Sofia 1989.
[2] EN54-7 Fire detection and fire alarm systems P.7
Point – type smoke detectors – Detectors using
scattered light, transmitted light or ionization

HUMAN-MACHINE INTERACTION OF ELECTRIC VEHICLE

Zlatko V. Sovreski[1,2], Petr Bouchner,[2] Stanislav Novotný[2], Misko Dzidrov[3] and Elizabeta Hristovska[4]

Abstract - **The paper provides overview of challenges of driver-car interaction in modern electric vehicles. Different aspects of human-machine interaction are reviewed. Analysis of method of driver-vehicle interaction through in-vehicle display and its influence on using eco driving modes is provided on the example of experimental studies in the field.**

Keywords - **HMI, Electric Vehicle, Hybrid Car, IVIS**

I. INTRODUCTION

One of the main problems of electric vehicles is limited battery capacity and longtime of battery recharge. Unlike in case with conventional cars energy management system is crucial for EV, where it is not a problem of ecology of driving, but it is also essential for general utilization of the EV. The range may vary on driving style and driving conditions in several tens of presents (for contemporary cars the difference is of 100 km). Most of the general rules of eco-driving can be applied to electrical vehicles too. Similarly to conventional cars, the eco-driving techniques may be applied in EV:

- To monitor the fuel use
- To drive smoothly and mind the traffic ahead
- Not to use the break and gas pedals aggressively
- Try to avoid speed oscillations when possible, i.e. on highways. Turn on cruise control when possible
- Not overuse air conditioning
- Use the air conditioner smartly – not under- or overuse
- Check the tire pressure
- Not to overload car with unnecessary load and not to increase aerodynamic drag like one from the roof rack, or open windows

This problem could also be solved with the help of application of automated vehicle systems.

[1,2] Zlatko V. Sovreski is with the University "St. Climent Ohridski" Bitola - Faculty of Technical Sciences[1] - Department of Traffic and Transportation Sciences, Czech Technical University in Prague - Faculty of Transportation Sciences - Czech republic, and E-mail: zlatkosovre@yahoo.com[2] and zlatko.sovreski@uklo.edu.mk,
[2] Petr Bouchner and Stanislav Novotny is with the Czech Technical University in Prague - Faculty of Transportation Sciences[1]- Czech republic, E-mail: xnovotny@fd.cvut.cz
[3] Misko Dzidrov is with the University "Goce Delcev" Stip, Faculty of Mechanical Engineering, E-mail: E-mail: misko.dzidrov@ugd.edu.mk - Republic of N. Macedonia
[4] Elizabeta Hristovska is with the University "St. Climent Ohridski" Bitola - Faculty of Technical Sciences[1] - Republic of N. Macedonia, E-mail: elizabeta.hristovska@uklo.edu.mk

II. SYSTEMS CURRENTLY IN USE

According to time scaling by the IEA (International Energy Agency), we are now living the third age of electrical vehicles. Quite a variety of HMI (Human-machine interaction) systems have been developed since the beginning of development of electrical vehicles by now. Let us assume that vehicle-to-driver (and vice versa) communication can be subdivided into display systems (the information displays and gauges), starter system, pedaling and charging system. Display systems are usually a wide variety of designs. Besides the intuitive design, interface needs to provide useful and relevant information sufficient to keep driver confident in car functioning.

II.1 Different aspects of EV driver-car interaction

There two main streams in approaches to design. The first one pursuing the goal to make the appearance of the systems maximally similar to those in ordinary fuel car with exploitation of mechanical gauges with arrows where and displaying EV systems operation related information instead. The familiar appearance makes the interaction intuitive and transition from fuel to electric cars more comfortable. Placement and manner of delivery of information is to be carried out on the assumption of the general knowledge and standards. According to ANSI (American National Standards Institute) [2] the indicators standardization is not electric vehicle specific, general rules for car displays apply here. Thus, instead of tank fuel level, there is an indicator of battery capacity, energy use and recuperation is usually associated with the revolutions counter, and navigation display usually contains map of refueling stations. The location of the gauges is usually as it were in a fuel car (Mitsubishi I Miev, Nissan Leaf).Sometimes the gauges location varies, like in Tata Vista EV, where the main meter is located at navigation panel place, and in Smart for two Electric Drive energy use and battery controls are above the navigation panel. Also design of indicators can be different (Ford Focus Electric). In those systems the information representation differs – battery charge can be expressed in percent, in km or in scale value. Another approach is quite forward in design, different in location of interface components and even uses the external and portable devices as in-vehicle display (BMW i3).Most of electric vehicles have also a display showing info related to eco-ing.

Fig. 1: Examples of EV charging stations: corporate station at Transportation Faculty, CTU in Prague (left),Tesla public charging station with solar energy source (upper right), home charge station by Easy charge (bottom right)

II.2 Propulsion system interaction

Due to regenerative breaking vehicle feedback from pressing the pedals in EV is different. Vehicle decelerates (with possibility to even stop at intersections) by depression of drive pedal only and starts on after idling by its pressing thus making it possible to drive with one pedal only. There is no clutch, to there are only two pedals which can be rather inconvenient for drivers who are used to mechanical gear.

II.3 Audio feedback

We know the importance of audio stimuli, car engine in particular, which is described in relation to automobile simulator engine in [3] and [4]– for driver and, in our context, also for pedestrians', cyclists 'and blind people safety. There exists quite a variety of possibilities for improvement of car to human interaction quality due to possibility of computer simulation of car engine sound. Through proper legislation, it is possible to regulate the sound volume inside and outside the car, solve the problem of sound proofing and make adaptation to electrical vehicles in general more gradual. Obligatory simulation of engine sound in electrical and hybrid vehicles has already been implemented in the USA [5] and in Europe and is referred to as Electrical Vehicle Warning Sound. According to legislation, the sound (Approaching Vehicle Sound for Pedestrians System – Nissan leaf) is activated while driving bellow the certain speed (50 km/h or 20 mph range in city) and is deactivated at higher speeds, when enough noise is generated by tire friction and aerodynamic drag. Being a necessary solution for pedestrians' safety, it does not intrude into the car cabin acoustic, which is an asset for driver and passengers comfort. Manufacturers of applied systems suggest pedestrian warning and in-cabin sound modules, the later can be switched on and off with possible volume adjustment (Electrical Vehicle Electronic Sound System by EVEESS^{TM}, Sound Box by S.M.R.E. Engineering etc.).

II.4 Charging

In electric vehicles, visual interface is concentrated around efficient energy consumption. The basic reason of it is the range anxiety phenomena, which is a fear of electric vehicle driver to run out of "fuel" and not be able to charge the

vehicle in time. The network of charging stations is not as dense as one of the fueling stations. Users, while switching from fuel cars, are to change their habits of route planning. The possibility of driving right after refueling is not available in case of EV, where recharging takes hours while driving electric vehicle requires certain route planning. That is why the infrastructure of charge stations is a very important factor for convenience of vehicle exploitation. Besides, the proper delivery and management of information regarding the charging stations is very important as the factor influences driver's route planning. Charging station solutions for EVs are proposed by special companies (General Electric, Eaton, Siemens etc.) as well as by electric cars manufacturers (Tesla Motors etc.) and can be available on the market as public charge stations, corporate stations for companies supporting the development of green technologies, and for private use at homes. The architecture of station can imitate the fuel station (public solution), parking spot (corporate and public solution) or look like a socket (more or less private solution). For reference please see Figure 1. EV battery can be charged manually at the station or remotely via existing remote control applications for cell phone having your car connected to the socket.

III. RELATED RESEARCH

III.1 Audio feedback

Experimental assessment of EV or hybrid HMI has been conducted by researchers from Sweden [6]. The selected group of participants had no, or little experience of driving electric or hybrid car. Goal of experiment what functional information should be in EV, how it can be presented and what approach of the interface (i.e. traditional, or innovative) should be preferred. Both stationary and while driving interface testing was performed. Probed were allowed to provide comments and opinion during the tests, relevant notes were made for analysis. Besides, number of errors and driver surveying were collected. The tested systems were: battery charge symbol and need of charge, message about limited performance of vehicle, eco meter. Two phases of experiment covered testing of two different concepts. During the first phase the classical (similar to fuel car) approach of HMI was presented for testing, while in phase 2 the one was developed and enhanced according to the reflections collected from participants in phase 1. Interface design changed from imitation of classical gauges, location of information changed too with moving of speedometer to the left and adding more information in the bottom and on the right. The interface looks more like battery device interface. During the experiment drivers related to information by habit and at first understood it intuitively as if they were in a fuel car. Among the compared systems were battery charge level, eco meter, distance to goal indicator, and readiness of EV electric consumption, warnings and some others. The information characteristic to EV (or hybrid) only was accepted better by drivers in the second phase. While in the first they were looking more to where the things should be. The most familiar systems in both cases were the battery level, distance to goal

(and empty) and speedometer, which is quite understandable as the first is familiar from mobile devices and the second – partially, from navigation system interfaces.

Don Norman in his book specifies for basic principles of development of sound of electric cars and hybrids: it shall convey an alerting function (vehicle is somewhere near), it shall serve for orientation regarding vehicle direction of movement, it shall not be annoying and it shall be designed in the terms of certain standards so that it is able to be interpreted as a car sound in relation to the term of skeuomorphic [7].

Cooperative research by a Laboratories Vibrations Acoustique, INSA-Lyon, France and Institute für Psychologies, Technische Universität Darmstadt, Germany was made to compare sound perception by sighted (100 participants) and blind people (53 participants) of diesel and electric car sound [8]. Nine variants of electric car sound were proposed to propends. The task set for participants was to determine from which direction the vehicle was coming in pedestrian crossing in dry and rainy weather (rain sound added) with traffic noise simulation. Better results in response time of vehicles detection were detected in dry weather, and for EV some respondents could identify the approaching car at unsafe (under 7.5 m) distance. There was not a big difference between responses of blind and sighted people.

Visual interface for electric vehicle Efficiency of interface in motivating driver for eco-driving can be assessed in laboratory with the help of vehicle simulator. The emissions, pedals use, acceleration/deceleration can be measure in simulated driving environment.

One of the proposed approaches described in [9], where a system providing static and dynamic feedback is developed. The system will provide voice notifications (warnings) when driver is not driving in eco-mode and driver will be given recommendations on driving after the measurement. In experiment five non-eco types of driving were predefined: rapid acceleration and deceleration, high engine revolutions rate, high speed with high amplitude on highways or freeways and idling while being parked. Visual interface of the proposed system indicated CO2 emissions. The experiment results showed decrease in CO2 emissions. There were three tests run, both static (3.43%) and dynamic (5.45%) feedback showed CO2 emissions reduction.
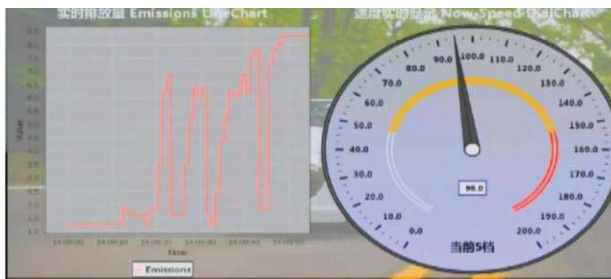
Fig.2: Proposed eco-driving mode notification [11].

Another research proposes interface with notifying driver about eco related information in trucks (Figure 3) [10].In this experiment drivers were testing a notification system and were to evaluate the kinds of notifications. The design of experiment was to identify which of the presented information was found useful and helpful for the drivers. There were two types of information on display – permanent (average fuel meter, speed guidance, acceleration/deceleration guidance, and coasting guidance at crests and changing (speed, feed forward advice and performance feedback indicating how well the driver performed in the last event (the stars and text under the speedometer). After the experiment drivers chose the most useful among the presented information – see Figure 4.

Other than visual types of notification have been tested by [11]. There were three systems proposed: Two hap tic accelerator pedal systems (hap tic forced and hap tic stiffness) and one multimodal visual–auditory. The system suggested information about fuel efficiency and position of acceleration pedal. Depending on the section speed limit, probed were advised to select acceleration pedal position angle: increase (15% depression), decrease (0% depression) or maintain their speed (7% depression) – see Figure 5 for reference. Three system notification types showed comparatively similar outcomes, where all systems have successfully improved Eco manner of driving, with the only difference for deceleration with hap tic force system, which was more efficient. Visual-vocal interface was considered more loading by the drivers' subjective opinions.
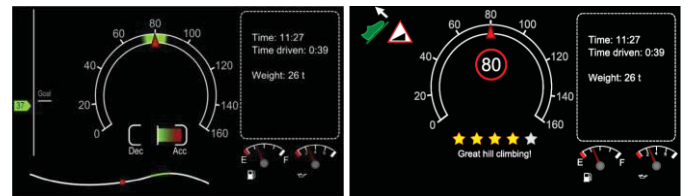
Fig. 3: Permanent and changing information about eco-driving [10].

Number of participants who selected the different eco-driving constituents.

Eco-driving constituent	Number of participants
Speed guidance (continuous)	12
Average fuel meter (continuous)	14
Coasting guidance at crests (continuous)	9
Acceleration/deceleration guidance (continuous)	9
Speed alert (intermittent)	8
Feedforward advice (intermittent)	15
Performance feedback (intermittent)	13

Fig. 4: Notification selection by propends. [10].
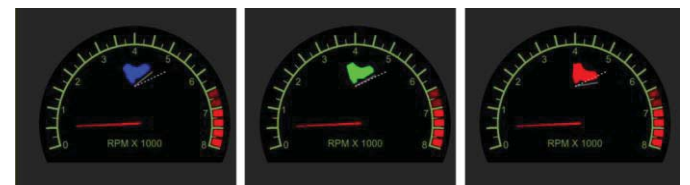
Fig. 5: Acceleration pedal advise in [7] - left is insufficient, right is excessive, middle is normal.

IV. CONCLUSION

Motivation of eco-driving is proved to be efficient method of overall driver awareness of necessity for driving smartly, save natural resources and decrease contamination of the

ecosphere. Real-time dynamic feedback appears to be more efficient method for motivation of eco-driving. Efficiency of notification and study of the target audience of such systems is very important. Average consumer on a vehicle market will be found to be used to classical outline of the in-vehicle system interface. For many drivers over 50 the visualization is found to be unclear or distracting due to vision issues with age. That is why it is important to perform sufficient research on visual design and representation of useful information: more types of notifications (like hap tic an acoustic) shall be considered for design by car manufacturers.

REFERENCES

[1]   Yoshizawa, S., et al., Development of HMI and Telematics Systems for a Reliable and Attractive Electric Vehicle, SAE Technical Paper, 2011-01-0554, 2011, doi:10.4271/2011-01-0554.

[2]   Electric Vehicle Standards Panel, Standardization Roadmap for electric vehicles, Version 2.0, American National Standardization Institute, 2013.

[3]   Hajný M., Module of Audio Subsystem of Car Simulator,(in Czech), Diploma thesis, FEE CTU, Prague, 2006

[4]   Ing. Petr Bouchner, Driving Simulators for HMI Research, PhD. Thesis, Institute of control and Telematics, CTU, Prague, 2007

[5]   U.S. Department of Transportation Proposes, New Minimum Sound Requirements for Hybrid and Electric Vehicles, Press Release, 2013 (source http://www.nhtsa.gov/ last accessed on May 29, 2015).

[6]   Helena Strömberg, et al., Driver Interfaces for Electric Vehicles, AUI 2011 Proceedings, 2011

[7]   *Don Norman,* The Design of Everyday Things: Revised and Expanded Edition. *Available from Basic Books, a member of the Peruses Books Group. U.K. MIT Press, 2013.*

[8]   Etienne Parizet, et al., Auditory warnings for electric vehicles: Detestability in normal-vision and visually-impaired listeners. Applied Acoustics, Volume 86, December 2014, Pages 50-58

[9]   X.Zhao, Y. Wua, J. Rong, Y. Zhang: Development of a driving simulator based eco-driving support systém.Transportation Research Part C: Emerging Technologies. Elsevier Ltd. Contents lists available at Science Direct.

[10]   Carina Fors, Katja Kircher, Christer Ahlström: Interface design of eco-driving support systems – Truck drivers 'preferences and behavioural compliance. Transportation Research Part C. September 2015. 2016 Elsevier Ltd.

[11]   Ing. Zlatko V. Sovreski, Ph.D., Doctoral dissertation., Study and application of hydrogen as an alternative propulsion in vehicles in urban mass transportation, CTU in Prague, 2013.

Mathematical Modeling and Simulation of Hybrid Electric Vehicle

Gordana Janevska [1], Mitko Kostov[1] and Goran Stojanovski[1]

Abstract – **The paper presents a Hybrid Electrical Vehicle modeling based on multi-physics approach. The integrated simulation model of a series-parallel HEV is developed including the electric drive system, PI controller, vehicle load model, and gear box. The system step response is simulated and analyzed. The simulation results confirm the validity of the model.**

Keywords – **Mathematical modeling, Simulation, Hybrid electric vehicles.**

I. INTRODUCTION

Hybrid Electrical Vehicles (HEVs) are vehicles with many electric components compared to conventional ones. In fact, the power train consists of electrical machines, power electronics and electric energy storage system connected to mechanical components and to an Internal Combustion Engine (ICE). The approach for a new vehicle design has to be multidisciplinary in order to take into account the dynamic interaction among all the components of the vehicle and the power train itself. In order to find the correct size of the components, the best energy control strategy and to minimize the vehicle energy consumption the vehicle designers need modeling and simulation since prototyping and testing are expensive and complex operations. Developing a simulation model with a sufficient level of accuracy for all the different components based on different physic domains is a challenge.

A HEV uses both an ICE and an electric motor/generator for propulsion. According to the architecture of hybrid propulsion, there are three basic layouts of HEVs:

- **series hybrid** in which the ICE and the electric motor are connected in series, and only the electric motor is providing mechanical power to the wheels;
- **parallel hybrid** in which the ICE and the electric motor are connected in parallel, and both the electric motor and the ICE can deliver mechanical power to the wheels;
- **Power-split hybrid** (**series–parallel HEV**) which combines both parallel and series hybrid architectures. This architecture is more efficient overall, but at the cost of more complicated control systems.

In this paper, the series-parallel hybrid form is studied. The paper is organized as follows: Section II details the components used in the HEV model. This includes the formulations for the electric drive system, vehicle load model, and gear box. Section III describes the development of the HEV model using Matlab/Simulink software package. Section IV provides simulation results, and Section V concludes the paper.

II. SERIES–PARALLEL HEV MODEL ARCHITECTURE

The series–parallel HEV (Fig. 1) incorporates the features of both series and parallel HEVs. Therefore, it can be operated as a series or parallel HEV. In comparison to a series HEV, the series–parallel HEV adds a mechanical link between the engine and the final drive, so the engine can drive the wheels directly. When compared to a parallel HEV, the series–parallel HEV adds a second electric motor that serves primarily as a generator.
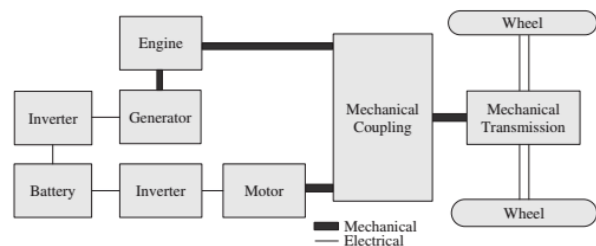
Fig. 1. The architecture of a series–parallel HEV

A. Electric Drive System

The high voltage *battery* is the main power source for the electric motor drive. For the purpose of this modeling, an ideal battery source is assumed, which is modeled by using a simple voltage source with zero source impedance.

The main function of *inverter* is to convert the DC voltage to an AC voltage. The inverter is a major part of the motor drive circuit and it allows controlling the current in the motor, hence the motor produces smooth torque output. In this paper, inverter is modeled as an ideal three-phase current source from the controller.

Main function of an electric drive is to convert electrical energy into mechanical energy. This is accomplished by the use of an electric machine or a motor. Here, as part of the system modeling, a *Permanent Magnet Synchronous Motor* (PMSM) is used.

Motor torque as a function of speed is given by

$$\tau = \tau_s - K \frac{1}{\omega_{NL}} \omega \ , \tag{1}$$

where τ_s is the stall torque, ω_{NL} is no load rotational speed, and K is a constant. Motor power as a function of rotational speed is

[1]Gordana Janevska, Mitko Kostov and Goran Stojanovski are with the Faculty of Technical Sciences at St. Kliment Ohridski University, Bitola, Makedonska Falanga 37, Bitola 7000, Macedonia, E-mail: gordana.janevska@uklo.edu.mk

$$P = \tau \cdot \omega \quad . \tag{2}$$

A PMSM used in the electric drive can be modeled using the steady state equations. PMSM motor phases is configured as a wye connection.

State equations for the motor phase current are:

$$\frac{di_a}{dt} = \frac{1}{L}(v_a - i_a R_a - v_{e\_a} - v_n) \quad , \tag{3}$$

$$\frac{di_b}{dt} = \frac{1}{L}(v_b - i_b R_b - v_{e\_b} - v_n) \quad , \tag{4}$$

$$\frac{di_c}{dt} = \frac{1}{L}(v_c - i_c R_c - v_{e\_c} - v_n) \quad , \tag{5}$$

where v_a, v_b and v_c are phase drive voltages, R_a, R_b and R_c are motor phase resistances, v_{e_a}, v_{e_b} and v_{e_c} are the phase back-emf voltages, and v_n is the neutral point voltage.

A commonly used technique to achieve the phase alignment and produce the maximal torque is to change the frame of reference from stationary A,B,C frame to rotational D,Q frame of reference. This frame transformation, known as Park's Transformation, converts an AC signal into a DC signal to simplify the analysis and controls.

Since d-axis produces no torque, the measured and controlled d-axis current should be 0. On the other hand, q-axis produces the maximum torque, so the q-axis current should be the actual commanded current.

The block diagram of the model is shown in Fig. 2.
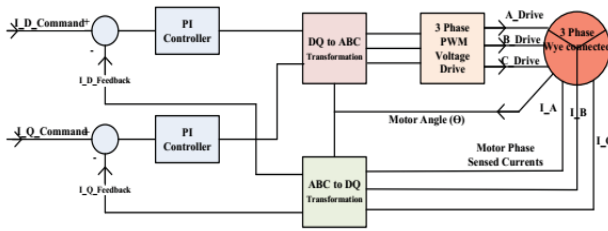
Fig. 2. Closed loop current control

The conversion from stationary frame to the d-q rotating frame leads to the dynamic state equations of the phase currents:

$$\frac{di_d}{dt} = \frac{1}{L_d}(v_d - i_d R_s + \omega_m L_q i_q) \quad , \tag{6}$$

$$\frac{di_q}{dt} = \frac{1}{L_q}(v_q - i_q R_s - \omega_m L_d i_d - \omega_m \lambda_m) \quad , \tag{7}$$

and the motor torque is given by

$$T_e = \frac{3}{2}\frac{P}{2}\left[\lambda_m i_q + \left(L_d - L_q\right) \cdot i_q i_d\right] , \tag{8}$$

where P is the number of poles in the machine.

B. Vehicle Load Model

Vehicle load can be a function of multiple variables, including aerodynamics. In this study, only the parameters that provide the largest effects on the vehicle in terms of load are considered.

The torque on wheels can be calculated as

$$T_{wheel} = r f_{wheel} = r M_T \frac{dv}{dt} =$$
$$= r M_T \frac{d(r\omega_{wheel})}{dt} = r^2 \frac{d(\omega_{wheel})}{dt} \tag{9}$$

where r is the radius of the wheels, f_{wheel} is the force on the wheels, i.e. accelerating mass of vehicle, M_T is the total mass, $M_T = M_{veh} + M_{tire}$, v is the velocity of the vehicle, and ω_{wheel} is the rotational speed of the wheel.

The load torque can be calculated as

$$T_L = F_T \cdot r \quad , \tag{10}$$

where F_T is the total force and it can be determined by

$$F_T = f_{wheel} + f_{rr} + f_d + f_{slope} + f_{bear} \tag{11}$$

The tire rolling resistance f_{rr} is given by $f_{rr} = \mu_{rr} M g$, where μ_{rr} is the rolling resistance coefficient. Viscous drag f_d can be calculated as $f_d = (1/2)\rho A C_d v^2$ where ρ is the air density, A is cross section area and C_d is the viscous drag coefficient. Force due to slope effect f_{slope} is given by $f_{slope} = M g \sin(\theta)$, where θ is the slope angle. Force due to bearing friction f_{bear} is given by $f_{bear} = (k_b/r)\omega$.

The rotational equation of vehicle tire is determined as

$$\frac{d\omega}{dt} = \frac{1}{J}(T_e - T_L) \quad , \tag{12}$$

where J is the moment of inertia, and T_e is the motor torque.

C. Gear Box

The problem that occurs in the case of hybrid powertrain configuration is to find the best gear ratios and arrangement that can cover more kilometers with minimal use of electricity from the battery. The problem of minimizing fuel consumption depends on several factors such as gear ratio, engine torque throughout the cycle.

Reducing the gear ratio allows the shaft to spin at a lower rate, and increase the torque, yet get all the power from the motor drive. Since the motor can spin at high rate, the gear ratio allows the motor to be at low torque. This mathematically can be written as

$$P = \tau_m \omega_m = \tau_w \omega_w = \tau_w \frac{\omega_m}{G} \quad , \tag{13}$$

where ω_w is the wheel speed, τ_w is the torque at the wheel, and G is the gear ratio.

Speed ratio can be defined as ratio between input engine speed and the output speed of the vehicle. By changing the connection of the planetary gears, the efficiency of the vehicle can be increased [3].

The selection of an appropriate gear ratio can be useful in efficient use of the power of the drive, and minimizing the energy consumption from the power source [4]. Optimizing the gear ratios can be useful for saving energy during travel, resulting in more mileage with one charge of the battery.

III. Integrated Simulation Model

The system simulation model is developed on the basis of the theoretical mathematical models shown earlier. The five main parts of the overall model are motor model, vehicle model, gear box model, controller model, and model of power balance and energy conservation.

The paper considers the following blocks: the controller, motor model, gear box and vehicle load model. Power-energy model is used to calculate and store the values of input and output power and energy during the drive cycle. The integrated simulation model of the system is shown in Fig. 3.
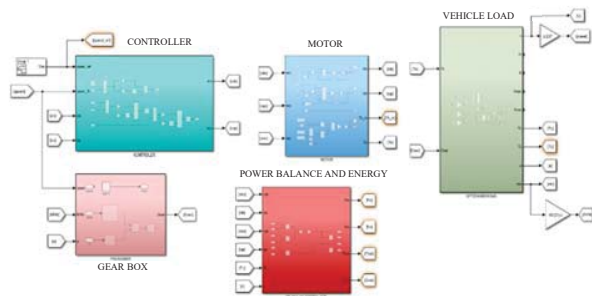
Fig. 3. Integrated simulation model

A. Motor Simulation Model

A PMSM motor simulation model, developed in Simulink on the basis of the Eqs. (6), (7) and (8), is presented in Fig. 4. At this model, v_{ds} and i_{qs} are inputs, and i_{ds} is the output.
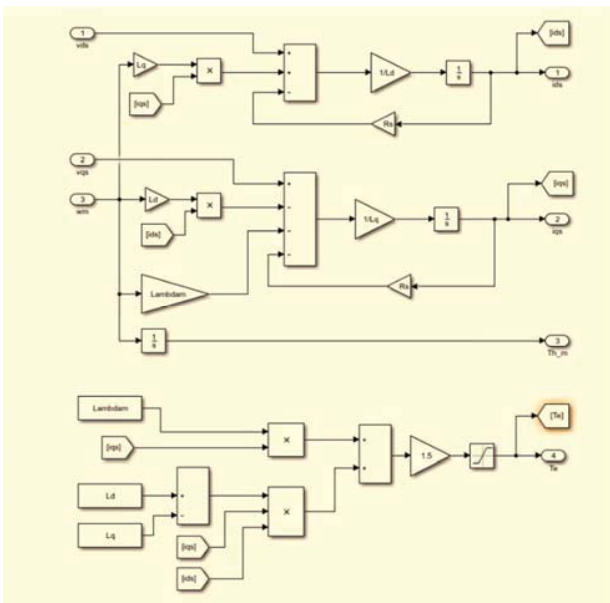
Fig. 4. Simulation model of PMSM motor

B. Model of the Controller

PI controller is used for vehicle speed control. The d-axis current is controlled to zero, so only the q-axis current produces torque. Kp and Ki gains are chosen so as to provide a critically damped step response, as well as a reasonable steady state error less than 2%. The simulation block diagram of the controller is shown in Fig. 5.
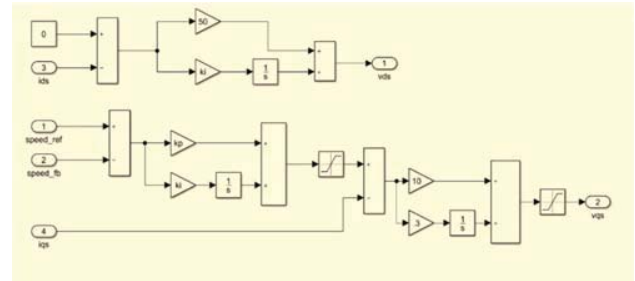
Fig. 5. Simulation model of PI Controller for speed and current

C. Load Modeling

The load model is developed based on the key parameters of the vehicle and the environmental conditions. The simulation model of vehicle load developed in Simulink (Fig. 6) is based on the mathematical model given in Section. II B.

Vehicle weight of 1500 kg is used on the basis of the passenger fleet averages. Although the viscous drag coefficient is a function of the vehicle aerodynamics, an average value of 0.5 is used. The tires and rotating parts weight is taken to be 50 kg, the tire loss coefficient is 0.01, $A_{Front} = 3.7$ m^2 and $r_{tire} = 0.3$ m.
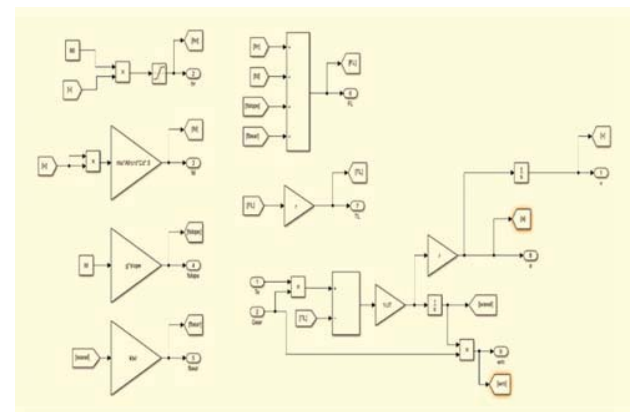
Fig. 6. Simulation model of vehicle load

D. Gear Modeling

The main objective of the simulation is to find a combination of gear ratios that will produce the most energy efficient drive cycle, and provide the highest possible mileage. For the basic model, nominal gear ratios are selected so that sufficient power can be generated and delivered to the wheels. In order to find the optimal gear ratio for the lowest energy consumption, a range for each gear is used. The range is selected so that the

vehicle can provide reasonable torque and power to move under the loads. The simulation is made using random integers within the range of each gear and determining the energy used during the drive cycle.

A state model for this gear transition used in the simulation is shown in Fig. 7. The used model ensures that the engine speed does not exceed the set value of 5000 RPM. The model
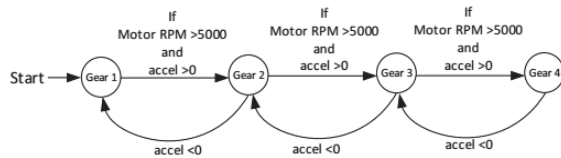
Fig. 7. Gear transition

also allows the both, up-shifting during acceleration, and down-shifting during deceleration.

IV. SIMULATION RESULTS

One of the best ways to test a control system is to input a step function. Understanding the system's step response is important to determine how well the system is designed and tuned.

In this paper the simulation is performed using step function for vehicle speed as an input, not only to determine the parameters of the system in the setting of PID loop, but also to optimize the gear ratios.

The system step response for two distinctive cases is considered, i.e. in a case of acceleration of the vehicle, as well as in a case of vehicle decelerates. For the both cases, the diagrams of speed and acceleration of vehicle, the motor current, the torque, as well as the input and output power are determined (Figs. 8 and 9, respectively).
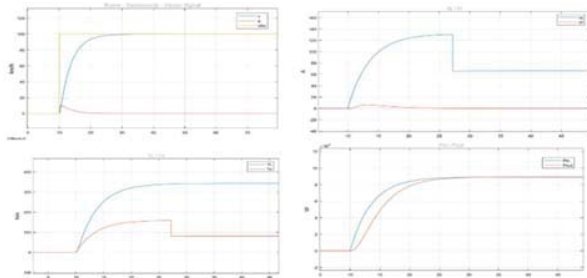
Fig. 8. System step response due to acceleration

Step response shows the settling time i.e. how much time it takes for the vehicle to achieve desired speed. Step response also shows the steady state error of the system control loop. According to the simulation results, the system settling time is 20 seconds, achieving steady state commanded input in this duration. As it was already noted, this is important to confirm that the PI gains have been correctly selected.

Initially, the motor torque is high due to low speed, and the input power drawn from the battery is high due to initial high load requirement, mainly due to the required acceleration. In the beginning the acceleration increases, since the vehicle starts from zero speed. As the speed reaches steady state, the acceleration decreases to zero, and input and output power become theoretically equal.
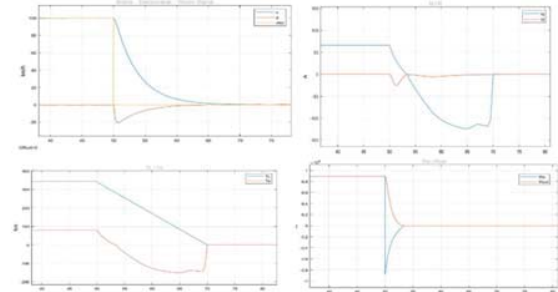
Fig. 9. System step response due to deceleration

When the set input speed is zero, the system goes into regenerative braking mode. During this time as the vehicle decelerates, and the speed goes down to zero, the mechanical system is putting power back into the battery. This can be clearly seen in the plot as the negative input power.

On the basis of the simulation results, it can be concluded that the system is performing well in a case of step input.

V. CONCLUSION

In this paper, the series-parallel HEV modeling based on multi-physics approach is presented. The system integrated simulation model, including the electric drive system, PI controller, vehicle load model, and gear box, is developed using Matlab/Simulink. Through a performed simulation the system step response is determined. The simulation results show that the system is performing well when step input is given, which confirms the model validity.

REFERENCES

[1] Ch. Mi, M.A. Masrur, D. W. Gao, *Hybrid Electric Vehicles: Principles and Applications with Practical Perspectives*, John Wiley & Sons, Ltd, 2011.

[2] S. Massey, *Modeling, simulation and control of hybrid electric vehicle drive while minimizing energy input requirements using optimized gear ratios*, Michigan Technlogical University, 2016.

[3] G.-S. Hwang, "Power Flows and Torque Analyses of an Independently Controllable Transmission with a Parallel Type", *Proceedings of the World Congress on Engineering, London, U.K.*, vol. 3, pp. 2020 - 2024, 2011.

[4] M. H. Prins, C. W. Vorster and M. J. Kamper, "Reluctance Synchronous and Field Intensified-PM Motors for Variable-Gear Electric Vehicle Drives", *IEEE Energy Conversion Congress and Exposition*, pp. 657-664, 2013.

[5] G. Janevska, *Modeling and Simulation of Mechatronic Systems*, Tempus project DRIMS, 2012.

C-STAR Simulator Teaching Experience

Zoya Hubenova, Konstantin Metodiev[1]

Abstract – **In the paper hereby, some issues related to training, evaluation, and selection of UAV operators are considered. In particular, the training process is carried out by means of simulator C-STAR, SimLat Ltd, and group of trainees. Each trainee is required to perform exercises at the pilot and the payload station and, eventually, take a test. In this way, the trainees' abilities are estimated and the eligible candidates are selected for further training.**

Experimental data are processed upon candidates' training completion and presented in the paper graphically. Conclusions are made as to what previous experience is preferable and others.

Keywords – **operator, UAV, simulator, training**

I. INTRODUCTION

Year after year, a growing number of companies set about developing and building unmanned aerial vehicles (UAV). For example, the number of unmanned airplanes in the US Air Force for the past 10 years has grown 136 times, i.e. from 50 units in 2000 to 6.8 thousand in 2014. In Europe, the belief that armed unmanned complexes may prove to be a good substitute for combat aviation is firm. Russia is trying to catch up with Europe, the United States, and Israel in creating an unmanned aviation, and in the next three years has increased the development of different types unmanned airborne systems.

The number of unmanned aerial systems for civilian use is increasing too. However, development of civil UAVs is delayed due to the potential users being interested in fully-built systems and purchasing related services. What is more, a legislature basis for UAV certification has yet to be developed which is the reason why it is remarkably easy to obstruct the UAV integration into existing air traffic management systems, [1], [2].

Ensuring safety is of paramount importance in preparation and performing flights by UAV. At certain technological level it is possible for the civil air traffic control to monitor the UAV flights. Monitoring might also be carried out by military air traffic services, [3]. One of the main factors for high level of emergency is the human factor. Unfortunately, the majority of "experts" running this type of equipment, do not have relevant qualification which is frequently encountered issue. Obviously, staff training process can have a significant impact on safety. This makes it necessary to develop identical training requirements and programs for UAV operators, as well as reliable training tools depending on the different categories of UAVs.

A basic tool for training is the simulator which is used to provide individual and crew training and improve trainees' expertise.

[1]Department of Aerospace Control Systems, Space Research and Technology Institute, Georgy Bonchev str., BAS campus, Sofia 1113, Bulgaria, e-mail: komet@space.bas.bg

II. MATERIALS AND METHODS

A. Basic description of simulator C-STAR

The training system C-STAR of Israeli firm SimLat, installed at the Department of Aerospace Control Systems, provides an educational tool for pilots and payload operators alike who work within a crew. The training system is a virtual environment which closely resembles a real flight routine. Initially, the simulator is meant to train crews but it is also possible for both pilot and payload operator to carry out flight simulation separately. In this case, it is the instructor who replaces the absent crew member. The working environment is user friendly. The interface is understandable and accessible to everybody who is somewhat skillful in computers.

At the instructor station, the flight scenario and mission control are carried on. During the flight session, various malfunctions might be simulated, such as on-board instruments failure or severe weather. The instructor sets path and behavior pattern of ground, water, and aerial objects, the UAV flight route, flight task type (targets discover and tracking, aerial scan, etc.) and simultaneously logs flight monitoring data. The software tool PANEL facilitates instructor in relation to further data processing by creating charts, graphs and evaluating the trainee.
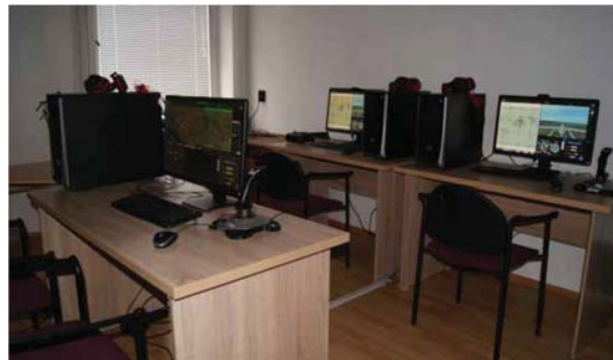
Fig. 1. C-STAR system common view

B. Training process stages

The Space Research and Technology Institute possesses a base version of the training system C-STAR, Fig. 1. It is thoroughly discussed in paper [4]. Crew members were taught in a "Laboratory for Selection, Training, and Control of UAV Operators" by Institute employees who had attended training course arranged by SimLat and been given certificates for instructor proficiency. All flight scenarios used for training process were developed by the researchers at the Institute.

In 2018, employees from Ministry of Interior Affairs and Ministry of Defence attended course at the C-STAR simulator. The course lasted five days and included theoretical study and practical exercises. Trainees were required to take a final exam. All trainees were between 20 and 30 years old. Among them, few had professional pilot experience.

Initially, the trainees were shown basic flight skills at the pilot station. This included take-off, following an initially set flight route, approach for landing with short/long lag, landing. Then all trainees were required to perform exactly the same flight procedures so as to gain some skills. In order to hinder the trainee in addition, the autopilot stabilization ability was occasionally switched off so that Euler angles limitations were no longer available.

The training continued at the payload operator station. It included discovery and tracking of a mobile ground target. The trainee was required to complete this task as quickly as possible. Meanwhile, the airplane was performing automated flight, i.e. no pilot was initially present. According to our experience, this particular flight task seemed more difficult and it took the trainees more efforts to succeed because the target kept on altering its position. The airplane manoeuvring only worsened the situation for it happened to shift the camera FOV.
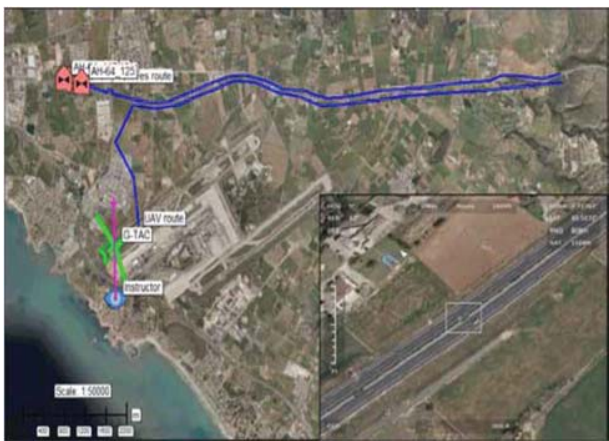
Fig. 2. Flight task routes

In Fig. 2, the aforementioned task scenario for is depicted. The airplane route is visible, so is the route of the ground mobile target (tank). In the lower right corner, a screenshot taken by the sensor video channel is placed. It depicts the target followed by the onboard sensor after successful lock. The sensor keeps on following the target automatically no matter how intensively the airplane performs a manoeuvre.

Eventually, the trainees were told to form crews and undertook team work at both stations. It should be noted that the stations share same video channel which (in our humble opinion) makes successful task completion by the crew less likely. The SimLat staff however declined our proposal of introducing an alternative video channel.

During the training process, various malfunctions were simulated in addition such as autopilot or cooling system failure, communication link loss, etc. The trainees were also put to a test of bad weather conditions, i.e. turbulence, wind shear.

III. RESULTS

Upon taking the exam, trainees are given a report produced by the PANEL tool. Exemplary reports are shown in Fig. 3-5. The results are average for excellent, acceptable, and unsatisfactory trainee performance. Each chart is divided into two main sections: a flight area map (left half) and a plot sector. The plot contains four curves and multiple grey bars which provide information about trainee's performance.

Fig. 3. Excellent result achieved by a professional pilot
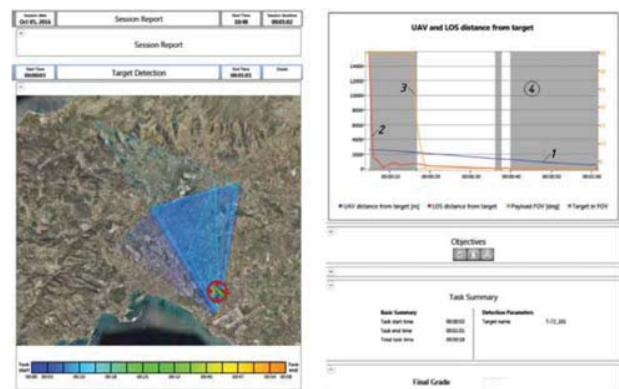
Fig. 4. Acceptable result achieved after initial training
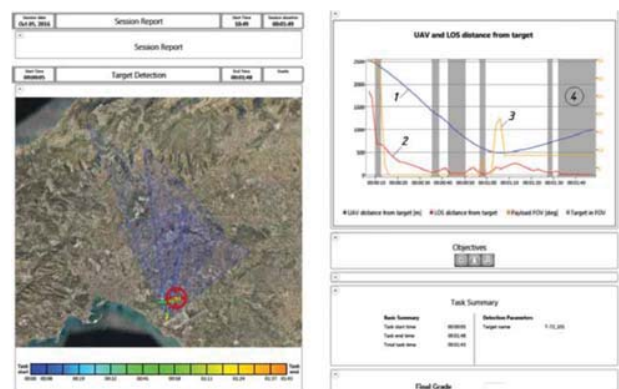
Fig. 5. Unsatisfactory performance

The charts show the payload operator's effectiveness during flight task performance. The curves stand for following:

- Curve 1 – distance between the UAV and the target, m;
- Curve 2 – distance between line of sight and the target, m;
- Curve 3 – onboard sensor field of view, degrees;
- Gray areas 4: target lies within the sensor FOV.

We therefore deduce that successful operators are either skillful in driving vehicles (airplanes in particular) or play computer games on a regular basis.

With the benefit of hindsight, we should have carried out additional research by means of a wider range of trainees in terms of initial experience. The flight sessions should have been extended. It is also advisable to take into account the trainee's opinion about personal performance, i.e. a self-evaluation report.

IV. FURTHER STUDY

Yet another way of getting a clear, deep, and perhaps sudden understanding of the human-machine interaction in this particular study case is making use of an eye-tracker. Having been installed, such a device presumably provides information about where the trainee's eyes are focused during the flight. This information can be used to gain deep insights into trainees' attention and also design new pilot / operator interface or improve potentially erroneous one.

Fig. 6. Stationary eye tracker and heat map

An exemplary stationary eye tracker is shown in Fig. 6. The heat map (red and green spots) is data visualization technique showing location, order, and time spent looking at different screen locations, [5]. For the flight simulator case, these locations may vary depending upon flight conditions. For instance, the trainee would not be expected to make use of the video channel if visibility is low.

V. CONCLUSION

The simulator training program is an integral part of the operators' complex training process, which focuses on the practical training. Upon completion of the course, candidates are supposed to have developed habits for quick situational assessment, making choice of control strategies, making optimal solutions in rapidly changing environment. The formation of these qualities and technical culture implies individual approach towards each trainee.

The professional training of UAV operators includes conducting selection of candidates as well as monitoring the acquired habits and knowledge based on simulators and training systems. For applicants who are assessed it should be determined whether they are skillful enough to perform specific work and be keen on performing better on duty.

Program implementation for candidates training (both professionals and amateurs) is focused on training and certification of UAV operators applicable to different spheres, i.e. environmental monitoring, mapping, infrastructure remote inspection, advertising, professional orientation of adolescents and young people. The flight simulator stage is an essential part of the training program which is suitable mainly for beginners. In this regard, experiments conducted in order to study activity of operator working with the C-STAR simulator include:

- Investigating operator's activity during UAV control and performing different types of flight tasks;
- Evaluation of the algorithm for correct data logging by the SimLat's PANEL tool;
- Making comparison between training data obtained by trainees with and without prior experience in UAV control;
- Picking out applicants that meet initially set requirements.

The UAV operator's job is mainly related to cognitive processes in terms of summarizing and analyzing a significant amount of information. Meanwhile, it is necessary for the operator to acquire skills quickly so as to perform complex procedures and actions, make decisions within a limited time interval, and communicate and interact within the crew simultaneously, [6].

Information processing is a basic approach in cognitive psychology. In this case, the human cognitive system is considered a system having devices for inputting, storing, outputting the information considering the bandwidth ability (by analogy to the computer). The methods of cognitive psychology are designed for qualitative and meaningful analysis of such psychological processes, such as motivation, decision making, goal setting, information processing, etc.

At present, UAV control is mainly carried out in semi-automatic mode. Most operations such as following flight path, aerial photography, area scan, ecological monitoring, search and rescue, inspection of power lines, wind turbines, railroads, anti-poaching, etc., are performed in automatic mode, which is the reason why the operator is able to monitor the flight activity having been located at a significant distance from the UAV. Nevertheless, in the final phases of flight, such as take-off and landing, operators are often forced to switch to manual mode. More than 70% of the military UAV losses occur due to subjective errors [7].

Significant difficulties are experienced by the operator during take-off and landing phases. This might be explained by the intuitive operator's response who is not present on board and does not receive visual information. The operator is unable to hear the engines running, sense a load factor during steep decent or climb, use side vision for orientation, etc. Instead, the operator solely observes the situation through the onboard camera. For these reasons, the operator is forced to interpret the flight instruments readings. The task of the operator, who is initially included in the control loop, boils down to detecting and recognizing the ground targets and also tracking the target by means of the onboard sensor. Naturally, the UAV capability and effectiveness are determined to a great extent by operator's ability to process visual information and implement accurate and highly coordinated actions upon target detecting and tracking.

ACKNOWLEDGEMENTS

REFERENCES

[1] Getzov, P., Z. Hubenova, G. Sotirov, "Professional Selection and Training of Operators of Unmanned Aerial Complexes," Third International Scientific Conference Science, Education, Innovation, Shumen, Bulgaria, 2014, Vol. 1, p. 50-63, in Bulgarian

[2] Concept of Air Surveillance and Intelligence with Unmanned Aviation Systems by the Armed Forces of the Republic of Bulgaria, Ministry of Defense, 2012, in Bulgarian

[3] Williams, K.W., "A Summary of Unmanned Aerial Aircraft Accident/Incident Data: Human Factors Implications," U.S. Department of Transportation, Federal Aviation Administration, 2004

[4] Getzov, P., "Bulgarian Space Research in the Context of Bulgarian Membership in European Space Agency," Twelfth Scientific Conference with International Participation, Space, Ecology, Safety, 2-4th of November, 2016
http://www.space.bas.bg/BG/magasin/SES/PROCEEDINGS%20SES%202016.pdf

[5] https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/interpret/working-with-heat-maps-and-gaze-plots/

[6] Langevin, S., B. Joseph , S. Straussberger , B. Guiost , G. Boy, "Human-centered design methodology: an example of application with UAVs mission"
http://conferences.telecom-bretagne.eu, 2008

[7] Aviation Accident History
http://www.planecrashinfo.com/cause.htm

Real Overview on Implemented p-FMEA Methodology in a Real Industrial Entity after a Year from the last Follow Up

Ivo Kuzmanov[1], Roberto Pasic[2], Zore Angelevski[3], Ilios Vilos[4]

Abstract – **This paper presents only a part from a real overview which presents an implemented p-FMEA into industrial entity after a year of its real implementation and usage on a daily base. The entity is a key player into the production of hot stoves in the country and one of the key players into the Balkans. The reason of this research is because of the extensive application research and also an implementation effort made, over the year which gave a real productive and profit benefit to the industrial entity. So, the data presented into previous papers presents the real benefit from the implementation into the years 2017-2018, and this paper is an overview which presents what happened with the company and the implementation process after a year since the last follow up, when the expert influence was done. The main aim of the paper is to present the real benefits of the p-FMEA as a method, but also to present what really happens with companies when the collaboration with an expert is done, or when the company doesn't understand the real benefit from the same one as a key method for quality improvements.**

Keywords – ***p-FMEA, production system, Quality Assurance, Quality Control, overview.***

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 North Macedonia. The same one works into the metal cutting industry, or to be more precise the same one is one of the largest producer of hot stoves and fireplaces for home use in R. North Macedonia, but on the same hand one of the key players in the Balkan market for years. One of the key elements which presents in best light the industrial entity, is the fact that this one is one of the key players in the market of fireplace and hot stove production in the Balkan area since the 90ties, and has a constant production more than 60 years with a small stagnation of production processes during the period of 90ties. At this stage the company has made a real investment into the machinery over the past 10 years and became one of the key players into the market, but on the other hand has seen the bad side from the concurrency which produces stoves which work on pallets. Still one of the key

points that represent this company as a real market player is that the same one has a tradition over 60 years, and has a constant production on new products (fancy ones) for the market. On the other hand the industrial entity has unique capability – to produce as much as the market demands because of its machinery investment and also because of the market role and cooperation with raw material producers. One of the key positive points at this stage is the number of produced pieces and the number of employees for this kind of producer which makes the company one of the key player into the Balkans, maybe even more – because there is a situation in which some companies buy products from this company and then sell them to foreign markets. When we are talking about the benefits or the positive sides from the company we must say that they use several CNC machines in several key production stages which bring the company real production benefits and a competitive advantage, but also a productivity which is quite bigger than the competitors.

But, so far we just present the positive side of the company. Now the market is changing from minute to minute, so the production of hot stoves and fireplaces is fearing a worldwide change – at first the demand of governments and regulative to influence on reduction of pollution, but also to compete with producers which produce stoves that use pallets as a fuel (trend at the market at the moment). So the company is facing real market turbulence and all of the efforts are up to the line to solve or to find a new product or a new market which will save the company profits. This is the main reason why this research was done and why the same one is at this stage.

So, in this point the main important thing is that the real implementation of the FMEA method was made and the real benefits into the year 2018 were with the following benefits: less waste materials in production stages, financial benefits, mind change, real management commitment, significant reducement of non conformities, implemented problem solving techniques on a week level, less production expenses etc. As well as the FMEA was implemented on a daily and month stage also significant reduce of waste materials were detected, significant quality improvements were detected, significant reducements of expenses were detected, but also bigger profits and ideas from the internal workers were spotted.

But the market had some significant changes during the year 2017 – 2019 and now the company is facing some problems such as: global market is changing from a day to day, there are some demands from the authority about the pollution, the mindset of the customers is changed and everybody is looking for inverters (as electricity heating) or stoves that work with pallets (as a less pollution and easier way of heating). So, the company and its management is in a

[1]Ivo Kuzmanov is with the Faculty of Technical Science at the University St. Kliment Ohridski Bitola - UKLO, Bitola 7000, R. Noth Macedonia, e-mail: ivo.kuzmanov@tfb.uklo.edu.mk

[2]Roberto Pasic is with the Faculty of Technical Science at the University Kliment Ohridski Bitola - UKLO, Bitola 7000, R. North Macedonia

[3]Zore Angelevski is with the Faculty of Technical Science at the University Kliment Ohridski Bitola - UKLO, Bitola 7000, R. North Macedonia

[4]Ilios Vilos is with the Faculty of Technical Science at the University Kliment Ohridski Bitola - UKLO, Bitola 7000, R. North Macedonia

situation where they are facing a change management solutions and the FMEA method is not on the top management strategic key points.

That is way the papers is done, as a result of an extensive research done into the same entity, just to see what really happened with the implemented FMEA method from one hand, but also to see what happened with the company when they face change management. So, the starting hypothesis from this point of view is that there will be some bigger RPNs on the same criteria's and that the level of quality of the final products will be not on such high level as into the year 2018 regarding that the company is facing some problems.

At the end of this part it's more than important to say that there were several papers published previously, but also the team is working on several similar project in different stages, from implementation, to follow up, to the stage where the same one has no benefits due to several issues.

And another key point at this stage is that from the beginning phase of the implementation there was a multidisciplinary team which was working on this project (conducted from different persons – university professor, managers, different shift managers, workers from different work departments, workers from the warehouse and even an customer), then the first follow up into the year 2018 was done only based on a daily work from internal team members from the company when the same one was still a part from the strategic plans of the company, and now at the end we have a situation where there is a second follow up or an overview when the team should be working with this method but we will see the results in the following paper.

II. SHORT OVERVIEW OF THE FMEA METHODOLOGY – WHAT IS IT AND WHY THE SAME ONE SHOULD BE USED

The presented method used at the initial research and used after the same one on a daily base (or in some cases used monthly) was the FMEA methodology. We should consider at this stage it is a thing that must be presented, so the readers could get a real picture about the method which was implemented and was a part of the daily activities in a large period of time (since 2016 till the end of the year 2018). On the other hand the same one was used as a p-FMEA or so called process oriented Failure Method Effective Analysis. So that really means that the method was used to specific process. At this stage it is more than important to present the same one and its real meaning so that we could see what are the real benefits of the same one, what was achieved in the past period into the company and at the end because of some reasons (market and customer ones) what is the situation in the moment. But at first in the following part of the paper the basic information of the FMEA are presented at first.

FMEA as a quality control and quality improvement oriented method is a worldwide known and recognized by companies as a method which will improve the quality, will reduce the problems, will deal with spotted 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 production problems, non-conformities and errors, but also as a method that creates potential solutions for the spotted production problems (related with raw materials, production, machinery, people, documents flow etc.). So the full method name is worldwide known as Failure Mode Effect Analyses. There are some modifications depending on the stage the same one is used as: p-FMEA or product / process FMEA, d-FMEA or design FMEA. But the general idea of the same one regarding in which stage is used is for:

- Detection of potential design or production problems which has a significant influence to the system, to the quality, to the work effectiveness and in total to the overall system productivity,
- Evaluating the potential and spotted problems and effects of each spotted and even detected problem /error or non conformity and their real or potential 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.

So in this stage we could conclude that the people that deal with quality and productivity will consider the implementation of a huge amount of methods such as: QFD, FMEA, FMECA, OEE, SPSS, SPC, ABC, KANBAN, KAIZEN, JUST in TIME and several other methods. But, 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 (although several other are as well) and that the same one is the most commonly used one for continuous improvements based on a constant usage of the same method. The improvements could be spotted in all of the production stages from the raw material department, try pre-production stages, try production stages on each machine, till the final product and post selling processes, seen as improvements spotted by the customers as well. It's a situation where the entity (the team which is working on the same one, but also every other employee) could spot all of the potential non-conformities, could evaluate the same ones, could divide the non-conformities to a priority or no priority ones at that stage for the system, but also could provide a process of several alternative actions which could reduce the influence of the same ones to the system. The provided alternative actions are also a thing that will be evaluated during the future FMEA processes and evaluations so the team could see if the same one brought a real effect to the process.

The implementation and the working approach of the FMEA method is based on a team work, process of evaluation of the system (but a real one – regarding how bad is it in that stage), and after the same ones (as activities) process of creation of real tabular views which actually are a multiplied numbers from three relevant factors. At this stage this tabular views are maybe one of the key elements why companies use this method, because the same one presents a real overview of the problems related to the production stages.

So, the key elements (factors) 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:

$$RPN = S \text{ (severity)} \times O \text{ (occurrence)} \times D \text{ (detection)}$$

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. At this stage, at first a worker with a real understanding and a real experience is a must have as a person to the FMEA team so that a real benefit is aimed from the method.

The multiplication of the three key factors could give the team a highest RPN number up to the number of 1000 (which is a situation that nobody wants). So, the final thing that is worth to mention is the solving approach, which is also one of the key things why companies choose this method. 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 COMPANY AND THE PRODUCTION STAGES INTO THE SAME ONE

At this stage because it is a company that has more than 3 year experience (at first non formal – only as a pilot project, but at the end as a real method used into the same one) it is more than important to present in short term the same one, it capabilities and also to present only a small part from the production stages. So, the same one is a company that has a market share and experience in the hot stove and fireplace production more than 60 years, and has been to different stages. From state one till 100 % private one. But during the period especially in the period from the late 80ties till the middle of the 90ties it has a significant reducement of the

production and had a period when even the same one has been under a key. Then since the same one changed the property from state to a private one, and since the market demands were to get a quality hot stove that will last, it has a process of transformation and after a lot of investments especially into the CNC machines and automated processes, has became one of the key players on the market, as well as into the Balkan's. Now there is a situation in which more than 150 employees are a part of the production processes, with more than 20 different products on the sale line.

But what is more than relevant in this key stage (year 2019), and because of the regulations and customer changed demands, the company has a "bad period". This is a situation mainly because of the upraising demand for pallet stoves and inverter technology as a new way of heating. Even in some parts a gas heaters are a main heating during the winter period. So in that situation the company is facing a situation in which the demand is reducing, so all of the activities are on a different level. Even there is a consideration to reduce the number of employees. So that situation is mainly interesting, because at the stage 2017-2018 the company was using FMEA and was enjoying the benefits from the same one. On the other hand now we have a situation in which the company's strategic goals are completely different. So, that was also one of the main points why this paper is written.

And before we present what really happened in the past when the FMEA was used, and what is happening now, we should have a real picture about the sub processes into the production stages of the company. So 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 here into the overview as into the follow up and the first implementation of the p-FMEA into the production stage the same production characteristics were also taken under consideration. So everything is the same just in a mater to get a real picture what is happening. The following characteristics were taken into considerations:

- Methodology of work
- Documents used for planning the work (work orders, customer demands)
- Machines – same ones as before

- Raw materials and other materials used while production (same raw material producers)
- Human factors (employees) – with some changes (some of them from before are retired, some are gone – or find another job, but also we have some new employees)
- Measurement instruments (some of them aren't calibrated at the moment and at this stage is a problem that was seen immediately)
- Work conditions (in different shifts) – there are some modifications as well regarding the new employees
- Customer demands (some are the same, some are new, and also there are customers which are trying to get a lot of discount because of the previous mentioned things)

So, having in mind that all the same characteristics were taken under considerations, but changes were spotted immediately, even ones which will have a bad influence to the production stages, at first we get an impression that the FMEA will have worse RPNs then before. In this stage especially human factor and measurement instruments were the first thing that we spotted and that could or should be a part of the new FMEA. But this time, the time necessary and given for creation of FMEA matrix was quite shorter especially because the management was considering getting more and more products at same time (shift) so they could get a lower price of the product. That was one thing more that was a problem.

Also during the last FMEA there was a company consideration to switch a part of the production processes so that they could produce pallet stoves, but over a year, nothing happened. And at the end we had a situation where it will be a must do for a shorter period of time. That is also a problem.

At the end the first impression is that we had a company where in a year period of time, a lot of problems accrues, and maybe the FMEA method could provide solutions if it is used again.

IV. Presenting the Results from the Overview

This is the main part from the overview and the presented paper. In this part of the paper we should once more present that this paper presents a small segment from an extensive research done in several parts since the year 2016 till 2019. So in the first period of time we had a situation where the subject was working on a daily base but without FMEA method, then a part when the subject started to use the method, then a part where the subject used FMEA on a daily base, and now finally when the same one is not used. So, previous published papers represents the real situation in each stage, but this paper takes into consideration the last tabular views (in a moment of active use of the p-FMEA on a daily base) and the real situation at the moment (February 2019) when the subject due to various reasons is not using the same one.

So, we could get a real picture about the situation and the real benefit from the FMEA method as a quality control method, but also a strategic one, we have done also a FMEA process in the present time so we could compare what happened.

Also on other hand so that we get a real comparison, regarding the subject, we used the same process as before (in which in the past FMEA was used). We took under consideration the process – Transferring done pieces to warehouse, as a sub process which is quite important for the production and even for some processes is the final process. So we could compare things we have shown two tabular views. The first one is also presented into previous published papers but is a starting point from which we could get a real picture what really happens when the process is under p-FMEA. Also this tabular view num. 1, shows us that even then some side effect happened.

TABLE I
PRESENTING ONE PROCESS UNDER FMEA

FOLLOW UP POINT IN YEAR 2018

PROCESS	POTENTIAL FAILURE	NUS EFFECT	RPN
Transferring done pieces to warehouse	Damaged piece	Replacing time sequences which are long, but compared to previously far more faster	10
	Long time required for transferring	Production delay and free work force with nothing to do at the moment	30
	Conditions which are not appropriate for the product into the warehouse	Damaged piece which has passed all of the production stages	6

Seeing this tabular view, we could say at first that although FMEA is used, some mistakes are spotted, but if compared with previous we could say that the benefits were seen in reduce which was more than 50%.

On the other hand, after the overview we made another table which is presented as following in which we could see that there are a lot of mistakes (more even than the past) in a moment when the company decided not to use the FMEA anymore. So, we present the table as well and if seen and

compared with previous we could conclude that the real implementation of the p-FMEA on a daily use brought a lot of benefits to the company in the past. Now we have a new situation and if we see the tabular view num. 2 we could conclude that a lot more problems we have in present time. Also if we compare that situation with the situation in which the company is in the moment, we could say that the company is not facing only market problems, but also internal problems that could be seen as loss of raw materials, loss of final pieces-produced ones, loss of quality (non conformities) and finally loss of money (loss of profit). Now, before we could present even more, first let's see the tabular view num. 2.

TABLE II
PRESENTING FAILURES AND REASONS WITH APPROPRIATE RPNs

PRESENT TIME – YEAR 2019

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	80
Long time required for transferring	Production delay and free work force with nothing to do at the moment	Transport equipment which is old, OR THERE ISNT ANY Workers who are only standing and not doing anything	120
Conditions which are not appropriate for the product into the warehouse	Damaged piece which has passed all of the production stages	Old building which was renovated in the past, but not as they should be renovated	100
Workers who do not what to do even when they are at working places	Damaged pieces in production processes	Not enough training for the workers, or no team leader appointed	40
New employees – almost every week there are new employees	Damaged pieces in production processes	Not enough training for the workers, or no team leader appointed	60

If we compare both tabular views presented, we could immediately conclude that there are a lot more problems than before in present time. On other hand because FMEA matrix was really done but in short time, just to compare the things and to present a real picture (how was in the past – how is now), and if we just compare the RPNs we could see the problems. Here we must say that maybe some of the problems which occurred are not presented, because the lack of time to do a real matrix again – It was a rush up, regarding the management decision that they should produce regarding to spend time for such activities. But, if we just see the RPN numbers, we could conclude that things happen more often, with a larger influence to the system and to the processes. On other hand there is a situation where nobody is taking care if these kinds of things could be prevented. And some of them really can be prevented.

Also another relevant information from before is or to be more precise are the previous activities which were just an idea how things could get better. This data are also presented in past papers. Having in mind that some activities were selected as a real must do for the entity, and a lot of time was invested into that process, we could conclude that there are a lot real problems in this subject at this stage. But first to get a clear view, let's see the previous things which were selected as a must do (in the year 2018 for the following period). They were done in a better time, a time when the company used the FMEA on a daily base. So here they are:

- Training for the workers especially for the process of transferring
- Special two week training activities for the new employees in each case
- Quality check done by workers on direct machines as a pre-process, actually before they start to use raw material (piece by piece)
- Generating workers which will be responsible for the transfer of 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 (forklifts etc.)
- Replacing the older transport equipments – the ones that they have at the moment (not automated processes)
- Follow up after doing the same ones

So if we see just the things presented previous and see what is happening at the moment there are some things which should be done, so the company could have a significant improvement into processes such as:

- Special training activities for new employees (each worker depending the job position to have different training)
- Creating a model for cooperation with workers - which will keep the workers into the company
- Showing what is happening into processes (gathering data which will be really analyzed by the managers and some actions will be taken)

- Buying new (or used, but in good condition) transport equipment (especially forklifts – electrical, on fuel and hand forklifts)
- Quality check in each stage (worker will be the first check point before and after work activities)
- Improving for workers – in term to get a higher and better paid work positions – as a general motivation to keep them into the company
- Creating a short term strategy with to do activities, deadlines and responsible persons
- Creating a strategy which will transfer the production into some new and more attractive for the market (stoves which will work on pallets – especially because the company has the equipment and the potential for such a big step)
- Repairing a part of the warehouse and using the same one for the done pieces (creating small modern warehouse)
- Creating a work flow without a warehouse
- Generating PDCA cycle in each production stage

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 significant financial cost at first, but seeing the final result it will be a long term benefit, and the same one will return as a profit on long term base. Maybe the first things to do are: to get a new or used but in good condition equipment (especially forklifts), to find a way to keep the workers and to motivate them, to create a work flow without a warehouse and to make quality check in each stage so the non conformity is less and less in stages.

This will be a something that hopefully will be followed into the future, and some future possible publications could be prepared as well. It's quite interesting to see what really will happen.

V. Conclusion

This paper is only a small segment from a real overview into industrial entity which had a real benefit from the implementation of the p-FMEA as a method and now due to several reasons has a situation where not only the same one has not significant benefits, but also is in a situation where the same one should consider its future. So, because the paper is only a small part from what was really seen, researched and really done into the company, maybe in the future some other publications will come from this paper and the overview done. But the main point of the paper is to present that the implementation of FMEA as a method in a full production stage and a real market share with customer demand is more than real, and also with real benefits. At this stage we could say that the entity had benefits is several different ways: from production savings in a matter of raw material, till effective production, till maximization of profits made as a benefits, till enrollment of worker ideas in every production stage, etc. But only one moment (changes into the regulations, market

changes and also customer needs changes) could influence on the implemented p-FMEA and also on the key strategy for the future. So that situation could turn over the "management eyes" to a complete new way of seeing things, and change the company way of doing things from a company where "future methods" are everyday activity to a situation where the company is considering only about the company's future and cares only about profits (if any).

So, this is one good example (maybe a bad one in practice) of how a real implemented FMEA method with benefits could be set as a non relevant method at the moment, regarding market changes. But on the other hand there is also the moment where the company should buy new equipment and where the company should change the main products if they want to have a market share, so the FMEA method from implemented will come to a stage non relevant, and again (hopefully) in future maybe will come to a situation where it will have a key role in industrial processes. That is maybe a thing that should and could be followed up, and from which several new papers could be published in future.

At this stage, from this example, we could conclude that the FMEA method and its implementation and benefits are a thing that is complementary with the way of doing things for the company, but also with the market share at the time. So, when the company has a real market share and incomes, methods could be implemented, but when there is a situation in which the company has market share problems, the future of the company and the profits are the priority things – so that is everything that is talked about in future. At the very end of the paper, a future research into the subject could get us to new similar publications.

References

[1] Ivo Kuzmanov, pFMEA methodology – overview activities done into real industrial entity, February 2019

[2] Ivo Kuzmanov, Silvana Angelevska, Roberto Pasic, Ilios Vilos, Presenting follow up on Implemented pFMEA Methodology into Industrial Entity as a Quality Control Methodology used on a daily base, ICEST Conference, june 2018, Sozopol, R. Bulgaria

[3] Ivo Kuzmanov, Step by Step to a Successful Implementation of p-Fmea, Lap Lambert Academic Publishing, internationally published Book, 2018

[4] Ivo Kuzmanov, pFMEA methodology – follow up activities done into real industrial entity, February 2018

[5] 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.

[6] Ivo Kuzmanov, pFMEA methodology – follow up activities done into real industrial entity, June 2017

[7] 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.

[8] Ivo Kuzmanov, research conducted into real entity, 2016 -2017

[9] Ivo Kuzmanov, FMEA methodology – internal documents for application into real entities, 2016

Approaches and principles of building functional-sustainable ergatic systems

Zoya Hubenova[1], Filip Iliev[2], Antonio Andonov[3]

Abstract – **The subject of this article is to carry out analysis of the risky information systems and the problem of their functional sustainability. The research is related to the modeling, forecasting, and control of highly responsible critical processes in large artificial systems. Such control is related to the concepts of information, organization, functioning, and purpose. An approach towards building a control strategy in risky ergatic systems based on the principle of functional homeostasis is shown.**

Keywords – **Ergatic system, Homeostasis, Information, Operator**

I. INTRODUCTION

At the present stage of the development of society in today's techno-sphere, large ergatic systems acquire global distribution and dominating role. They are characterized by a wide range of self-organizing capabilities, behavioral freedom, a great deal of internal information and memory. Tendencies towards increasing scale and complexity of the technical systems grow. For example, development and creation of such systems as large-scale transport, energy, aviation, and others are associated with consuming huge amount of resources having been exploited for the lifetime of a whole generation and even further. Essential properties of such complex systems are conflict liability and low predictability of consequences, which is often unrecognizable and yields risky conditions and crises. These can evolve into a possible and constantly developing conflict occasionally related to catastrophic consequences.

The ergatic systems (ES) are a set of a large number of hierarchically dependent complex subsystems comprising teams of people and technical means of a certain degree of organization and autonomy. Particularly, the ES are brought together in accordance with the current hierarchy of objectives through the means of organization, while, commonly, the ES are combined through energy tangible assets and information relationships in order to provide purposeful functioning of the system as a whole [1, 2]. The means of organization include *control nodes* the decision-making process takes place in and *actuators* turning information about decisions taken into actions which in turn are aimed at achieving the control objectives. Generally speaking, in these systems, both control nodes and actuators can be complex man-machine complexes. For the purpose of quantitative analysis of large systems, it is necessary to formalize and evaluate all components included in the definition of large systems, i.e. degree of organization, hierarchy of objectives, information, autonomy, relationsships between man and machine within the control nodes [3, 4]. The ergatic system, in accordance with the objectives of functioning, must provide for the fulfillment of matter-energy processes. However, its integral functioning, development, and existence are determined by the processes of transmission, processing and transformation of information.

The process of informatisation is accompanied by a widespread dissemination of information-seekers, advisers, designers and other systems that have already entered into different areas of human activity. The constantly increasing need of processing the growing amount of information automatically, the development of computing techniques, and the activating role of the man as an element of complex ergatic systems, make it necessary to investigate further so as to increase their effectiveness. The purpose of this article is to propose a conceptual-evaluation apparatus that forms the principles of building functional-sustainable ergatic systems.

II. RISKY INFORMATION SYSTEMS AND A PROBLEM FOR FUNCTIONAL SUSTAINABILITY

As a result of extensive society informatization, the risky control functions are placed under the surveillance of automated systems. This process gives rise to the problem of providing the functional sustainability of complex ergatic systems using hybrid human-machine technologies for information processing. [5]. It is of immediate relevance, therefore, that the problem of automated information systems providing proper functioning of the control systems should be characterized by: presence of risk in solving functional tasks, territorial and information distribution, concentrating of restricted access information, use of biological and electronic technologies for processing information, semantic accessibility for information impact, time constraints of the control cycle, and others. These properties determine the complexity of technological processes for information processing and the potential danger of violating their functional sustainability.

Therefore, automated information systems can be considered to be a component of risky control systems. This in turn lets the systems be individualized as a class of risky information systems demanding strict requirements to be met

[1]Zoya Hubenova is the Space Research and Technology Institute–BAS, Acad. G. Bonch-ev Str., bl. 1, 1113 Sofia, Bulgaria, E-mail: zhubenova@space.bas.bg

[2]Filip Iliev is with the Faculty of Telecommunications and Electrical Equipment in Transport at Higher School of Transport of Sofia, 158 Geo Milev Str., Bulgaria, E-mail: fgi@mail.bg

[3]Antonio Andonov is with the College of Telecommunications and Post, 1 acad. Stefan Mladenov str., 1700 Sofia, Bulgaria, E-mail: andonov@vtu.bg.

in terms of the functional sustainability. These requirements are due to dangerous consequences that may arise after functional failures. Again, the risky information systems are allowed to be defined as a class of ergatic systems implementing information processes in risky control systems. The risk itself boils down to potential danger of impairing the systems' functional sustainability for full or partial system failure yields significant economic, political, military, ecological, moral and other losses.

Functional sustainability can be defined as a property of the risky information systems, expressing their ability to implement certain information functions (processes of information processing) under conditions of external and internal destabilizing factors. The criticality in this case is characterized by the high level of information risks depending on the amount of damage the information system destabilization might lead to [6].

III. ASPECTS OF THE SYSTEM-INFORMATION ANALYSIS OF THE FUNCTIONAL SUSTAINABILITY OF RISKY INFORMATION SYSTEMS

Providing for functional sustainability of the information systems used in risky applications is a complex problem requiring working out systematically a solution of interrelated tasks related to development of theoretical positions, methods for automatic modeling, and analysis of complex systems. All this makes it possible to take the systems apart, to build reliable models of the information architecture and the information processing, to define requirements for the functional sustainability to meet, and to evaluate its implementation. Methods for system-information analysis of the functional sustainability include the following three steps:

1. An initial step appears to be study of information structure of the system under consideration. As a result of this step, main levels of the control hierarchy and related information streams are distinguished.

2. The next task of the analysis is essentially a matter of determining functioning purposes and building a relevant hierarchy. The quality of functioning depends on the fulfillment of certain set of criteria. The assigned objectives are actually images of parameters and criteria in the system state space, in which the system functioning is being monitored and the control commands are being applied.

3. Main goal of the analysis is development of a summarized criterion that make it possible to assess the system functional sustainability.

At the presented material, the degree of uncertainty (entropy) of the system's functioning in terms of defined purposes is proposed to be used as a generalized criterion. In this way it is possible to report all system states related to purposes, and also to determine the probabilities of occurrence of different situations [7, 8]. In this sense, formation of the control structure ensuring functional sustainability will be reduced to distribution of information resources in risky information subsystems between the possible situations within the big system, i.e. between a finite number of possible

situations, determined by significant changes in its properties or changes in the environment. In doing so, the search for a solution boils down to how optimally one can allocate the information resource between the baseline situations determining the main modes of functioning and the risky situations. The latter, in case of improper control and adverse conditions, may evolve so as to yield loss of functional sustainability, i.e. termination of the system existence as a whole.

Risky information systems are self-organized ergatic systems, characterized by the presence of technological areas with automatic, automated, and intellectual control. The last circumstance considerably complicates the analysis in the problem area, for the properties of functional sustainability are ambiguous in relation to information systems with different levels of complexity of the structural organization.

One of the most promising approaches used during building management strategy in risky ergatic systems is based on the principle of functional homeostasis, according to which: "Any system which is under the impact of other systems has the property of preserving, within certain limits, a set of sustainable functional behaviors in solving its own private or common tasks." [9, 10]. Researches in this area show that homeostasis is an integral concept describing the structural organization of systems. This organization is characterized by a high degree of internal freedom which permit the system's resources to be used rationally so as to achieve purposes in terms of the external uncertainty. This complex structural organization of homeostatic systems covers the following aspects:

- purposes of control;
- a process of purposeful functioning;
- the system as a loop of a homeostatic control.
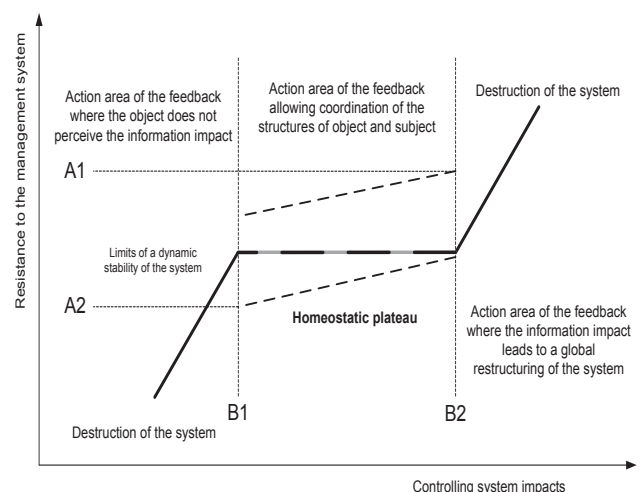
Fig.1. Homeostatic plateau

The essence and purpose of homeostasis is to provide constancy of the internal parameters of each system, the preservation of its integrity, and ability to survive in a dynamically changing environment. Then, functional sustainability of the system class under consideration can be defined as a dynamical equilibrium within the tolerances of

the homeostatic plateau (Fig. 1). The inconsistency between the subject and the object of the control effect getting the system out of the borders of this area, yields functional instability and information destruction of the system. This in turn produces an inability to adapt or modify the target function of the system as a result of global structural reorganizations.

Homeostatics is formed on the verge of various sciences and disciplines such as informatics, system analysis, biology, medicine, psychology, philosophy, artificial intelligence, economics, ecology and etc. The essence of homeostatics is studying the mechanisms of hierarchical control of complex systems providing constant vital functions, parameters, and modes of the development. It should emphasized that homeostasis is interpreted differently in cybernetics and homeostatics. In traditional cybernetics, in terms of purposes, homeostasis is seen as some stable control of the object state. It is provided by the fact that any difference between the controlled target state and the control goal is made up for by a negative feedback. In homeostatics, the homeostasis is seen as a dynamical constancy of the parameters, functions, and modes of development of the controlled object in the course of the purposeful or the unintended existence. This dynamical constancy is maintained at the expense of managing the internal contradictions existing or incorporated in the object. Maintaining homeostasis is performed in the hierarchical control structure consisting of three control loops, as the purposes in two of them is contradictory which is why they are targeted by the third control loop. A homeostatic system is a system that consists of a controllable and controlling part, as the latter appears to be a homeostasis. The homeostasis is a structure of control of material objects containing straight, inverse, and cross links. These provide for, in the course of their work, homeostasis maintenance, i.e. the dynamical constancy of vital functions and system parameters.

Essential properties of large ergatic systems are conflictuality and low predictability of consequences caused by different contradictions during the control processes in the individual subsystems. These are commonly not taken into account yielding critical situations that can evolve as a possible and constantly evolving point of a conflict, occasionally related to fatal consequences. In this respect, with the creation of new technologies in modern conditions and the improvement of responsible technological processes, reliability issues in technics, its safety, and questions about discipline, order, and organization acquire paramount importance [11, 12]. There is a sustainable trend in increasing number of accidents, incidents, and catastrophes worldwide. This is due to a number of reasons: while the technique evolves, the danger obviously grows faster than the human ability to resist; the accident toll increases; not only do people tend to get used to the danger but also to break the rules. In most cases, dangerous situations occur in case of close interaction between the technique and the operators.

IV. CONCLUSION

Based on the ideas of homeostatic control, a large number of behavioral models can be formulated which are able to control internal contradictions. This yields a high level of sustainability and adaptability to influence of various adverse factors. The basic idea is how much the active contradictions make use of the system's resources (energy ΔE, substance ΔV, information ΔI), moving in this way towards dangerous situations. It is also important to estimate how far these resources can be released in order to preserve purposefully systems sustainability and maintain homeostasis. In this way a contradiction is "set" inside the system in advance which is an additional possibility for carrying out system control in case of emergency and other severe situations.

An essential feature of ergatic systems is the strong relationship between the general functioning of the system and the behavior of its subsystems and elements. The integrity and dynamical sustainability of such systems are to a large extent guaranteed by full or partial compensation of opposed trends based on the principle of homeostasis. Starting from the idea of a single scientifically methodological apparatus based on the principles of functional sustainability, objective assessments of the information systems used in risky applications can be made.

REFERENCES

[1] Salvendy Gavriel, Handbook of Human Factors and Ergonomics [Hardcover] Hobokenq New Jersyq2012

[2] Зараковский Г., Павлов В., Закономерности функционирования эргатических системах., М., Радио и связь, 1987.

[3] John P. Van Gigch, Applied General Systems Theory, Harpercollins College Div; 1978

[4] System Engineering Handbook, Edited by Machol R. McGRW Book Company, New York, 1995

[5] Andonov A., Z. Hubenova, Functional sustainability of information-control complexes in critical applications, Sofia, 2011, p. 175, ISBN 978-954-12-0192-3.

[6] Andonov A., G.Cherneva, Z. Hubenova, The Problem of functionally Stable Information in Railway Transport, Proceedings of the V International Scientific Conference "Theoretical and Practical Issues in Transport", 2010, p.298-300

[7] Remington R., Shafto M., Freed M., Making, Human-Machine System Simulation a Practical Engineering Technique, HF Journal, N3, 2001.

[8] W. Ross Ashby, An Introduction to Cybernetics, (Chapman & Hall, London), Gloucester.

[9] Горский Ю.М. и др сб. „Гомеостатика живых, природных, технических и социальных систем", М., 1998.

[10] Степанов А.М. Основы медицинской гомеостатики, Воронеж: НПО "МОДЭК", 2001

[11] Getzov P.S., Angelov P.S., Hristov P.L. A Method for Control and Diagnostics of an Aircraft Automatic Control System using Simulation. International Conference AIRDIAG'91, 1991, p. 8, Poland.

[12] Cybernetics and Applied Systems, edited Constantin Virgil Negoita, Copyright, 1992

[13] Thomas B. Sheridan, William R. Ferrell, Men-Machine Systems: Information, Control and Decision Models of Human Performance, Hardcover, 2002

Interoperable and Safety Aspects of the Platform-Train Interface

Kalin Mirchev[1] and Denitsa Kireva-Mihova[2]

Abstract – **This article analyses the interoperability and safety requirements for the platform-train interface (PTI). Discrepancies between the European requirements and the Bulgarian national rules have been identified, as well as the impact of these inconsistencies on passengers safety is analysed.**

Keywords – **Safety, Interoperability, Platform-train interface.**

I. INTRODUCTION

Technical Specifications for Interoperability (TSIs) are standards produced in accordance with the Railway Interoperability Directives. To achieve interoperability of the national railways with the European railway system, compliance with the TSIs is required. [3]

Safety is one of the essential requirements for achieving interoperability of the railway transport. Boarding and alighting passengers on train is a process of increased hazards and potential of incidents at the PTI. The Railway Safety Directive requires monitoring of the Common Safety Indicators (CSIs) at European level and in case of drastic deterioration of an indicator, proper measures are to be introduced to improve it. Some of the CSIs monitored by European Union Agency for Railways (EUAR) and related to the PTI safety are: [2]

- Total number of persons seriously injured in accidents to persons caused by rolling stock in motion;
- Total number of passengers seriously injured in accidents to persons caused by rolling stock in motion;
- Total number of passengers seriously injured in accidents to persons caused by rolling stock in motion relative to train km;
- Total number of passengers seriously injured in accidents to persons caused by rolling stock in motion relative to passenger train km;
- Passengers seriously injured in accidents to persons caused by rolling stock in motion relative to passenger km;
- Total number of persons killed in accidents to persons caused by rolling stock in motion.

Passenger means any person, excluding members of the train crew, who makes a trip by rail. For accident statistics, passengers trying to embark/disembark onto/from a moving train are included [2].

Results of various reports show an increase in incidents with passengers at the PTI, some of them fatal. This type of incidents account for most of the total passenger fatality risk on the railway network and about one-fifth of the overall passenger fatality and weighted injury risk [12]. To manage such risk we must acknowledge the PTI as part of the system that makes up the railway and not an isolated issue. Spatial studies for the needs of the German railways (DB) and the British railways have been devoted to that issue. In [9] the safety integrity level related to the hazards raised at the PTI is calculated for an individual required by the National Safety Authority of Germany. In the study three types of vehicles are taken into account for calculation of the tolerable individual risk by using three different approaches per vehicle. As a result, the authors recommend the installation of monitoring devices to reduce the risk. Statistics of passenger accidents at the PTI, as well as analysis of the factors affecting the risk is presented in [12] for the United Kingdom. A web based Platform Train Interface Risk Assessment Tool is developed by RSSB in order to manage the safety level case by case. Similar risk model [14] for the needs of DB determines the risk and the necessary safety measures on the platform.

One of the main hazards at the PTI comes in the gap between the platform and the footstep of the rolling stock. This paper analyses the status of the technical rules defining the platform offset from the track centre with highlight to the Bulgarian case, where the possible lack of compliance is identified.

II. SYSTEM DEFINITION

The PTI is identified by the fixed installation system from one side, which is the platform and from the other side the rolling stock. The position of the platform, in particular the platform edge, in reference to the track is defined by a platform height measured from the rolling surface and a platform offset measured from the track centerline. The various rolling stock manufactures lead to great variety of systems serving the passengers by boarding and alighting on train. As a result comes the gap between the platform and the rolling stock. Figure 1 depicts the platform-train interface definition. [6]

Platform offset is the distance between the upper surface of the platform edge and the running edge of the nearest rail on the track adjacent to the platform, measured parallel to the plane of the rails.

Step is the vertical distance between the platform surface and the train stepping surface.

Gap is the horizontal distance between the edge of the platform and the edge of the train stepping surface.

[1]Kalin Mirchev is with Tinsa EOOD, 3A Nikolay Haytov Str., Sofia 1113, Bulgaria, E-mail: k_mirchev@tinsabg.com
[2]Dnitsa Kireva-Mihova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kireva@tu-sofia.bg
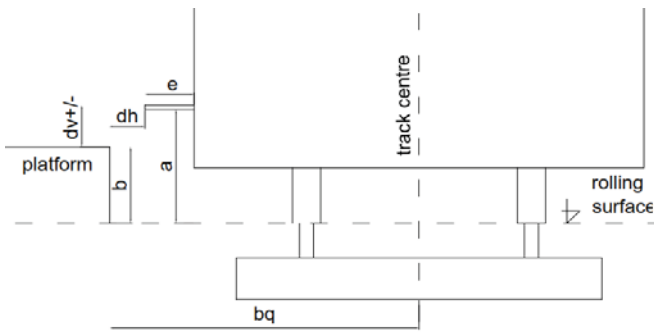
Fig. 1. Platform-train interface

where,

a   the height difference between the rolling surface and the boarding step

b   height of the platform

dh   distance from the platform edge to the boarding step

dv+/- the height difference between the platform surface and the boarding step

bq   distance from the platform edge to the track centreline (platform offset)

e   depth of the step, if applicable

III. RULE-BASED APPROACH ANALYSIS

At European level, the legislation applicable to determine the position of the platform edge is [4] and [6]. As per p. 4.2.9.2 and p. 4.2.9.3 of the technical specifications for interoperability relating to the 'infrastructure' subsystem (TSI INF) [4] the nominal height (h_q) shall be 550 mm or 760 mm for radius of 300 m or more, while the platform offset shall be calculated for the installation limit gauge (b_{qlim}) on the basis of the reference clearance profile G1 as defined in chapter 13 of [6] (Figure 2).
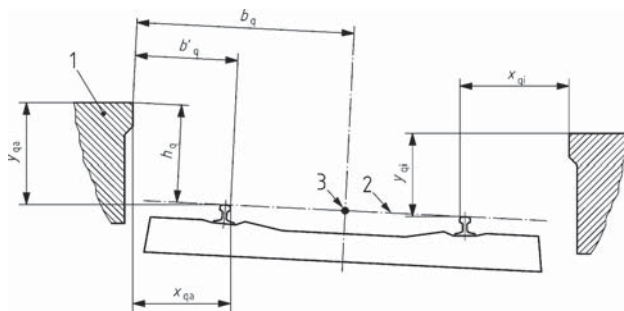
Fig. 2. Installation of the platform

Chapter 13 of [6] states that *"the platform shall be installed as close as possible to the passenger coaches whilst ensuring the safety of the rail traffic. It is important to limit the gap between the vehicle steps and the platform edges in order to provide acceptable stepping distances for passengers."* Following the provision of [6] the simplified formula for calculation of the platform offset in straight level line is

$$b_{qlim} = b_{cr} + \Sigma_{2cin} \qquad (1)$$

where,

$b_{cr} = 1620$ for heights from 550 mm to 760 mm according to the reference clearance profile G1;

Σ_{2cin} is calculated for installation limit gauge and takes into account the following coefficients of the allowances: track position error, cross level error, as well as oscillations; loading dissymmetry and suspension adjustment dissymmetry of the vehicle.

Based on the recommended coefficients of the allowances, the calculated value for b_{qlim} is 1650. For some administration this value may vary up to 1670, due to different coefficients of the allowances.

When the track along the platform is situated in curve, the platform height and the platform offset is changing based on the relative values y_{qi} and x_{qi} for the inside of the curve and y_{qa} and x_{qa} for the outside of the curve. The formulas for calculation of these measurements are determined in chapter 13 of [6]. Table 1 presents an example of how the platform edge installation changes as a function of curve and cant.

TABLE I
PLATFORM EDGE INSTALLATION CHANGES

Curve R=1400 m; Cant D=100 mm		
	b_{qlim}	y_q*
Platform inside	1690 mm	490 mm
Platform outside	1620 mm	610 mm

* It should be noted that the TSI INF [4] requires calculation according to [6] only for the platform offset. For the platform height [2] defines nominal values of 550 and 760 mm. Application of the requirements of the standard for calculating the platform height, in case the platform is situated in curve, depends on the national rules.

The step position for vehicle access and egress is defined in p. 4.2.2.11. of [5] as *"It shall be demonstrated that the point situated in the central position on the nose of the access step of each passenger access door on both sides of a vehicle in working order with new wheels standing centrally on the rails, shall be located inside the surface identified as 'step location"*. The step location defined in [5] is visualized as shaded area on Figure 3 a).

a)   b)
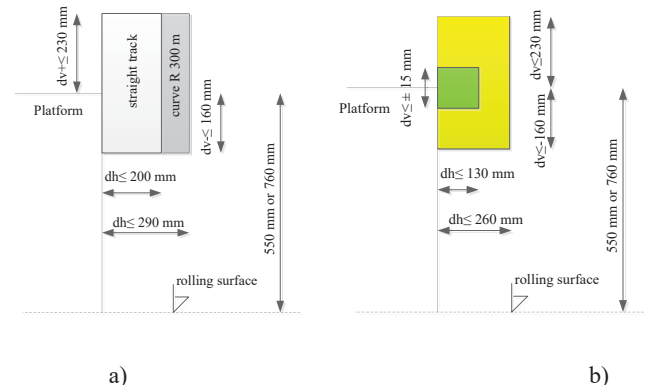
Fig. 3. Step location

The values of *dh*, *dv+* and *dv-* on Figure 3 a) is considered as reference requirements, which satisfy interoperable

operation of the line. More detailed study for the risk analysis of the PTI [9] determines the green area shown on Figure 3 b) as not critical for the access/egress step. Beyond that area some measures must be taken.

At national level, the platform installation is defined by Ordinance No. 55 [7] and Ordinance No. 58 [8]. The platforms shall be installed along the track in straight level line or in curve with radius equal or greater than 600 m, without cant or with cant equal or less than 100 mm. They are classified as law, semi-high and high with a platform height and platform offset as given in Table II.

TABLE II
PLATFORM CLASSIFICATION

	Platform height*	Platform offset*
Law	300 mm	1650 mm
Semi-high	760 mm	1750 mm
High	1100 mm	1750 mm

*Values valid for platforms on a straight level track. For platforms on a track with a curve these values change according to the rules defined in [8].

The position of a low and a high platform is also defined in [4], p. 7.7.3 as a *specific case* for Bulgaria.

New projects for modernization of the railway lines in Bulgaria have to meet the TSI requirements, as well as the national rules. The practice shows that in the common case the platforms are installed at platform height of 550 mm and platform offset of 1750 mm, as the nominal value for the platform height of 550 mm is not defined in the Bulgarian legislation.

IV. RISK-BASED APPROACH

The rule-based approach analysis identified cases, where the technical requirements for an interoperable gap are not satisfied. In such cases additional measures are needed to maintain the required safety integrity level.

Deviation from the interoperability requirements affects safety requirements, posing an additional danger to passengers. The risk of these hazards should be further analyzed by establishing a risk model taking into account all risk factors. Proposing the risk-based approach different factors to the risk have been identified, such as:

- type of the platform
- width of the platform
- visibility indicators
- weather
- light conditions
- profile of the passengers (age, gender)
- intoxication
- day of the week/time of the day.

By interaction of the two systems, from the one hand PTI and from the other hand a passenger, the hazards appear on the boarder of the system (Figure 4). In case of external event, e.g. the train starts moving, an accident is possible. Such accident could be fatal or not. The scope of that approach is the hazard of accidents due to the gap.
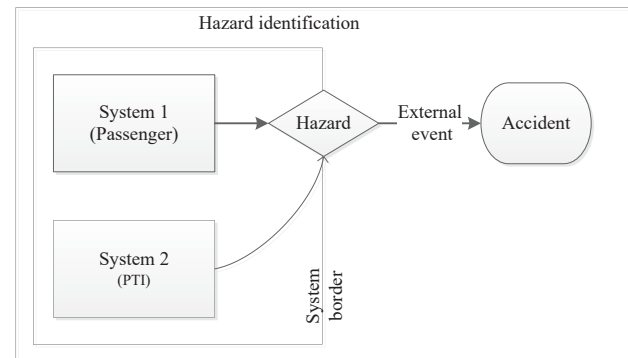
Fig. 4. Risk-based model

For the risk assessment the estimated severity (S) and the estimated frequency (K) shall be taken into account. K is function of:
- the frequency and the duration of the exposition of the hazard (F);
- the probability of the hazard to happen (W)
- the possibility to avoid the accident or to reduce the damage (V)

$$Estimated\ risk = (S)X\ (K)\begin{cases}F\\W\\V\end{cases} \qquad (2)$$

$$Estimated\ risk = \frac{SxFxW}{V} \qquad (3)$$

We can consider V as factor of reduction, which is reciprocal to the possibility to avoid the accident. A fault tree diagram is applicable for determination of V (Figure 5).
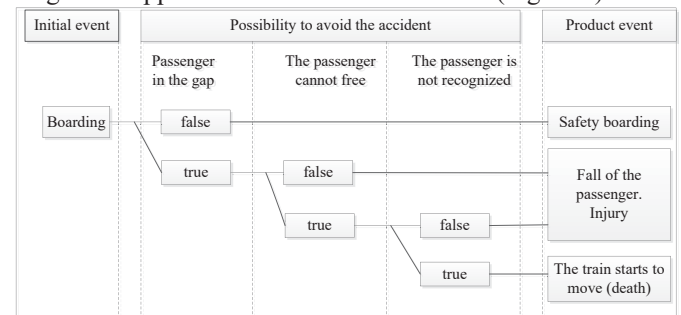
Fig. 5. Fault-tree diagram

Applying the equation (3) for a common evaluation of the tolerable individual risk we can go to the equation (5).

$$THR = \frac{TIR}{SxFxWxV} \qquad (5)$$

where,
THR – Tolerable Hazard Rate
TIR - Tolerable Individual Risk of death per an individual
Explicit for the Bulgarian railways the following hazards have been identified to be used in the risk assessment process:
- opening a door earlier
- opening a door on the wrong side

- a door remains open
- forcibly opening a door by a passenger
- a boarding step doesn't open or closes earlier

V. Conclusion

The platform-train interface poses risk to passengers (including members of the public) and staff. This risk is comprised of high likelihood but low severity hazardous events (slips, trips and falls) and low likelihood but high severity hazardous events (dragging, falling from the platform and being struck by the train, and being struck by the train when standing on the platform).

The differences between the national technical requirements of Bulgaria and the European rules for PTI parameters are identified. How this nonconformities affects the interoperability of the railway transport and the safety of the passengers are analysed.

Assuming the presented analysis of the applicable rules for determination of the platform installation in Bulgaria, the following conclusion could be made:

- The current national rules allow a big variety of platform installation.
- There is a gap between the European and the national legislation concerning the platform offset.
- The missing national rule for platform height of 550 mm is assumption for different approaches for determination of the platform installation in curves.
- The track condition and the maintenance plan are not taken into account.

All these factors are prerequisite for a step position beside the uncritical area, where the safe operation of PTI is assured. (Fig. 2 b). The rule based approach does not take into account significant factors for the PTI operation like passenger flow, staffed/unstaffed stations, safety measures taken in the station. All these factors confirm the need of risk assessment tool for the PTI case by case. Such tool could consist of a control system, which will manage the risk case by case. Input of the tool will be the external conditions, which have an impact to the risk, and the type of the operation train. Based on the risk assessment results, the outputs will lead to explicit established measures in order to reduce the risk to an acceptable level.

Acknowledgement

References

[1] Annual Safety Performance Report, *A reference guide to safety trends on GB railways*, 2017/18
[2] Railway Safety Directive (2004/49/EC)
[3] Railway Interoperability Directive (2008/57/EC)
[4] Regulation (EU) No 1299/2014 of 18 November 2014 on the technical specifications for interoperability relating to the 'infrastructure' subsystem of the rail system in the European Union.
[5] Regulation (EU) № 1300/2014 on the technical specifications for interoperability relating to accessibility of the Union's rail system for persons with disabilities and persons with reduced mobility.
[6] EN 15273-3:2013 Railway applications - Gauges - Part 3: Structure gauges
[7] Ordinance No.55 of 29.01.2004 for the design and construction of railway lines, railway stations, railway level-crossings and other railway infrastructure elements.
[8] Ordinance No.58 of 2.08.2006 on the rules of the technical operation, traffic operations and signalling in the railway transport.
[9] Risikoanalyse zur Einstiegssituation bei Schienenfahrzeugen in Abhängigkeit des Abfertigungsverfahrens, Wiesbaden, im November 2009, Die Ingenieurwerkstatt GmbH.
[10] Validierungeines semi-quantitativen Ansatzes zur Risikobeurteilung in der Eisenbahntechnik, Dipl.-Ing. Sonja-Lara Bepperling, Technischen Universität Carolo-Wilhelmina zu Braunschweig, 2008.
[11] Technische Regeln Spalt, Verbund der Bahnindustrie in Deutschland, 2010.
[12] Risk at the platform-train interface, RSSB, 2013.
[13] An introduction to the Platform Train Interface Risk Assessment Tool, RSSB, 2016.
[14] https://www.ebp.ch/de/pdf/generate/node/893.

Control of production and inventory in the automotive industry for multi customer and multi products

Ivan Djordjevic[1], Gordan Stojic[2] and Dobrila Petrovic[3]

Abstract – **Different factors have an uncertain impact on the material flow time in the enterprise, such as production time, time to store goods and preparation time for delivery. A new model of aggregate production planning (APP) is developed as a fuzzy linear programming (FLP) model with experiments on real world data.**

Keywords – **Aggregate production planning, fuzzy optimization, uncertain production, uncertain customer demand.**

I. INTRODUCTION

Aggregate production planning (APP) is one of the most important part of operations management in competitive supply chains. It concerns matching supply with forecasted customer demand over a planning period, which is usually one year in practice. Generally, the aim is to determine required resources, which include production rate, warehouse levels, work force level, overtime, etc., in such a way as to meet customer demand.

In the literature, it has been assumed most often, that all the parameters which are associated with the APP process, such as customer demand, processing times, production capacities etc., are deterministic in nature (for example, [1]). In order to handle uncertainties which characterise real world APP environments, and a randomness in customer demand, in particular, various stochastic optimisation models have been proposed [2]. Furthermore, one can find in the literature that different types of uncertainties encountered in APP problems, such as imprecise demand, production capacities with tolerance, fuzzy processing times can be specified by production managers using imprecise linguistic terms. They have led to the development of a number of fuzzy APP models and applications of fuzzy optimisation techniques[3].

In this paper, we propose a new fuzzy model for optimal APP in the presence of uncertainty. The novelty of the model is that the objective is to minimise the fuzzy total time required for production, storing manufactured products and their preparation for delivery to the customer. We introduce uncertain factors to take into consideration uncertainty in customer demand which is forecasted and can fluctuate around these values and uncertainty in manufactured quantities. As all the time parameters listed above, customer demand deviations

and the parameters which describe the output of manufacturing process are fuzzy, both the associated objective function and constraints become fuzzy, too. We adapt and apply one of the methods for transforming the fuzzy linear programming optimisation model, with the fuzzy objective function and fuzzy constraints, into a crisp optimisation model with both the crisp objective function and crisp constraints [4].

The paper is organized as follows. Literature review on APP models and methodologies used and methods of modelling uncertain APP parameters is presented in Section II. Problem statement is given in Section III. The fuzzy aggregated production and inventory planning model are described in Section IV, while Section V contains case study and analyses of results of different experiments carried out using the proposed model. The conclusion is given in Section VI.

II. LITERATURE REVIEW

It is well recognised in the literature that treating uncertainty in APP models in an appropriate way brings an advantage to handling real world APP problems and brings them nearer to the practice [3]. Majority of the APP models handle uncertainty using a classic probability theory approach, and consider only one type of uncertainty which is based on randomness and frequency of a random event occurrence.

Linear mixed integer programs (MIPs) were developed to solve two production planning problems with demand uncertainty [5], when the manufacturer had a flexibility to accept or reject an order. The MIP method in production planning problems for multi-period and multi-items in make-to-order manufacturing system was used in [6]. Avraamidou and Pistikopoulos [7] developed a bi-level mixed integer linear programming model for a supply chain under demand uncertainty.

A very important issue in modern production planning is energy consumption. Today, the most manufacturers invest significant money assets to optimise and reduce energy consumption. In [8], a multi-objective linear programming problem with three objective functions including operational expense, energy expense and carbon emission, was analysed.

Zadeh proposed a new approach to handle different types of uncertainty, by introducing the concept of fuzzy sets [9]. It has been demonstrated in the literature that fuzzy sets can be successfully applied to modelling uncertainty where available information is vague or cannot be defined precisely due to the limited knowledge. One can find some good examples in the literature on how fuzzy sets are applied in supply chain management problems, for example in supply chain partners' collaboration [10], in MRP (material requirement problems) [3], in serial supply chains [11], etc. Tang et al. considered both uncertainty in customer demand and

[1]Ivan Djordjevic is with the Faculty of Technical Sciences University of Novi Sad, 6 Square of Dositej Obradovic, Novi Sad, 21101, Serbia, E-mail: ivan.djordjevic87@gmail.com.
[2]Gordan Stojic is with the Faculty of Technical Sciences University of Novi Sad, 6 Square of Dositej Obradovic, Novi Sad, 21101, Serbia. E-mail:gordan@uns.ac.rs.
[3]Dobrila Petrovic is with the Faculty of Engineering, Environment and Computing University of Coventry, Priory Street, Coventry, CV1 5FB, United Kingdom, E-mail: d.petrovic@coventry.ac.uk;

production capacity and modelled them as fuzzy values in a multi-product APP model [12] A fuzzy multi-objective mixed-integer non-linear programming model for a supply chain was proposed in [13]. Fuzzy customer demand was considered in three objective functions that minimised the total supply chain cost, total maximum product shortages, and the rate of changes in human resources.

III. PROBLEM STATEMENT

A problem is to generate the optimum aggregate production and inventory plan for a supplier for a given planning time horizon. The supplier operates in a "make-to-order" manner and has to prepare a production and inventory plan in such a way as to satisfy customer demand and optimise an associated performance measure in the considered time horizon.

The planning time horizon is discretised into a series of subsequent discrete time periods. The APP determines 3 quantities to be generated for each time period in the planning time horizon: (1) optimal production quantity to be manufactured, (2) the safety stock quantity that should be kept in the warehouse and (3) the quantity that should be delivered to the customer.

If the same production line is used for manufacturing of different products for more than one customer, an efficient use of the production line is of paramount importance for the production process.

All these uncertainties have to be taken into account when generating the optimal production and inventory plan.

IV. FUZZY AGGREGATED PRODUCTION AND INVENTORY PLANNING

A. Notation

The following notation is used:
i – index of a time period in a planning horizon, $i = 1,...,n$,
D_i – customer demand in period i, $i = 1,...,n$,
$\tilde{t}_p$ – fuzzy production time per unit of product (in minutes), with trapezoidal membership function $\tilde{t}_p = (t_{p\,1}, t_{p\,2}, t_{p\,3}, ...)$,
$\tilde{t}_s$ – fuzzy warehouse storing time per unit of product (in minutes), with trapezoidal membership function $\tilde{t}_s = (t_{s\,1}, t_{s\,2}, t_{s\,3}, t_{s\,4})$,
$\tilde{t}_t$ – fuzzy preparation time for shipping to customer per unit of product (in minutes), with trapezoidal membership function $\tilde{t}_t = (t_{t\,1}, t_{t\,2}, t_{t\,3}, t_{t\,4})$,
$\widetilde{w}_i^d$ - fuzzy factor for uncertain customer demand deviation from forecasted value in period i, $i = 1,...,n$, with triangular membership function $\widetilde{w}_i^d = (w_{i\,l}^d, w_{i\,m}^d, w_{i\,u}^d)$,
$\widetilde{w}_i^p$ - fuzzy factor for uncertain production quantity output in period i, $i = 1,...,n$, with triangular membership function $\widetilde{w}_i^p = (w_{i\,l}^p, w_{i\,m}^p, w_{i\,u}^p)$,
T^l – minimum "days of inventory" in the warehouse,
T^u – maximum "days of inventory" in the warehouse,
C – machine capacity.
Decision variables:

P_i – quantity manufactured in period i,
Ss_i – safety stock in period i,
Q_i – quantity delivered to customer in period i.

B. Fuzzy APP LP model

The objective is to minimize the total material lead time $\tilde{Z}$ including the production time $\tilde{t}_p P_i$, warehouse time $\tilde{t}_s Ss_i$ required for storing safety stock of manufactured products and time for preparation of delivery to customers $\tilde{t}_t Q_i$, as follows:

$$min \, \tilde{Z} = \sum_{i=1}^{n} \tilde{t}_p P_i + \tilde{t}_s Ss_i + \tilde{t}_t Q_i \qquad (1)$$

The following constraints are considered:

Uncertain customer demand $\widetilde{w}_i^d D_i$ in each time period i is satisfied using the uncertain production $\widetilde{w}_i^p P_i$ or safety stock Ss_i:

$$Ss_i + \widetilde{w}_i^p P_i \geq \widetilde{w}_i^d D_i, \, i = 1, ..., n \qquad (2)$$

The safety stock Ss_{i+1} in each time period $i + 1$ is equal to the stock in the previous period Ss_i increased by uncertain production in the previous period, $\widetilde{w}_i^p P_i$, and reduced by uncertain customer demand, i.e., quantity delivered to the customer in the previous period, $\widetilde{w}_i^d D_i$:

$$Ss_{i+1} = Ss_i + \widetilde{w}_i^p P_i - \widetilde{w}_i^d D_i, \, i = 1,...,n \qquad (3)$$

Installed machine capacity C produces uncertain $\widetilde{w}_i^p P_i$ units per period i:

$$\widetilde{w}_i^p P_i \geq 0, \, i = 1,...,n \qquad (4)$$
$$C \geq \widetilde{w}_i^p P_i, \, i = 1,...,n \qquad (5)$$

The safety stock Ss_i in period i is defined by a supplier's target to cover between T^l and T^u days of uncertain customer demand $\widetilde{w}_i^d D_i$ in that period:

$$Ss_i \geq T^l \widetilde{w}_i^d D_i, \, i = 1,...,n \qquad (6)$$
$$T^u \widetilde{w}_i^d D_i \geq Ss_i, \, i = 1,...,n \qquad (7)$$

The delivery Q_i in each period i must be equal to uncertain customer demand $\widetilde{w}_i^d D_i$ in order to operate with the maximum service level - 100%.

$$Q_i = \widetilde{w}_i^d D_i, \, i = 1,...,n \qquad (8)$$

Decision variables P_i, Ss_i and Q_i in each time period i are non-negative:

$$P_i, Ss_i, Q_i \geq 0, \, i = 1,...,n \qquad (9)$$

C. From the fuzzy APP optimization model to a crisp APP optimization model

We applied a method developed by Jimenez et al [4] to transform the fuzzy APP model into a crisp APP model. We adapted it in such a way as to handle fuzzy parameters in the objective function with trapezoidal membership functions..

The transformation includes 3 steps as follows.

Step 1. The decision maker specifies the feasibility degree β of constraint satisfaction he/she is ready to accept. Let us assume that the lowest feasibility degree that the decision maker is ready to consider is *Neither acceptable nor unacceptable solution* - $\beta = 0.5$ of course, it can be changed to any other feasibility degree β from interval [0, 1].

The crisp optimisation model is solved iteratively for each feasibility degree $\beta = 0.5., 0.6, \ldots, 0.9, 0.95, 0.99$ and 1 where each solution is β-feasible, i.e., the minimum of feasibility achieved for all constraints is β. The β-feasible solution Pi, Ssi and $Qi, i = 1, \ldots, n$ are found as follows.

First, fuzzy parameters $\tilde{t}_p, \tilde{t}_s$ and $\tilde{t}_t$ in the objective function are mapped into their crisp expected values. They are calculated as the middle points of the Expected intervals.

Step 2. The decision maker specifies tolerance thresholds to obtained fuzzy objective function values achieved for different β-satisfaction of constraints. The shortest time $\underline{Z}$ will be achieved for the lowest constraints' satisfaction $\beta = 0.5$ and the longest time $\overline{Z}$ for the highest constraints' satisfaction $\beta = 1$. We assume that the tolerance function $\tilde{G}$ is linear between these two tolerance thresholds, the shortest time $\underline{Z}$ and the longest time $\overline{Z}$. The membership function is:

$$\mu_{\tilde{G}}(z) = \begin{cases} 1, & z < \underline{Z} \\ \frac{\overline{Z} - z}{\overline{Z} - \underline{Z}}, & \underline{Z} \leq z \leq \overline{Z} \\ 0, & z > \overline{Z} \end{cases} \quad (10)$$

We propose the following formula to calculate tolerance $K_{\tilde{G}}(\tilde{Z}(\beta))$ to obtained objective function value $\tilde{Z}(\beta)$ when the feasibility of constrains is β.

$$K_{\tilde{G}}\left(\tilde{Z}(\beta)\right) = \frac{\overline{Z} - EV(\tilde{Z}(\beta))}{\overline{Z} - \underline{Z}} \quad (11)$$

Step 3. Balance between the feasibility degree of constraints β and the satisfaction degree of solution, $K_{\tilde{G}}\left(\tilde{Z}(\beta)\right)$, is calculated as:

$$\beta \cdot K_{\tilde{G}}\left(\tilde{Z}(\beta)\right) \quad (12)$$

The solution $Pi, Ssi, Qi, i = 1, \ldots, n$ which achieves the highest balance $\max\limits_{\beta=0.5, 0.6, \ldots, 0.9, 0.95, 0.99, 1} \beta \cdot K_{\tilde{G}}\left(\tilde{Z}(\beta)\right)$, is recommended.

V. CASE STUDY

We considered a first tier supplier in the automotive industry located in Serbia, which has become an increasingly important industrial sector in the recent years. The factory supplies window regulators to a number of European car manufacturers. We analysed one production line which manufactures multi products for two different customers. All products belong to the same product family. They are packed in two types of plastic containers specified by the customers. The developed fuzzy APP model is applied to determine the minimal time required for production and logistics processes. The planning horizon is selected to be a period of 12 weeks. Customer demand forecast for 12 weeks is a typical mid-term forecast used in the automotive industry for production planning. A longer period of customer demand has huge uncertainty and is not reliable for sustainable production planning.

The result of fuzzy APP model is presented in Table I. The calculation is performed using formulas (1-12) and simplex method of classical LP solver. An algorithm is developed in software Visual Studio 2015 in C++ programming language. The performance of computer: Intel processor i3-2120 (3M Cache, 3.30 GHz), 8G RAM memory (2133 MHz).

TABLE I RESULTS OF THE FUZZY APP MODEL

Feas. degr. β	Decision variables			Fuzzy objective function value				Tolerance $\mu_{\tilde{G}}(z)$	Balance $K_{\tilde{G}}(Z()$	Objec. func. value z
	$\sum P_i$	$\sum Ss_i$	$\sum Q_i$	z_1	z_2	z_3	z_4			
0.5	188776	123372	194883	54839	58009	62338	68889	0.733	0.3665	61382
0.6	194107	126022	198581	56235	59489	63927	70646	0.666	0.3995	62948
0.7	199546	128672	202279	57654	60992	65541	72429	0.598	0.4183	64538
0.8	**205098**	**131323**	**205977**	**59094**	**62519**	**67180**	**74241**	**0.528**	**0.4225**	**66153**
0.9	210766	134141	209674	60562	64074	68851	76089	0.458	0.4118	67799
0.95	213644	135620	211523	61305	64863	69698	77027	0.422	0.4006	68634
0.99	215968	137007	213003	61909	65503	70386	77791	0.393	0.3886	69313
1	216552	137365	213372	62061	65664	70559	77983	0.385	0.3852	69484

Fuzzy factor $\tilde{w}_i^p$ is symmetrical triangular fuzzy number and modeled by logistics expert in enterprise as 10% of production output deviation (0,9, 1, 1,1). Fuzzy factor $\tilde{w}_i^p$ is symmetrical triangular fuzzy number and obtained as previous calculation based on customer demand deviation in enterprise in period of 12 weeks before testing time window of 12 weeks:

$$f_i = \frac{D_i}{\sigma}, i = 1, \ldots, 12 \quad (13)$$

Where f_i is demand fluctuation, D_i customer demand prediction different for every week i, and σ is standard deviation of D_i, $i = 1, \ldots, 12$. Production time $\tilde{t}_p$, time for safety stock storing in warehouse $\tilde{t}_s$, and time for preparation of shipment to customer $\tilde{t}_t$ are measured in enterprise and presented as nonsymmetrical trapezoidal fuzzy numbers: $\tilde{t}_p = (0,20, 0,21, 0,23, 0,25)$, $\tilde{t}_s = (0,020, 0,023, 0,028, 0,04)$, $\tilde{t}_t = (0,075, 0,077, 0,082, 0,086)$ minutes per unit product. The target of safety stock keeping days used in calculation is between $T^l = 3$ and $T^u = 5$ days. The machine capacity is $C = 19000$ pcs/week for 5 working days in a week.

The optimal value of objective function in fuzzy APP model is 66153 minutes for whole time window of 12 weeks (5512 minutes/weekly; 1102 minutes/daily; 18,4 h/daily) for feasibility degree $\beta = 0,8$ (Table I). Testing is performed in enterprise for both common used strategies of production/inventory planning and compared with realized production plan in enterprise and result of fuzzy APP model. For comparing purpose an initial safety stock value in week 1 for both strategies and realized production plan is supposed $Ss_1 = 10000$ pcs. The result of fuzzy APP model for week 1 is $Ss_1 = 17031$ pcs.

TABLE II COMPARING WITH TWO STRATEGIES OF PRODUCTION AND INVENTORY PLANNING

Week, i	Prod.planning with using maximal prod. capacity (C=19000 pcs/week)			Prod. planning with using safety stock target (3 days of coverage)			Results of experiment from fuzzy APP model			Realised production plan in enterprice		
	Ss_i	P_i	Q_i	Ss_i	P_i	Q_i	Ss_i	P_i	Q_i	Ss_i	P_i	Q_i
1	10000	19000	16516	10000	15962	16516	17031	17070	16516	10000	17450	16516
2	12484	19000	15744	9446	16424	15744	10585	16585	15744	10934	17800	15744
3	15740	19000	16877	10126	16424	16877	11426	19588	16877	12990	17750	16877
4	17863	19000	16122	9673	19667	16122	14137	19089	16122	13863	18350	16122
5	20741	19000	22030	13218	15651	22030	17104	12564	22030	16091	17740	22030
6	17711	19000	11398	6839	13637	11398	7638	17272	11398	11801	17850	11398

7	253 13	190 00	151 30	907 8	161 85	151 30	135 13	194 43	151 30	182 53	159 50	151 30
8	291 83	190 00	168 88	101 33	199 43	168 88	178 26	181 06	168 88	190 73	168 40	168 88
9	312 95	190 00	219 80	131 88	180 87	219 80	190 44	182 45	219 80	190 25	175 30	219 80
10	283 15	190 00	154 92	929 5	159 52	154 92	153 09	182 64	154 92	145 75	174 40	154 92
11	318 23	190 00	162 58	975 5	163 28	162 58	180 81	182 62	162 58	165 23	175 30	162 58
12	345 65	190 00	163 74	982 4	115 50	163 74	200 85	176 10	163 74	177 95	183 60	163 74
Σ	275 033	228 000	200 809	120 576	195 809	200 809	181 778	205 098	200 809	180 923	210 590	200 809

In Table II are presented data for both strategies of production and inventory planning in enterprise. For every strategy and real-world data the delivered quantity is the same $\sum Q_i = 200809$ pcs of goods for period of 12 weeks. The quantity of safety stock $\sum Ss_i$ and produced parts $\sum P_i$ are different. They are highest in strategy of using maximal capacity 275033 and 228000 pcs, respectively. The safety stock quantity is lowest in strategy with 3 days of inventory, $\sum Ss_i = 120576$ pcs.

An objective function value (Table III) which represents the total time required for all production and logistics operation in delivery goods to customers is the highest and in same time the most unfavorable in strategy with maximal capacity usage; it is 74476 minutes for 12 weeks (6206 minutes weekly, 1241 minutes daily, 20,7 hours daily). The lowest total time is in strategy with using safety stock target 62760 minutes for 12 weeks (5230 minutes weekly, 1046 minutes daily, 17,4 hours daily). The outcome of fuzzy APP model shows better result than strategy prod. planning using maximal prod. capacity and realized prod. plan in enterprise. Comparing with strategy with using safety stock target of 3 days the outcome is worse because the safety stock target used in fuzzy APP model is between 3 and 5 days. Normally the objective function is higher.

TABLE III OBJECTIVE FUNCTION VALUE AND INVENTORY COVERAGE

Prod.planning with using maximal prod. capacity (C=19000 pcs/week)		Prod. planning with using safety stock target (3 days of coverage)		Results of experiment from fuzzy APP model		Realized production plan in enterprise	
Objec. funct.	Inven. cover.	Objec. funct.	Inven. cover.	Objec. funct.	Inven. cover.	Objec. funct.	Inven. cover.
74476	6,9	62760	3	66640	4,5	67822	4,6

VI. CONCLUSION

The review of the published APP models in literature showed that they did not consider and analysed the material flow time in the APP problems. All of the APP models have been developed to minimize operational cost in manufacturing, dealing with impacts on production, inventory or delivery costs. However, in some industrial sectors, the material flow time is a very important factor and cannot be neglected, because it has a big impact on the total measure of manufacturer performance. We consider a real world APP problem in the automotive industry and develop a fuzzy LP model which considers a material flow as the measure of performance. Results obtained using the proposed APP model are better compared to the practical results; the total material flow time is

shorter using the proposed APP model. Practical application of the APP model in the factory would contribute to optimised production and inventory plan with higher customer satisfaction with the service level. Finally, the cash flow in the factory can be much improved.

REFERENCES

[1] S. J. Nam and R. Logendran, "Aggregate production planning - a survey of models and methodologies", European Journal of Operational Research 61, 255-272, 1992.

[2] B. R. Feiring, "Production planning in stochastic demand environments", Mathematical and Computing Modelling 15, 91-95, 1991.

[3] J. Mula, R. Poler, J. P. Garcia-Sabater and F. C. Lario, "Models for production planning under uncertainty: A review", International Journal of Production Economics 103, 271-285, 2006.

[4] M. Jimenez, M. Arenas, A. Bilbao and M. V. Rodriguez, "Linear programming with fuzzy parameters: An interactive method resolution", European Journal of Operational Research 177, 1599-1609, 2007.

[5] T. Aouam, K. Geryl, K. Kumar, N. Brahimi, "Production planning with order acceptance and demand uncertainty", Computers and Operations Research 91, 145-159, 2018.

[6] H. Salamati-Hormozi, Z-H. Zhang, O. Zarei, R. Ramezanian, "Trade-off between the costs and the fairness for a collaborative production planning problem in make-to-order manufacturing", Computers & Industrial Engineering 126, 421-434, 2018.

[7] S. Avraamidou and E. Pistikopoulos, "A Multiparametric Mixed-integer Bi-level Optimization Strategy for Supply Chain Planning Under Demand Uncertainty", IFAC-PapersOnLine, Volume 50 Issue 1, 10178-10183, 2017.

[8] M. Modarres and E. Izadpanahi, "Aggregate production planning by focusing on energy saving: A robust optimization approach", Journal of Cleaner Production 133, 1074-1085, 2016.

[9] L. A. Zadeh, "Fuzzy sets", Information and Control 8, 338–353, 1965.

[10] M. Wulan and D. Petrovic, "A fuzzy logic based system for risk analysis and evaluation within enterprice collaboration", Computers in Industry 63, 739-748, 2012.

[11] D. Petrovic, R. Roy and R. Petrovic, "Supply chain modelling using fuzzy sets", International Journal of Production Economics 59, 443-453, 1999.

[12] J. Tang, D. Wang and R. Y. K. Fung, "Fuzzy formulation for multi-product aggregate production planning", Production planning and Control 11 (7), 670-676, DOI: 10.1080/095372800432133, 2000.

[13] N. Gholamiana, I. Mahdavia and R. Tavakkoli-Moghaddam, "Multi-objective multi-product multi-site aggregate production planning in a supply chain under uncertainty: fuzzy multi-objective optimisation", International Journal of Computer Integrated Manufacturing, 29 (2), 149–165, 2016.

MEASUREMENT SCIENCE AND TECHNOLOGY

Analysis of Measurement Uncertainty in Wireless Sensor Network based Power Quality Measurement

Milan Simić[1], Goran Miljković[2], Dragan Živanović[3], Dragan Denić[4], Zivko Kokolanski[5]

Abstract – **Metrological evaluation of software supported experimental system based on wireless sensor network (WSN), applicable in measurement of electrical power quality (PQ) parameters and disturbances, is presented in this paper. System includes PC based generator of reference waveforms, software application for measurement of standard PQ parameters and two microcontroller based wireless sensor modules for transmitting and receiving of measurement results. Measurements of basic quality parameters for reference waveforms are performed using the control software application in LabVIEW environment. For communication is used wireless sensor network based on the communication standard IEEE 802.15.4 (Zigbee). Metrological assessment includes the calculation of measurement uncertainty components and final presentation of measurement uncertainty budget. Procedure for calculation of Type A, Type B, combined and expanded measurement uncertainty, according to the Guide to the Expression of Uncertainty in Measurement, is supported using the professional instrumentation with high accuracy levels.**

Keywords –**Measurement uncertainty, Wireless sensor network, Power quality measurement, Virtual instrumentation.**

I. Introduction

Having in mind the challenges of the modern smart grids and great importance of PQ problems, in the recent years special attention is given to development of microprocessor based sophisticated measurement systems for PQ monitoring. In the last decade especially attractive are virtual instruments, which are well suited for development of flexible computer supported measurement systems. Virtual instruments can be successfully used for research and scientific purposes [1-3]. In order to satisfy the specified level of the measurement uncertainty and characteristics, devices for measurement of PQ parameters must be followed by appropriate metrological evaluation. Measurement instruments can be used as single devices at specific points in power distribution network. Alternatively, number of separated devices can be combined

in distributed measurement system, for permanent monitoring and analysis of quality in power distribution networks.

Advances in wireless communications and electronics lead to development of wireless sensor networks (WSN), as low cost, low power and multifunctional sensors. These WSN sensor nodes are small and able to sense various data, process this data and communicate with each other or with central base station. Basic advantages of WSN systems are: self organizing capabilities, short range broadcast applications, changing of network topology in order to avoid fading and potential failures, including significant savings in energy consumption, transmission and computing capabilities [4].

WSN based networks can be used in various military applications, environmental measurements of various physical quantities, home intelligence, prevention of many disasters, surveillance, medical care purposes and others [5], [6].

Significant segment in real-time implementation of WSN for various measurement purposes is metrological assessment. Procedure for analysis of measurement uncertainty shown in this paper includes calculation of measurement uncertainty components and presentation of total uncertainty budget in accordance with relevant document - Guide to the Expression of Uncertainty in Measurement [7], defined by International Organization for Standardization - ISO. For this purpose are used reference PQ signal generator for simulation of typical network disturbances and experimental WSN based system for measurement of standard PQ parameters, presented and described in the previously published papers [8], [9].

II. Configuration of Experimental System for WSN based PQ measurement

Hardware configuration of experimental system, developed for WSN communication based measurement of standard PQ parameters and disturbances, is presented in Fig 1. Software supported generator of reference PQ waveforms is applied for generation of reference three-phase voltage waveforms. This generator is presented and described in previously published paper [8]. It can generate long time and short time reference waveforms, including special functions for simulation of various network disturbances, typical for real electrical power distribution networks. Procedure includes computer with the LabVIEW control application and data acquisition board PCIe NI 6343, supported with standard connection block SCB-68A. Control software application implemented on PC platform performs acquisition, measurements, recordings and graphical presentations of measurement data (RMS voltage values, signal frequency values and levels of high-order harmonics).

WSN communication for transmission of measurement results includes two WSN modules (transmitter and receiver), with Cortex M3 architecture, called the SPaRCMosquito v.2.

[1]Milan Simić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milan.simic@elfak.ni.ac.rs.

[2]Goran Miljković is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: goran.miljkovic@elfak.ni.ac.rs.

[3]Dragan Živanović is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.zivanovic@elfak.ni.ac.rs.

[4]Dragan Denić is with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.denic@elfak.ni.ac.rs.

[5]Zivko Kokolanski is with the Ss. Cyril and Methodius University, Faculty of Electrical Engineering and Information Technologies, 1000 Skopje, Macedonia, E-mail: kokolanski@feit.ukim.edu.mk.

Fig. 1. Hardware configuration of experimental system for WSN communication based electrical power quality measurement

Communication and transferring of the measurement results between application software for PQ measurement and WSN module on transmitter side of experimental system is provided using standard USB interface. WSN base module is supported with communication standard IEEE 802.15.4, also known as Zigbee. On receiver side of experimental measurement system VISA drivers from LabVIEW software are used for providing USB communication and transferring of results from receiver WSN module to laptop computer for further data analysis.

Basic functions enabled using software supported PQ signal generator are: definition of nominal amplitude and frequency values, definition of signal sample rate and duration of final test sequence, generation of noise with Gaussian distributed amplitude, variations of nominal signal frequency value, slow variations of signal amplitude value with defined frequency of variations, definitions of DC offset, voltage swell and voltage sag and possibilities for generation of high-order harmonic components [8]. Definition of reference waveforms, with specified levels of signal disturbances, can be performed directly inside the control front panel and block diagram of software application. Each category of signal disturbances can be defined and generated by separate functional segments. Individual disturbances can be combined and unified in form of final complex sequence, according to basic requirements of relevant European PQ standard EN 50160 [10].

Separated segment of control functions is used for selection and variation of amplitude levels related to individual high-order harmonics. Front panel of LabVIEW virtual instrument, for presentation and measurement of basic parameters related to voltage test waveform generated with certain level of signal harmonic components, is presented in Fig. 2. Specific voltage waveforms are generated with nominal frequency value of 50 Hz and normalized RMS voltage value of 1 V. Shown control software is implemented on transmitter side of experimental system. This application enables simultaneous presentation of voltage waveform, tables with obtained measurement results and write box with chronologically measured values of quality parameters: RMS voltage values, frequency values, THD - Total Harmonic Distortion factor values and individual high-order harmonics. Measurement is performed in cycles. Time interval for each PQ measurement cycle is set on 1 second.

Fig. 2. Software application for data acquisition, measurement of PQ parameters and presentation of measurement results - transmitter
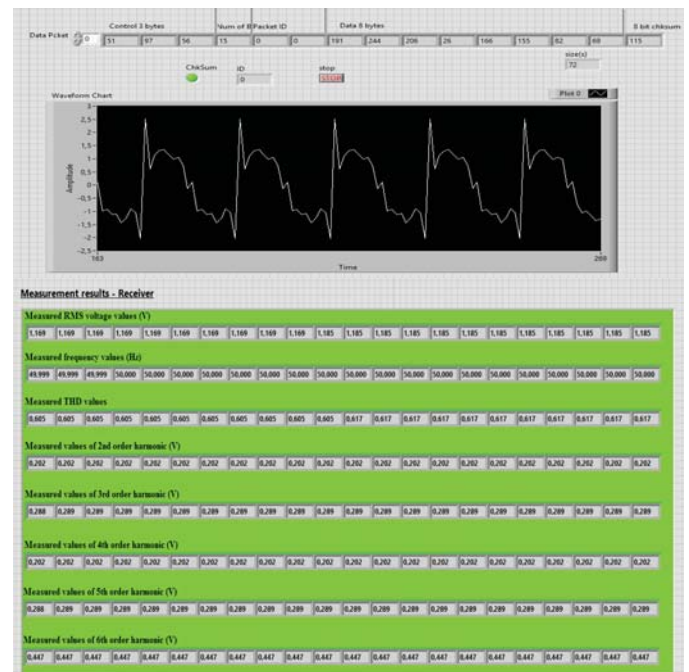
Fig. 3. Software application for presentation of received voltage signal and measurement results – receiver side

The WSN based network communication is applied for transferring of obtained measurement results to the receiver side of presented experimental system. LabVIEW software application implemented on the receiver side of experimental system, for presentation of received voltage waveform and received measurement results, is shown in previous Fig. 3.

III. CALCULATION OF STANDARD MEASUREMENT UNCERTAINTY COMPONENTS

Most important segment in metrological assessment of experimental system for the WSN based PQ measurement is calculation of measurement uncertainty components and final presentation of measurement uncertainty budget related to the PQ signal generator, applied as the source of reference voltage signals. Calculation of uncertainty is performed according to the standard document Guide to the Expression of Uncertainty in Measurement [7]. Voltage signal generated using the virtual instrument is physically reproduced by the data acquisition (DAQ) card containing the digital to analog (D/A) converter. Voltage signal generated on D/A converter output is usually standardized to reference voltage level, typically about ±10 V.

Procedure for uncertainty calculation is applied for each of measured signal quality parameters. In this paper is presented and analyzed method used for calculation of uncertainty in measurement of standard RMS voltage values. Measurement system applied for calculation of generator uncertainty is consisting of computer with DAQ card PCIe NI 6343 [11] for generation of reference voltage waveforms and 6 ½ digital multimeter Fluke 8846A [12] for measurement of DAQ card output RMS voltage values. Nominal RMS voltage value of the analog signals generated by DAQ card was set to 5V. Measurement of generator output voltage is performed for two values of signal frequency, 50 Hz and 1 kHz. In order to calculate Type A measurement uncertainty, 10 measurement cycles for each input signal frequency are performed. Interval between two successive measurement cycles is set to 5 min. RMS voltage values measured on the DAQ card output and calculated standard deviations, for two signal frequency values, 50 Hz and 1 kHz, are presented in Table I. Overall measurement uncertainty budget, including the calculations of standard, combined and expanded measurement uncertainty of applied PQ signal generator is presented in Table II.

Calculation of standard measurement uncertainty involves Type A uncertainty components (standard deviation of the mean values for measurement results) and Type B uncertainty components, including the multimeter uncertainty, multimeter resolution, DAQ card uncertainty and DAQ card resolution.

Standard deviation of the mean (Type A measurement uncertainty) is calculated according to the statistical methods applied on measurement results, using following equation:

$$u_A(V) = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} \left(V_{RMSi} - V_{AVER}\right)^2} \qquad (1)$$

Type B standard measurement uncertainties are calculated on the basis of data and measurement accuracies provided by specifications of instruments Fluke 8846A and DAQ card.

TABLE I
MEASUREMENT RESULTS AND CALCULATED STANDARD DEVIATIONS FOR MEASURED RMS OUTPUT VOLTAGE VALUES

No. of measurements	50 Hz V_{RMS} [V]	1 kHz V_{RMS} [V]
1	4.99308	4.99381
2	4.99298	4.99378
3	4.99318	4.99379
4	4.99311	4.99378
5	4.99305	4.99376
6	4.99318	4.99376
7	4.99300	4.99381
8	4.99309	4.99378
9	4.99297	4.99376
10	4.99315	4.99376
St. Deviation	0.00007	0.00002
St. Dev/$\sqrt{n}$	0.00002	0.00001
$V_{AVERAGE}$ (V)	4.99308	4.99378

According to instrument specifications [12], the multimeter absolute voltage uncertainty is $\Delta V_{MUL} = \pm (0.06\ \%$ of output value $+ 0.03\ \%$ of range value). Multimeter voltage resolution is $V_{MUL-RES} = 10\ \mu V$. Then, corresponding value of multimeter Type B uncertainty (u_{BMUL}) is calculated using equation:

$$u_{BMUL}{}^2(V) = u^2{}_{B1} + u^2{}_{B2} = \left(\frac{\Delta V_{MUL}}{2.58}\right)^2 + \left(\frac{1}{2}\frac{V_{MUL-RES}}{\sqrt{3}}\right)^2 \quad (2)$$

According to DAQ card PCIe NI 6343 specifications [11], DAQ card voltage absolute uncertainty value at full scale is $\Delta V_{DAQ} = 3.271\ \mu V$. Voltage resolution value of DAQ card is $V_{DAQ-RES} = 305\ \mu V$. Then, corresponding DAQ card Type B uncertainty (u_{BDAQ}) is calculated using the following equation:

$$u_{BDAQ}{}^2(V) = u^2{}_{B3} + u^2{}_{B4} = \left(\frac{\Delta V_{DAQ}}{2.58}\right)^2 + \left(\frac{1}{2}\frac{V_{DAQ-RES}}{\sqrt{3}}\right)^2 \quad (3)$$

Calculation of combined voltage measurement uncertainty value (u_{CDAQ}) is based on the previously calculated individual Type A and Type B measurement uncertainty values, using (1), (2) and (3), with following equation:

$$u_{CDAQ}(V) = \sqrt{u_A^2 + u_{BMUL}^2 + u_{BDAQ}^2} \qquad (4)$$

Finally, expanded measurement uncertainty value (u_{EXP}) is calculated for desired confidence probability level of 95% (value of coverage factor k is 1.96). Using the previously calculated value of the combined measurement uncertainty, expanded measurement uncertainty value is calculated as:

$$u_{EXP}(V) = k u_{CDAQ}(V) = 1.96 u_{CDAQ}(V) \qquad (5)$$

TABLE II
OVERALL MEASUREMENT UNCERTAINTY BUDGET - STANDARD, COMBINED AND EXPANDED VOLTAGE UNCERTAINTY CALCULATIONS

Uncertainty source	Type	Notation	Uncertainty for 50Hz [V]	Uncertainty for 1kHz [V]	Probability distribution	Coverage factor	Standard uncertainty for 50 Hz [V]	Standard uncertainty for 1 kHz [V]
Standard deviation	A	u_A	0.0000235	0.0000059	Normal	1	0.0000235	0.0000059
Multimeter uncertainty	B	u_{B1}	0.005995847	0.005996267	Normal	2.58	0.002323972	0.002324135
Multimeter resolution	B	u_{B2}	5.00E-06	5.00E-06	Uniform	1.732050808	2.88675E-06	2.88675E-06
DAQ card uncertainty	B	u_{B3}	3.27E-06	3.27E-06	Normal	2.58	1.26783E-06	1.26783E-06
DAQ card resolution	B	u_{B4}	0.000152588	0.000152588	Uniform	1.732050808	8.80967E-05	8.80967E-05
						Combined	0.00233	0.00233
						Expanded k = 1.96	0.00456	0.00456

The results presented in Table II suggest that expanded voltage measurement uncertainty values are ± 0.00456 V, for the both signal frequency values of 50 Hz and 1 kHz.

IV. CONCLUSION

Procedure for detailed metrological assessment of WSN based measurement of standard PQ parameters is described in the paper. This software oriented method includes calculation of standard, combined and expanded uncertainty components and presentation of overall measurement uncertainty budget. Evaluation of standard measurement uncertainty involves Type A uncertainty components (standard deviation of the measurement results) and Type B uncertainty components, including the specified instrument uncertainty and resolution values. Most important component in metrological assessment chain is PQ signal generator, including computer and data acquisition card, applied as the source of reference voltage waveforms. Metrological performances of WSN based PQ measurement system are evaluated for two signal frequency values, 50 Hz and 1 kHz. This solution can be implemented in distributed systems for electrical power distribution monitoring, with a number of remotely controlled measurement stations.

ACKNOWLEDGEMENT

REFERENCES

[1] J. G. de la Rosaa, A. A. Péreza, J. C. Salasa, J. M. Fernándeza, A. M. Muñoza, "A Novel Virtual Instrument for Power Quality Surveillance based in Higher-Order Statistics and Case-based Reasoning", Measurement, vol. 45, no. 7, pp. 1824–1835, 2012.

[2] D. Montenegro, M. E. Hernandez, G. A. Ramos, "A Realistic Generator of Power Quality Disturbances for Practicing in Courses of Electrical Engineering", Computer Applications in Engineering Education, vol. 23, no. 3, pp. 391-402, 2014.

[3] W. Zhu, W. Y. Ma, Y. Gui, H. F. Zhang, "Modelling and Simulation of PQ Disturbance based on Matlab", International Journal of Smart Grid and Clean Energy, vol. 2, no. 1, pp. 18–24, 2013.

[4] Y. Lim, H.M. Kim, S. Kang, "A Design of Wireless Sensor Networks for a Power Quality Monitoring System", Sensors, vol. 10, no. 11, pp. 9712-9725, 2010.

[5] I. F. Akyildiz, M. C. Vuran, Wireless Sensor Networks, John Wiley & Sons, August 2010.

[6] I. Mahgoub, M. Ilyas, Smart Dust: Sensor Network Applications, Architecture and Design, CRC Press, Taylor & Francis Group, 2006.

[7] ISO, Guide to the Expression of Uncertainty in Measurement, International Standard Organization, Geneva, 1993.

[8] M. Simić, Z. Kokolanski, D. Denić, V. Dimcev, D. Živanović, D. Taskovski, "Design and Evaluation of Computer-based Electrical Power Quality Signal Generator," Measurement, vol. 107, pp. 77-88, 2017.

[9] M. Simić, P. Planinšič, D. Denić, D. Gleich, D. Živanović, M. Malajner, "Software based Experimental System for Electrical Power Quality Measurement using the Wireless Sensor Network Modules", Facta Universitatis - Series Automatic Control and Robotics, vol. 16, no. 2, pp. 131-141, 2017.

[10] Power Quality Application Guide, Voltage Disturbances, PQ Standard EN 50160, Copper Development Association, 2004.

[11] PCIe NI 6343 Multifunction Data Acquisition Card DAQ - Specifications, National Instruments Corporation, Rev. 2014.

[12] Fluke 8845A/8846A Digital Multimeter – Users Manual, Fluke Corporation, Rev. 3, 3/11, July 2006.

Low-Cost Energy-Efficient Air Quality Monitoring System Using Sensor Network

Mare Srbinovska[1], Aleksandra Krkoleva Mateska[1], Vesna Andova[1], Maja Celeska[1] and Tomislav Kartalov[1]

Abstract – **The air quality in urban areas is a major concern in modern cities due to significant impacts of air pollution on human health and global environment. Implementation of green walls is one of the methods for decreasing air pollution in urban areas as plants absorb the particulate matter through their leaves and growing medium. The objective of the paper is two-fold: 1) to present an implementation of a low-cost and energy-efficient air quality monitoring system using sensor network that can be easily deployed in highly polluted areas; and 2) to examine the influence of experimental green wall setup to particulate matter concentrations in the air in an urban area in Skopje. Furthermore, the paper presents the preliminary results of the ongoing experiment developed to evaluate the impact of green walls in reduction of air polluting particles.**

Keywords – **Sensor network, Air quality improvement, vertical green walls.**

I. INTRODUCTION

The air quality in many countries has been adversely affected over the past years due to continuous economic and industrial development. The increasing level of air pollution from vehicles exhausts, chemical discharge from industries, dust and particulate matter from various sources and toxic gas leakage are affecting the citizens' health and ecosystems. With the aim to respond to the existing environmental issues, new methods for improving air quality in both indoor and outdoor environments are being developed. According to the World Health Organization (WHO), more than one million people worldwide die from the consequences of polluted air every year [1]. United States Environmental Protection Agency [2] defines the air quality by measuring specific gases that have the most significant effect on human health, including: ground – level ozone (O_3), carbon monoxide (CO), nitrogen oxide (NO), particulate matter (PM10). Other pollutants also include carbon dioxide (CO_2), various nitrogen and sulfur oxides, hydrogen sulfide (H_2S), volatile organic compounds, smog and other particulate matters. The measurement of these gases is related to the specifics of each gas as well as of the equipment that is used.

The problem of air quality is increasing around the world. The growing economies of Asia have had this problem during the past decade [3, 4, 5, 6] and some European countries have been tackling this problem as well [7].The air quality in Skopje has decreased significantly over the past years. Data from several measurement stations throughout the city have shown high levels of particulate matter [8], especially during the winter months. These levels a few times higher than normally acceptable and are cause of serious concerns of citizens and authorities. A few governmental agencies and non-governmental bodies examine the reasons, but so far, there is no definite list of responsible entities. The general reasons lay into fast urbanization that "attacks" every free inch of space. Furthermore, the city has been developing for use of cars rather than other means of transportation, although it has a good disposition for using bikes. Other problems are the unauthorized landfills, use of various fossil fuels for heating. Although steps are being taken to solve this problem, public pressure to provide solutions is growing. Therefore, the existing problem of air pollution and the possible ways to solve it is the main motive to start our work on the air quality measurement project.

Green walls are vegetated vertical surfaces where plants are attached to the surface through various mechanisms. Green walls allow for high density and high diversity vegetation on vertical areas [9]. The plants filter nitrogen dioxide (NO_2) and microscopically small particles (fine dust) from the air. Both represent a very serious health problem in the cities of both industrialized nations and developing countries. For air pollution improvement using various types of green infrastructure, the majority of studies have focused on pollutants such as the PM10[10], PM2.5[11] that have implications for the adverse health effects.

II. LITERATURE REVIEW

Research shows significant influence of green walls on various environmental problems. Authors in [12] investigate the potential of green walls and facades for saving energy. Furthermore, the research presented in [13] describes the potential for energy savings in buildings using green walls and double-skin green facades. Classification and technical aspects of the green roofs are explained in [14]. The study showed that green roofs could reduce annual energy consumption for interior heating and cooling. Among the reasons for this is that they are on average 15,5 °C cooler than black roofs in summer [15]. Buildings in northern climates, with high temperature extremes and shorter growing seasons, distinctly show the energy advantages of green roofs.

The aim of this paper is to present a low - cost energy-efficient system for air quality monitoring in order to analyze the influence of the green wall on air quality improvement. The

[1]Mare Srbinovska, Aleksandra Krkoleva Mateska, Vesna Andova, Maja Celeska and Tomislav Kartalov are with the Faculty of Electrical Engineering and Information Technologies at Ss Cyril and Methodius University in Skopje, Rugjer Boskovic 18, 1000 Skopje, Macedonia, E-mail: mares@feit.ukim.edu.mk, krkoleva@feit.ukim.edu.mk, vesnaa@feit.ukim.edu.mk,celeska@feit.ukim.edu.mk, kartalov@feit.ukim.edu.mk

paper presents results for evaluation of the impact of green walls for reduction of air polluting particles, more specifically the influence of the air flow through the green wall and the capability of the green wall to absorb particulate matter. Therefore, the investigated system uses sensor network for measurement of particulate matters (PM10, PM2.5), CO and NO_2.

III. DEVELOPMENT PHASES

The first phase in the implementation process was to select a proper location for installation of the system. The experimental location is near the building of FEEIT and a parking lot with frequent movement of people and vehicles. Before the installation of the sensor nodes, continuous measurements for air quality monitoring were done using reference measurement instrument. The second step was definition of system architecture and hardware specification. The process was followed by equipment procurement. The third step was related to the green surface construction and installation of the WSN. The aim was to construct the green wall and to install the sensor nodes for measurement that will provide information for the important parameters related to air quality (PM2.5, PM10, CO, NO_2).

A. Hardware setup

The sensor modules deployed in the experimental setup consist of two major hardware components – sensors and controller. The sensor nodes are the primary data collecting elements in the network. They consist of four sensors measuring the following parameters: PM2.5, PM10, CO and NO_2, which feed data to the Wi-Fi module for transmission.

TABLE I
MAIN CHARACTERISTICS OF THE SDS011 SENSOR

Measurement parameters	PM2.5, PM10
Range	0.0-999.9 μg/m3
Power supply voltage	5V
Maximum working current	220 mA
Sleep current	2 mA
Operating temperature	-20 °C-50 °C
Minimum resolution of particle	<0,3μm

SDS011 [16] is the PM2.5 and PM10 sensing unit. It uses the principle of laser scattering and can get the particle concentration between 0.3 to 10μm in the air. This sensor with the digital output and built-in fan is stable and reliable. Some of the characteristics of this sensor are: accurate and reliable, it has quick response with response time less than 10 seconds when the scene changes, easy integration with universal asynchronous receiver-transmitter (UART) output, and high resolution of 0.3μg/m^3. The main characteristics of the sensing elements are summarized in Table I. MiCS-4514 [17] is combined CO and NO_2 sensor. Detection of the polluted gases is achieved by measuring the sensing resistance of both sensors: RED (reduced) sensor resistance decreases in the presence of

CO and hydrocarbons, OX (oxygen) sensor resistance increases in the presence of NO_2.

The main characteristics of the sensing elements are summarized in Table II.

TABLE II
MAIN CHARACTERISTICS OF THE MiCS-4514 SENSOR

Measurement parameters	CO, NO_2
Maximum heater power dissipation	88 mW (RED sensor) /50 mW (OX sensor)
Voltage supply	4.9 V - 5.1 V
Relative humidity range	5 % - 95 %
Ambient operating temperature	-30 °C - 85 °C
CO detection range	1 ppm - 1000 ppm
Sensing resistance in air	100 kΩ - 1500 kΩ
NO2 detection range	0.05 ppm - 5 ppm
Sensing resistance in air	0.8 kΩ - 20 kΩ

ESP32-WROOM-32D [18] is a powerful module that covers wide range of applications, from low – power sensor networks, to the most demanding tasks like voice encoding, music streaming. At the core of the module is the ESP32-D0WD chip.

The controller has integrated on-board antenna and Wi-Fi module, that allows large physical range and direct connection to the internet through Wi-Fi router. The current of the ESP32 chip in sleep mode is less than 5 μA, which makes this module suitable for battery powered and wearable electronics applications. The main characteristics of the controller are presented in Table III.

TABLE III
MAIN CHARACTERISTICS OF THE EPS CONTROLLER

Controller	EP32
Power supply	2.7 V - 3.6 V
Operating temperature range	-40 °C - 85 °C
Operating current	80 mA
On-chip sensor	Hall sensor
On-board clock	40MHz crystal
Module interface	SD Card, UART, SPI, I2C, Motor PWM
Wi-Fi frequency range	2.4 GHz - 2.5GHz

B. Experimental setup

For the experimental setup we used old materials for the hedera helix construction in order to build a simple, cheap and easy to replicate system and to contribute to waste reduction. The green wall support construction is built of used materials (wooden boards and metal support structures from old furniture). Instead of using new plastic pots, old 6 L plastic water bottles are used. The implemented construction is shown on Figure 1. Wires that connect the wooden boards are used to support the growing hedera helix plants. It is chosen to be near the Faculty building, which faces a Faculty parking lot,

between two smaller classroom buildings. Students and staff frequent this location quite often.

The sensor network consists of 4 sensor nodes. As shown in Fig. 2, all sensor nodes are cased in plastic boxes. One of the sensor nodes is placed near the green wall construction, while the other three are placed near the parking zone and the classroom buildings. The fourth sensor is positioned to face the parking side, where a lot of cars and vehicles commute on daily bases. The measurement data from all the sensor nodes are uploaded to an open platform [19] and are available for online
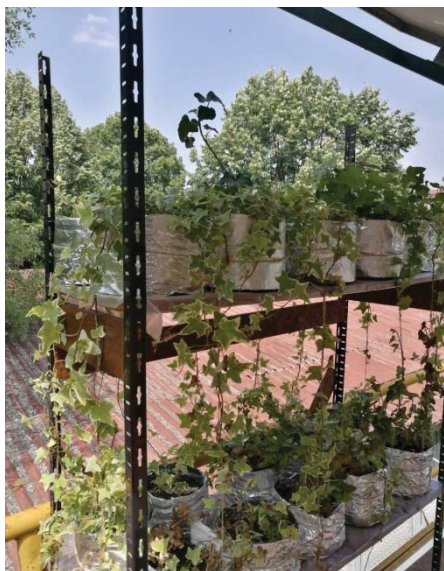
Fig. 1. Hedera helix green wall structure

Fig. 2. Sensor node box

monitoring and extraction for further analyses.

IV. ANALYSIS OF MEASUREMENT DATA

Table IV shows the maximum allowed values for various air pollutants [20]. These values are used as comparison reference for the measurements at the experimental setup. The table also provides a reference for the Air Quality Index (AQI), which may be calculated using the methodology described in [7]. As shown in Table IV, AQI values are divided into ranges, and each range is assigned with an adequate descriptor.

The experimental setup is designed in a manner that allows measurements to be taken near the green wall (sensor node 2) and in relative distance of few meters from the setup (nodes 1,

3 and 4). Node 4 faces the parking lot and is the most distant node from the experimental green wall. The disposition of the sensor nodes allows assessment of the influence of the green wall on the quality of air on the micro-location where the experimental setup is positioned.

TABLE IV
OVERVIEW OF THE MAXIMUM ALLOWED VALUES OF GASES AND PARTICULATE MATTER [20]

PM10 ($\mu g/m^3$)	CO ($\mu g/m^3$)	SO^2 ($\mu g/m^3$)	NO_2 ($\mu g/m^3$)	AQI	
0-50	0.0-10.0	0.0-350	0.0-200	0-50	Good
51-150	10.01-20.0	351-1500	201-400	51-100	Moderate
151-250	20.01-30.0	1501-2500	401-800	101-150	Unhealthy for sensitive group
251-350	30.01-45.0	2501-3500	801-1200	151-200	Unhealthy
341-450	45.01-70.0	3501-6500	1201-2400	201-300	Very unhealthy
420-500	70.1-100.0	6501-8500	2401-3200	301-400	Dangerous
501-600	100.01-120	8501-10000	3201-4000	401-500	Very Dangerous

V. RESULTS

The new series of measurements started in May 2018. The air pollution in this period of the year is not as high as during the winter months. After the first equipment tests, the setup was used to provide continuous measurements of all four parameters. However, this paper presents the measurements of PM2.5 and PM10, as these parameters were critical for the air quality in Skopje, which is already described in [3].

The values shown on Fig. 3 and Fig. 4 represent average concentrations of PM2.5 and PM10 per hour respectively, but the measurements are taken with rate of one measurement per minute. The sensor nodes are numbered as described above, node 2 being the closest and node 4 being the furthest positioned node from the green wall. The PM2.5 concentration is well below 25 µg/m3, being the relative concentration allowed in the air. It is clearly visible that the concentrations increase during the day, when there is more movement and are usually lower during the late night and early morning hours.

Similar conclusions can be drawn from the graph presented on Fig. 4. Namely, the concentrations of PM10 are lower than the reference in Table IV and the distribution of values follows similar pattern.

The graphs depicted in Fig. 3 and Fig. 4, show a typical cases during the winter months in Skopje. The graphs depict measurements of PM2.5 and PM10 concentrations respectively and are taken in a period of 24 hours, starting at 00:00 on 10.11.2018 until 00:00 on 11.11.2018.

First and foremost, it can be seen that the particular matter concentration values from nodes 2 and 3 are constantly
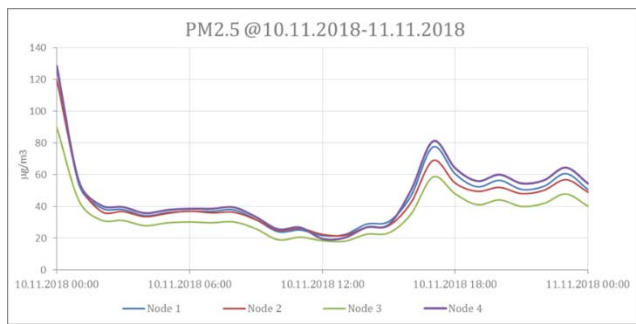
Fig. 3. PM2.5 concentration measurement for the 24-hour period from 10.11.2018 00:00 – 11.11.2018 00:00.
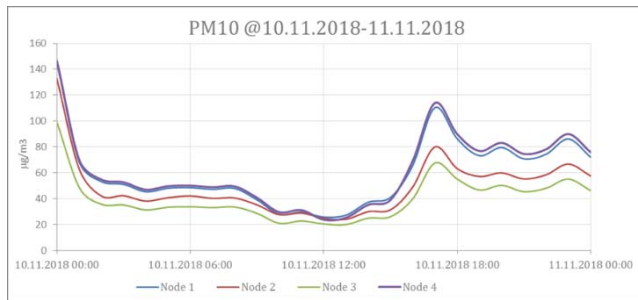
Fig. 4. PM10 concentration measurement for the 24-hour period from 10.11.2018 00:00 – 11.11.2018 00:00

relatively lower than those from node 1 and 4 which are furthest away from the green wall setup of the three nodes. This is observed in Fig. 3 and Fig. 4 and confirms the preliminary findings presented in [3] that the green wall contributes to lowering the levels of particulate matter in the surrounding air.

VI. Conclusion

The goal of this paper is to describe an experimental green wall setup and to present preliminary measurement results from an ongoing experiment designed with the aim to evaluate the influence of green walls on air quality on micro locations.

The presented experimental green wall setup is a low cost design, using over the counter and recycled materials and equipment. The collected data is available for online monitoring as well as off-line analyses.

The measurement results from a longer period of about 6 months indicate that particular matter concentrations tend to be constantly lower in the area nearer to the green wall structure.

The experimental green wall project enables continuous measurements of gaseous pollutants and PM concentrations. A continuous measurement campaign, using the WSN measurements, shall provide basis for analysis of the influence of the green wall in improving air quality, especially in terms of mitigating PM10 concentrations. Further considerations include analyses on how the size of the green wall structure impacts the scale of reduction of the particular matter concentration and the size of the area it affects positively.

References

[1] World Health Organization (WHO), Monitoring Ambient Air Quality for Health Impact Assesment, WHO Regional Publications, European Series, No.85, 1999.

[2] The United States Environmental Protection Agency (US EPA), http://www.epa.gov.

[3] M. Srbinovska, A. Krkoleva, V. Andova, M. Celeska, Wireless Sensor Networks Implemented in Vertical Green Walls for Air Quality Improvement, 12th Conference on Sustainable Development of Energy, Water and Environment Systems, 4-8 October, Dubrovnik, paper ID: SDEWES2017.0606.

[4] Bolton, C., Rahman, M. A., Armson, D., Ennos, A. R., Effectiveness of an ivy covering at insulating a building against the cold in Mancester, U.K, Journal of Building and Environment. Vol. 80, pp.32-35, 2014.

[5] Cuce, E., Thermal regulation impact of green walls: An experimental and numerical investigation, Journal of Applied Energy, Vol. 194, pp 247-254, 2016.

[6] Maleki, H., Sorooshian, A., Goudarzi, G., Nikfal, A., Baneshi, M.M., 2016. Temporal profile of PM10 and associated health effects in one of the most polluted cities of the world (Ahvaz, Iran) between 2009 and 2014. Aeolian Res. 22, pp. 135-140.

[7] Heal, M.R., Kumar, P., Harrison, R.M., 2012. Particles, air quality, policy and health. Chem. Soc. Rev. 41, pp. 6606-6630.

[8] Beevers, S.D., Westmoreland, E., de Jong, M.C., Williams, M.L., Carslaw, D.C., Trends in NOx and NO2 emissions from road traffic in Great Britain. Journal of Atmospheric. Environment, Vol. 54, pp. 107-116, 2012

[9] Bigazzi, A.Y., Figliozzi, M.A., 2015. Roadway determinants of bicyclist exposure to volatile organic compounds and carbon monoxide. Transp. Res. Part D, Journal of Transportation Environment, Vol. 41, pp.13-23.

[10] Coma, J., Pérez, G., de Gracia, A., Burés, S., Cabeza, L. F., Vertical greenery systems for energy savings in buildings: A comparative study between green walls and green facades, Journal of Building and Environment, Volume 111, pp 228-237, 2017.

[11] Alberto, A. , Ramos, N, Almeida, R., Parametric study of double-skin facades performance in mild climate countries , Journal of Building Engineering, vol. 12, pp 87-98, 2017.

[12] S. R. Gaffin, S.R,C. Rosenzweig, C.,J. Eichenbaum-Pikser, J., Khanvilvardi, R., Susca, T. et al, A Temperature and Seasonal Energy Analysis of Green, White and Black Roofs, Columbia University, Center for Climate Systems Research., New York, NY., 2010.

[13] J. Foster, A. Lowe, S. Winkelman, The Value of Green Infrastructure for Urban Climate Adaptation, Center for Clean Air Policy, February 2011. Available online at: http://www.ccap.org/docs/resources/989/Green_Infrastructure_FINAL.pdf.

[14] Fahmy M., Sharples S., Urban form, thermal comfort and building CO2 emissions - a numerical analysis in Cairo. Build Serv Eng Res Technol, vol.32(1) pp. 73-84, 2011.

[15] De Vito S, Massera E, Burrasca, G, Di Girolamo A, Miglietta M, Di Francia G, Della Sala D, "TinyNose: Developing a wireless sensor platform for distributed air quality monitoring applications" Sensors, IEEE on 26-29 Oct. 2008 , pp. 701 – 704.

[16] Laser PM2.5 Sensor specification, SDS011, Version V1.3 datasheet, Nova Fitness Co.,Ltd, 2015

[17] A1A-MiCS-4514 Version 2 datasheet, 2008

[18] ESP32-WROOM-32D & ESP32-WROOM-32U Datasheet V1.4, 2018

[19] Internet of Things open platform, https://thingspeak.com

[20] Air Quality Index (AQI), http://en.wikipedia.org/wiki/Air_Quality_Index

Practical Implementation of Voltage Unbalance Measurement

Lazar Sladojević[1], Lidija Korunović[2], Miodrag Stojanović[3] and Vladeta Milenković[4]

Abstract – **This paper describes practical implementation of the algorithm for evaluation of supply voltage unbalance according to IEC 61000-4-30 Standard for Electromagnetic compatibility and Power quality measurement methods. The algorithm was written in Python programming language and evaluated on real voltage signals. The results are presented and some issues are highlighted for future research.**

Keywords – **Voltage unbalance, IEC 61000-4-30, Power quality, Python**

I. INTRODUCTION

Power quality (PQ) is becoming increasingly important problem in electrical power systems. The deviation of supplying voltage from the ideal sine wave in many areas is now evident more than ever. This is most obvious in the areas with extensive usage of electronic devices. These devices generate higher order harmonics which impact negatively the supplying power grid, thus generating additional power and energy losses [1].

There is a total number of twelve power quality parameters defined in the IEC 61000 Standard for Electromagnetic compatibility, part 4-30 which considers Power quality measurement methods that should be addressed, measured and evaluated [2]. The standard defines measurement methods as well as measurement uncertainty and measurement range for each individual parameter, except for measurement of flickers, harmonics and interharmonics, which are described in separate standards. Two classes of measurement equipment are recognized, class A and class S and the main difference between them is the accuracy – class A is more accurate than class S.

This paper addresses the problem of real-time measurement of supplying voltage unbalance. Voltage unbalance can occur for different reasons. For example, it can be caused by unbalanced load or by the supplying grid itself [3]. As a result, huge losses in electrical drives can be observed [4]. Unlike some other power quality parameters such as voltage dips and swells, small unbalance usually doesn't produce immediate negative effects, but rather has long term consequences [5].

[1]Lazar Sladojević is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: lazar.sladojevic@elfak.rs

[2]Lidija Korunović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: lidija.korunovic@elfak.ni.ac.rs

[3]Miodrag Stojanović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: miodrag.stojanovic@elfak.ni.ac.rs

[4]Vladeta Milenković is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: vladeta.milenkovic@elfak.ni.ac.rs

For the measurement of unbalance, a data acquisition platform has been developed and used in conjunction with fast ARM processor for the necessary calculations. Unbalance measurement algorithm was entirely written in Python programming language. The goal is to achieve real time evaluation of voltage unbalance and to present the results of practical implementation of the methods proposed in standard IEC 61000-4-3.

II. PARAMETER DESCRIPTION AND CALCULATION METHODS

Supply voltage unbalance is a state of voltage waveforms in a three-phase system which deviates from the ideal model. The ideal model is characterized by three pure sine wave voltages with the rated magnitude and frequency whose phasors are angularly displaced from each other by 120 degrees. Therefore, voltage unbalance can either be influenced by different magnitudes of the voltages or by phase shifts that deviate from those of ideal model.

There are different methods for calculation of voltage unbalance [6]. According to IEC 1000-3-x series standards, the unbalance is calculated using the method of symmetrical components. This method is well known in electrical engineering and it is based on fact that every unbalanced (asymmetrical) three-phase system can be mathematically represented by the superposition of three balanced three-phase systems. These three systems are called the direct sequence system, the inverse sequence system and the zero sequence system, and have the same frequency as the original system. Assuming that phase A is the referent voltage phase, phase B lags by 120 degrees and phase C by 240 degrees relative to phase A in direct sequence system, while in the inverse system phases B and C lead by the same angles. In zero sequence system, there is no phase shift among voltage waveforms. Each phase phasor is a vector sum of corresponding direct, inverse and zero component.

The IEC standards define that the voltage unbalance is evaluated by the so-called Voltage Unbalance Factor (VUF). This factor represents ratio of the magnitudes of inverse sequence and direct sequence fundamental components of voltage, expressed in percent,

$$VUF = u_2 = \frac{U^i}{U^d} \cdot 100 \qquad (1)$$

In Eq. 1, U^i is the magnitude of inverse sequence component of fundamental voltage, U^d is the magnitude of direct sequence component of fundamental voltage, and u_2 is the alias for *VUF* which is used in IEC 61000-4-30 Standard. In standards defining the maximum allowed voltage unbalance in

transmission and distribution grid, these limits are given as maximum values of *VUF*.

On the other hand, ANSI and IEEE standards present the methods for voltage unbalance calculation that do not take into account the irregular phase shifts. They observe only the changes in magnitudes. Thus, the IEC method defined in IEC 61000-4-30 Standard is used in this paper for the calculation of voltage unbalance.

III. Algorithm Description

IEC 61000-4-30 Standard defines that the sliding 10 cycles window is used for unbalance calculation in 50 Hz systems. This means that unbalance *VUF* factor, Eq. 1, is calculated approximately every 200 ms. There are no requirements for the synchronization of this window with zero cross points of any of the phases, like for example in the case of voltage dips, swells and interruption detection. The accuracy class A of the standard states that measurement uncertainty for unbalance *VUF* factor value is equal to 0.15 %.

The detection method described above implicitly requires two things: fundamental magnitude calculation and fundamental cycle duration calculation. The fundamental magnitude calculation refers to calculation of magnitude of the fundamental component of voltage wave, i.e. component on the fundamental frequency in spectrum. This is necessary because voltage can also contain higher order harmonics which need to be minimized before the calculations. There are two ways to minimize higher order harmonics, as recommended by standard: the filtration using low pass filter or the Discreet Fourier Transform (DFT). The modification of the latter kind is applied in this paper. Namely, the corresponding algorithm uses the Python's built-in Fast Fourier Transform (FFT), which is a faster, optimized version of the DFT [7]. As required by the standard, the FFT is executed over each 10 cycles window.

Fundamental cycle duration is expressed as a number of samples N in one full cycle. This number can be calculated from the frequency of signal, f, and the sampling frequency, f_s:

$$N = round\left(\frac{f_s}{f}\right) \quad (2)$$

There are number of frequency estimation methods, some of those are explained in detail in [8–11], but for the application presented in this paper, a simple zero-crossing detection method is used. This method uses the linear regression over the samples around the first and the last zero cross of a signal. Samples involved in the linear regression are the ones with values between the fixed upper and lower threshold. This way, a variable number of samples is used for each linear regression, because the signal always contains some noise. This noise can cause multiple zero crosses and non-equal number of positive and negative samples. Here, linear regression behaves as some type of a filter, because it constructs the straight line which minimizes the sum of squared distances from all samples involved in regression,

$$\min \sum_{i=1}^{M} \left(U_i - \hat{U}_i\right)^2 = \min \sum_{i=1}^{M} \left(U_i - \left(a \cdot t_i + b\right)\right)^2 \quad (3)$$

In Eq. 3, $\hat{U}_i$ is the projection of sample U_i on the regression line, M is number of samples involved in linear regression, a is the line slope coefficient, b is line intercept coefficient, and t_i is the time point at which the sample was taken.

The approach is fairly simple and computationally effective, but can give errors when applied to highly distorted signals. In these cases, IEC 61000-4-30 Standard recommends filtration of a signal before the estimation of its frequency. However, primary goal of this paper is developing and evaluation of practical algorithm for voltage unbalance measurement, and for simplicity, that will be done on fairly clean, sine waveform signals. This assumption is not always fulfilled in practice, but processing of highly distorted signal requires better frequency measurement, and will be the subject of authors' future work. Nevertheless, the algorithm is not influenced significantly by harmonic distortion, since the calculations are performed on fundamental components of signals which are extracted from FFT.

IV. Experimental Results

For experimental purposes, a data acquisition and processing platform has been developed. The platform contains 16-bit Analog to Digital (A/D) converter with the 25.6 kHz sampling frequency for data acquisition and ARM Cortex-A8 AM335X microprocessor for digital processing of the gathered data. Platform is capable of sampling and processing of all three voltage channels simultaneously.

The algorithm itself was entirely written in Python programming language. Python was chosen because it is simple and open source programming language, yet very powerful when it comes to scientific computing. Python runs on Linux operating system which is embedded in the platform. The algorithm was developed to comply with class A measurement instrumentation. Sets of sampled data from each voltage channel are received in the packages of 25600 samples, which means they are sampled for one second before they are processed. This implies that the algorithm has to process the whole previous set of samples in time period shorter than one second, i.e. before the new samples are received. Therefore, the algorithm has to be highly optimized.

For test voltage generation, Omicron CMC 356 test set was used [12]. This test set is primarily designed for relay protection testing, but it can also serve well other purposes, such as the one described in this paper.

Many test cases have been created and examined, but only a part of the results is shown due to lack of space. Five characteristic cases are depicted in Figs. 1 to 5. The overview of the results is given in Table I, while the processing times for one-second data buffers are presented in Table II. Besides voltage unbalance measurements, signal frequency is estimated and FFT calculations are conducted each second. FFT is performed over 10 cycle time interval and fundamental phasor is extracted. All three phases' phasors are then used in Eq. 4 to derive symmetrical components for that time period.

$$\begin{bmatrix} U^d \\ U^i \\ U^0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & a & a^2 \\ 1 & a^2 & a \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} U_{AN} \\ U_{BN} \\ U_{CN} \end{bmatrix} \qquad (4)$$

In Eq. 4, U^d, U^i and U^0 are the phase A direct, inverse and zero sequence components, respectively, U_{AN}, U_{BN}, U_{CN}, are phase-to-neutral voltage magnitudes for phase A, B and C, respectively, and $a = e^{j \cdot 2\pi/3}$. Besides *VUF*, a similar zero-sequence factor is also requested by standard IEC 61000-4-30:

$$u_0 = \frac{U^0}{U^d} \cdot 100 \qquad (5)$$

TABLE I
PARAMETERS OF TEST CASES

Test case index	1	2	3	4	5
$U_{A\text{-}fund}$ [V]	230 ∠ 0°	230 ∠ 0	230 ∠ 0	230 ∠ 0	230 ∠ 0
$U_{B\text{-}fund}$ [V]	230 ∠ -120°	92 ∠ -120°	230 ∠ -120°	230 ∠ -120°	92 ∠ -90°
$U_{C\text{-}fund}$ [V]	230 ∠ 120°	230 ∠ 120°	92 ∠ 120°	230 ∠ 90°	92 ∠ 90°
U^d_{real} [V]	230 ∠ 0°	184 ∠ 0°	184 ∠ 0°	223 ∠ -9.9°	129.766 ∠ 0°
U^i_{real} [V]	0 ∠ 0°	46 ∠ -60°	46 ∠ 60°	39.6842 ∠ 135°	23.552 ∠ 0°
U^0_{real} [V]	0 ∠ 0°	46 ∠ 60°	46 ∠ -60°	39.6842 ∠ 15°	76.682 ∠ 0°
$U^d_{estimated}$ [V]	229.985 ∠ 0.01°	183.976 ∠ -0.01°	183.9899 ∠ -0.01°	223.0314 ∠ -9.88°	129.776 ∠ -0.01°
$U^i_{estimated}$ [V]	0.057 ∠ -111.3°	45.949 ∠ -60.06°	46.0208 ∠ 60.06°	39.6804 ∠ 135.1°	23.5613 ∠ 0.04°
$U^0_{estimated}$ [V]	0.019 ∠ 35.83°	46.0004 ∠ 60.02°	45.9845 ∠ -60.05°	39.6592 ∠ 15.01°	76.668 ∠ -0.04°
$u_{2\text{-}real}$ [%]	0	25	25	17.7913	18.1496
$u_{0\text{-}real}$ [%]	0	25	25	17.7913	59.0925
$u_{2\text{-}estimated}$ [%]	0.025	24.9756	25.0127	17.7914	18.1554
$u_{0\text{-}estimated}$ [%]	0.0083	25.0034	24.9930	17.7819	59.0774
u_2 estimation error [%]	0.025	0.0244	0.0127	0.0001	0.0058
u_0 estimation error [%]	0.0083	0.0035	0.0070	0.0094	0.0152

Fig. 2. Test case 2 a) phasors and b) symmetrical components
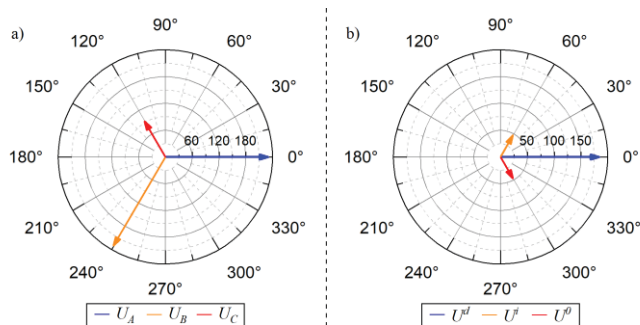
Fig. 3. Test case 3 a) phasors and b) symmetrical components
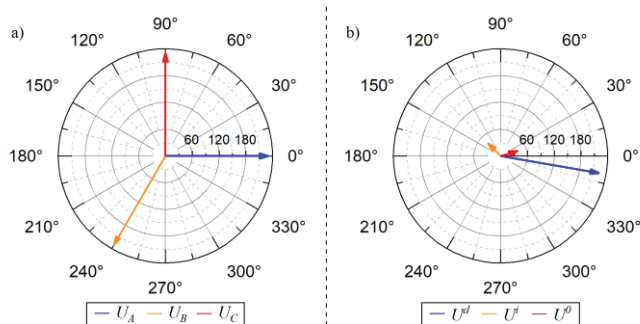
Fig. 4. Test case 4 a) phasors and b) symmetrical components
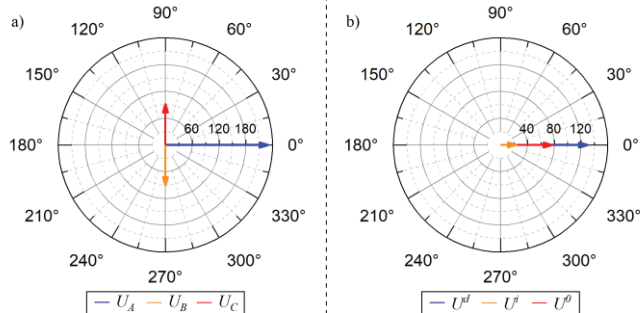
Fig. 5. Test case 5 a) phasors and b) symmetrical components

Fig. 1. Test case 1 a) phasors and b) symmetrical components

Each figure shows all three phases' phasors on the left side and measured symmetrical components on the right side. The values indexed as *real* in Table I are directly read from Omicron test software and these are the correct parameters of actually produced signals. Values for $u_{2\text{-}estimated}$ and $u_{0\text{-}estimated}$ are 150 cycles aggregated values whose calculation is mandatory for class A equipment. Note that presented cases are the most

extreme ones (some of them are also considered voltage dips), but it is known that the algorithms are most efficiently tested in extreme conditions.

Last two rows in Table I are the most important. They present absolute values of measurement errors for u_2 (*VUF*) and u_0. It is obvious that none of these values is greater than 0.15 %, which is class A uncertainty limit as described in applied standard. Similar values are obtained from other test cases, which are not presented here. Further, Table II shows that the average calculation time is shorter than one second, which is time available for processing of one buffer of data. This proves that the algorithm is well designed and capable of real-time class A voltage unbalance measurement.

TABLE II
PROCESSING TIMES FOR DIFFERENT TEST CASES

Test case index	Average processing time for one buffer of data [s]
1	0.8211
2	0.8205
3	0.8194
4	0.8254
5	0.8257
Total average time	**0.8224**

V. CONCLUSIONS

In this paper the software algorithm for measurement of supply voltage unbalance is presented. The goals of algorithm are to achieve real time measurement performance of this abnormal state in power systems, and to evaluate practically obtained results applying the method described in standard IEC 61000-4-30. For practical purposes, new data acquisition platform is developed and tested. Five test cases are examined and experimental results are presented. It can be concluded that the proposed algorithm is capable of real-time voltage unbalance measurements with very high precision which complies with class A instrumentation. In this paper clean sinusoidal voltages are examined, whereas further results will be obtained taking into account the presence of higher order harmonics. The research in this field and the improvement of developed algorithm will be the subject of authors' future papers.

REFERENCES

[1] R. Dugan, M. McGranaghan, S. Santoso, H. Beaty,, *Electrical Power Systems Quality*, Second edition, McGraw-Hill, 2004.

[2] IEC 61000 International Standard – Electromagnetic Compatibility, Part 4-30: Testing and Measurement Techniques, - Power Quality Measurement Methods, edition 3.0, 2015.

[3] L.Pierrat, R.E. Morrison, "Probabilistic Modeling of Voltage Asymmetry", IEEE Trans. on Power Delivery, Vol. 10. No. 3, pp 1614-1620, Jul. 1995.

[4] M. Kostic, A. Nikolic, "Negative Consequence of Motor Voltage Asymmetry and Its Influence to The Unefficient Energy Usage", WSEAS Trans. on Circuits and Systems, Issue 8, Volume 9, Aug. 2010.

[5] P. V. Santos Valois, C. M. Vieira Tahan, N. Kagan, H. Arango, "Voltage Unbalance in Low Voltage Distribution Networks", 16th International Conference and Exhibition on Electricity Distribution, 2001. Part 1: Contributions. CIRED. (IEE Conf. Publ No. 482), Jun. 2001.

[6] M. H. Albadi, A. S. Al Hinai, A. H. Al-Badi, M. S. Al Riyami, S. M. Al Hinai, and R.S Al Abri, "Unbalance in Power Systems: Case Study", 2015 IEEE International Conference on Industrial Technology (ICIT), Seville, Spain, 2015. pp. 1407-1411. DOI: 10.1109/ICIT.2015.7125294

[7] J. Cooley, J. Tukey, 1965, "An Algorithm for The Machine Calculation of Complex Fourier Series," Math. Comput. 19: 297-301, Apr. 1965.

[8] M. S. Sachdev and M. M. Giray, "A Least Error Squares Technique for Determining Power System Frequency," IEEE Trans. Power App. Syst., vol. PAS-104, no. 2, pp. 437–444, Feb. 1985.

[9] H. Karimi, M. Karimi-Ghartemani, and M. R. Iravani, "Estimation of Frequency and Its Rate of Change for Applications in Power Systems," IEEE Trans. Power Del., vol. 19, no. 2, pp. 472–480, Apr. 2004.

[10] M. Djuric, Z. Djurisic, "Frequency Measurement of Distorted Signals Using Fourier And Zero Crossing Techniques", Electric Power Systems Research Volume 78, Issue 8, pp. 1407-1415, Aug. 2008.

[11] E. Lavopa, P. Zanchetta, M. Sumner, and F. Cupertino, "Real-Time Estimation of Fundamental Frequency and Harmonics for Active Shunt Power Filters in Aircraft Electrical Systems", IEEE Trans. on Industrial Electronics, Vol. 56, No. 8, pp. 2875-2884, Aug. 2009.

[12] CMC 356 Reference Manual, Article Number VESD2003 - Version CMC356.AE.4

Sensor data fusion for determine object position

Emil Iontchev[1], Rosen Miletiev[2], Petar Kapanakov[3] and Lachezar Hristov[4]

Abstract: **The paper represents the algorithm which combines the sensor data to determine the object position in the space by Extended Kalman Filter (EKF). Two different filters are proposed – nonlinear Kalman filter and linear one with quaternions. Also the algorithm for rejection of the external magnetic disturbances is chosen to correct the magnetometer data. Another low-pass filter is applied to the accelerometer data. The proposed algorithms are tested on stationary and on moving platforms to calculate the rotation angles of the kid car seat.**

Keywords – **accelerometer, gyroscope, magnetometer, Kalman filter**

I. INTRODUCTION

Object position determination is the major task of the human life and it is solved by different methods. Nowadays, the outdoor position may be obtained easily by using the global navigation positioning systems (GNSS) [1]. The indoor navigation task also may be solved using inertial navigation systems (INS), beacons, etc. [2], but sometimes there is a necessity to combine the navigation systems to overcome some specific disadvantages, such as the integration error in the INS systems, which decision is proposed by Rudolf Kalman, [3] known as Kalman filter. This filter is also used to combine the sensor data from different measurement systems. The alternative of Kalman filter is the complementary filter [4,5] which may be implemented easily in the embedded system due to its low complexity. It uses the constant values for the low and high pass filters, while the Kalman filter calculates the sensor noise and dynamically changes these values. The Kalman Filter is, however, known to provide an optimal estimate of the unknown state for a linear dynamic system with Gaussian distribution.

The current paper discusses the Kalman filter algorithm which is used to calculate the rotation angles of the kid car seat. The algorithm verification is shown in the stationary and moving situations and the results are commented. By integrating data from accelerometers and gyroscope, mounted on the car seat, we were able to predict and calculate its position.

[1] Emil Iontchev is with the Higher School of Transport "T. Kableshkov" 158 Geo Milev Street, Sofia 1574, Bulgaria, E-mail: e_iontchev@yahoo.com
[2] 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
[3] Petar Kapanakov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: peshoteslata2@abv.bg
[4] Lachezar Hristov is with the Higher School of Transport "T. Kableshkov" 158 Geo Milev Street, Sofia 1574, Bulgaria, E-mail: Lachezar.Hristov@outlook.com

II. ALGORITHM DESCRIPTION OF THE DATA COMBINATION

The rotation angles calculation of the object (kid car seat) is accomplished by 9DoF inertial measurement system, which is capable of reading the linear accelerations and angular velocities and magnetometer data too. The data of these nine degrees of freedom are combined to calculate the three Euler angles in the inertial coordinate system. The transformation of the coordinate system to the inertial one is based on the rotation matrix Rxyz [6].

The combination of the sensor data is made using the Kalman algorithm, which is well known and is implemented in many systems [7,8]. The object mathematical model assumes that the actual measurement at any time is related to the current state and the system model may be written as [9]:

(1) $\mathbf{x_{k+1}} = \mathbf{A}\mathbf{x_k} + \mathbf{w_k}$

(2) $\mathbf{z_k} = \mathbf{H}\mathbf{x_k} + \mathbf{v_k}$

where, $\mathbf{x_k}$ – denotes the column vector (nx1) of state estimate, $\mathbf{z_k}$ - column vector (mx1) of measurements, A - state transition matrix (mxn) which is applied to the previous state, H – the observation matrix (mxn) which maps the true state space into the observed space $\mathbf{w_k}$ – model noise and $\mathbf{v_k}$ – measurement noise, which are normally distributed and are independent. In the filter algorithms the noise is represented by the matrices Q(covariance matrix of $\mathbf{w_k}$) and R(covariance matrix of $\mathbf{v_k}$), which is used to calculated the Kalman coefficient

(3) $\mathbf{K_k} = \frac{\mathbf{P_k^- H^T}}{\mathbf{H P_k^- H^T + R}}$.

The practical implementation of the Kalman Filter is often difficult due to the difficulty of getting a good estimate of the noise covariance matrices. Several methods for the noise covariance estimation have been proposed such as autocovariance least-squares (ALS) technique. The predicted (*a priori*) state estimate is calculated by (4):

(4) $\mathbf{\hat{x}_k} = (\mathbf{1} - \mathbf{K_k})\mathbf{\hat{x}_k^-} + \mathbf{K_k}\mathbf{z_k}$

In the same time the Q matrix is used to obtain the predicted (*a priori*) error covariance $\mathbf{P_{k+1}^-}$:

(5) $\mathbf{P_{k+1}^-} = \mathbf{A P_k^- A^T} + \mathbf{Q}$

It is better to choose the smaller Q values and bigger R values if we want to decrease the state variations.

The system model for calculation of the Euler angles with quaternions (q) is given by the equation (6):

(6) $\begin{Bmatrix} \mathbf{q_1} \\ \mathbf{q_2} \\ \mathbf{q_3} \\ \mathbf{q_4} \end{Bmatrix}_{k+1} = \left(\mathbf{I} + \Delta \mathbf{t}\frac{1}{2} \begin{bmatrix} \mathbf{0} & -\mathbf{p} & -\mathbf{q} & -\mathbf{r} \\ \mathbf{p} & \mathbf{0} & \mathbf{r} & -\mathbf{q} \\ \mathbf{q} & -\mathbf{r} & \mathbf{0} & \mathbf{p} \\ \mathbf{r} & \mathbf{q} & -\mathbf{p} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{q_1} \\ \mathbf{q_2} \\ \mathbf{q_3} \\ \mathbf{q_4} \end{Bmatrix}_k \right)$

where **p, q, r** denote the angular velocities from the gyroscope measurements.

The transition matrix is given by (7):

$$(7) A = I + \Delta t \frac{1}{2} \begin{bmatrix} 0 & -p & -q & -r \\ p & 0 & r & -q \\ q & -r & 0 & p \\ r & q & -p & 0 \end{bmatrix}.$$

The quaternion values are calculated from the Euler angles according to the equation (8):

$$(8) \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} = \begin{cases} \cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \\ \sin\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} - \cos\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \\ \cos\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} \\ \cos\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} - \sin\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} \end{cases}$$

If the Euler angles are used as state estimates the nonlinear model is used according to the equation (9) and (10):

$$(9) \begin{pmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{pmatrix} = \begin{bmatrix} p + q\sin\phi\tan\theta + r\cos\phi\tan\theta \\ q\cos\phi - r\sin\phi \\ q\sin\phi\sec\theta + r\cos\phi\sec\theta \end{bmatrix} + w_k \equiv f(x) + w$$

$$(10) z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{Bmatrix} \phi \\ \theta \\ \psi \end{Bmatrix} + v_k \equiv hf(x) + v$$

The difference between the above model and the linear model(1,2) is that the linear matrix equation has been changed into nonlinear form as following Ax_k to f (x) and Hx_k to hf (x). To implement an extended Kalman filter the Jacobian of equation (9) must be known. This is defined as following (11):

$$(11) A = \begin{bmatrix} \frac{\partial f_1}{\partial \phi} & \frac{\partial f_1}{\partial \theta} & \frac{\partial f_1}{\partial \psi} \\ \frac{\partial f_2}{\partial \phi} & \frac{\partial f_2}{\partial \theta} & \frac{\partial f_2}{\partial \psi} \\ \frac{\partial f_3}{\partial \phi} & \frac{\partial f_3}{\partial \theta} & \frac{\partial f_3}{\partial \psi} \end{bmatrix}$$

In both system models, the Euler angles, determined by the accelerometer data, becomes the measurement in Kalman filter. They are stable over time but substantially deviating from the actual values, especially when additional linear accelerations affect them (12):

$$(12) \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{w} \end{bmatrix} + \begin{bmatrix} 0 & w & -v \\ -w & 0 & u \\ v & -u & 0 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} + g \begin{bmatrix} \sin\theta \\ -\cos\theta\sin\phi \\ -\cos\theta\cos\phi \end{bmatrix}$$

where u, v, w denotes the velocities and p, q, r – angular speed and a_x, a_y, a_z are accelerometer data, g – gravitational acceleration. If the system is stationary or moving with constant velocity, the equation (12) becomes as simple as following (13):

$$(13) \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} = g \begin{bmatrix} \sin\theta \\ -\cos\theta\sin\phi \\ -\cos\theta\cos\phi \end{bmatrix}$$

The angles ϕ and θ are calculated according to (14) and (15)[6]:

$$(14) \theta_a = atan\left(\frac{-a_x}{\sqrt{(a_y)^2 + (a_z)^2}}\right)$$

$$(15) \phi_a = atan\left(\frac{a_y}{signa_z\sqrt{(a_z)^2 + \mu(a_x)^2}}\right)$$

The μ parameter is included in the equation to prevent the null values division. The ψ angle calculation cannot be estimated

from the accelerometer data so the magnetometer data have to be used. It is calculated according to the equation (16) [6]:

$$(16) \psi_a = atan\left(\frac{Y_h}{X_h}\right),$$

where Y_h and X_h denotes the Y and X magnetometer values respectively, which are calculated according to the magnetometer data values, corrected by the inclination angles by equation (17) and (18)[10]:

$$(17) X_h = (b_x - V_x)\cos\theta_a + (b_y - V_y)\sin\theta_a\sin\phi_a + (b_z - V_z)\sin\theta_a\cos\phi_a$$

$$(18) Y_h = (b_z - V_z)\sin\phi_a - (b_y - V_y)\cos\phi_a$$

The measured magnetometer data are distorted by the soft and hard iron effects so it is necessary to calibrate the data. The magnetometer data calibration is accomplished by the algorithm described at [11]. The six data values are used for X value calculation according to the equation (19):

$$(19) X = \begin{bmatrix} b_x[0] & b_y[0] & b_z[0] \\ b_x[1] & b_y[1] & b_z[1] \\ ... & .... & ... \\ b_x[5] & b_y[5] & b_z[5] \end{bmatrix}$$

The chosen data values are also used for parameter Y estimation according to the equation (20):

$$(20) Y = \begin{bmatrix} b_x[0]^2 + b_y[0]^2 + b_z[0]^2 \\ b_x[1]^2 + b_y[1]^2 + b_z[1]^2 \\ ... & .... & ... \\ b_x[5]^2 + b_y[5]^2 + b_z[5]^2 \end{bmatrix}$$

The results from the equations (19) and (20) are used for parameter β estimation according to the equation (21):

$$(21) \beta = (X^TX)^{-1}X^TY$$

The calculated β values then are used for calculation of the magnetometer data disturbances (22):

$$(22) \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

The obtained values are subtracted from the measured values and the calculated ones are used for the Euler angle estimation.

III. EXPERIMENTAL RESULTS

The validation of the proposed algorithm is accomplished with/via static and dynamic measurements with the MEMS inertial sensor ADIS16405 [12], produced by Analog Devices. The first test checks the static performance of the algorithm and compare the seat rotation angles with the measured ones. The test is also used to calibrate the magnetometer and compensate the soft and the hard iron effects. The calculated Euler angles before and after magnetometer calibration are shown at Figure 1 and Figure 2. The magnetometer curves in the XY coordinate systems before and after calibration are shown at Figure 3.
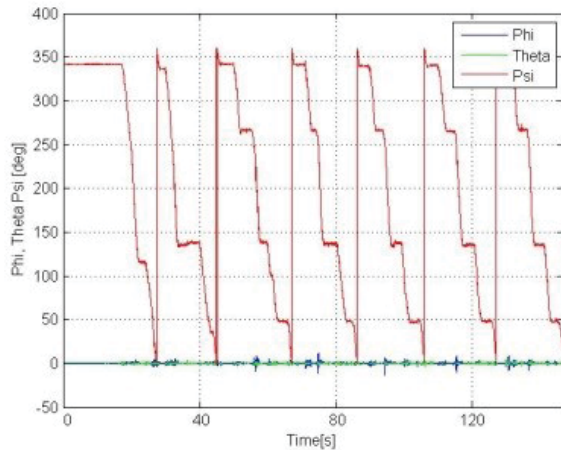
Figure 1. Euler angles before calibration
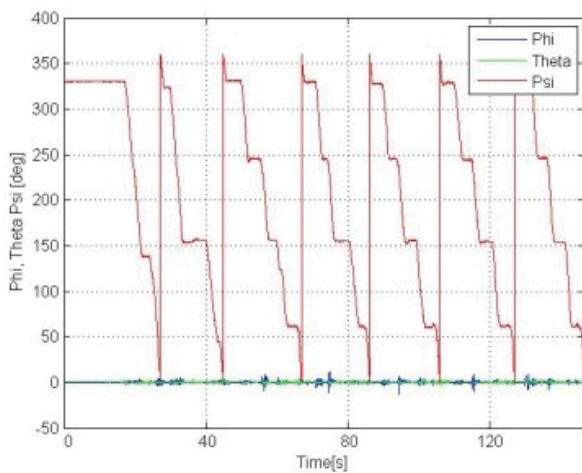
Figure 2. Euler angles after calibration
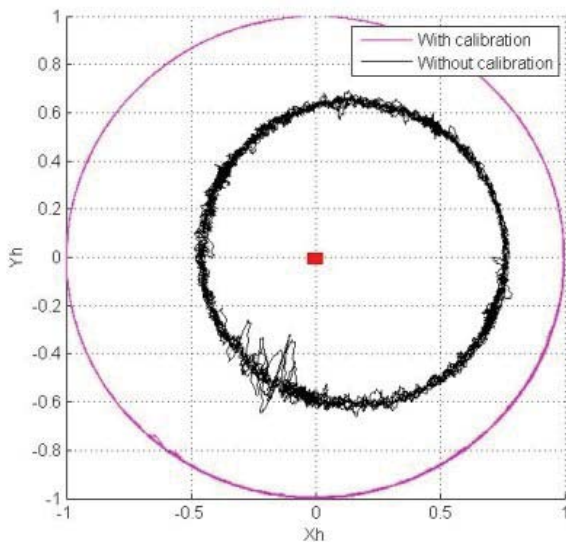
Figure 3. Representation of X_h and Y_h magnetometer data before and after calibration

The dynamic test is accomplished while the inertial module is installed on the kid car seat which is specially designed for

the measurements [13,14]. On the inertial module is mounted the mannequin with the size and weight of a child (Figure 4).

Figure 4. The inertial sensor position on the kid car seat
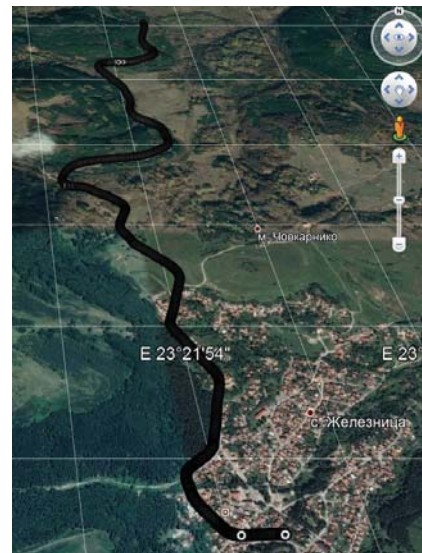
Figure 5. Experimental recorded track

The dynamic test is accomplished on a mountain road with lots of curves. The test road pieces are chosen in such way that the road direction is preliminary from north to south and the road curves are from east to west. The track and the speed are recorded with GPS receiver which is integrated in the inertial system design. The experimental track is shown at Figure 5.

The Euler angles are estimated from the linear accelerations, which are passed through low pass filter with cut-off frequency of 20Hz. Lower cut-off frequency lead to the increased latency of the Kalman filter output towards the acceleration data. Figure 6 represents ϕ angle values when the cut-off frequency is changed from 5 to 20Hz.
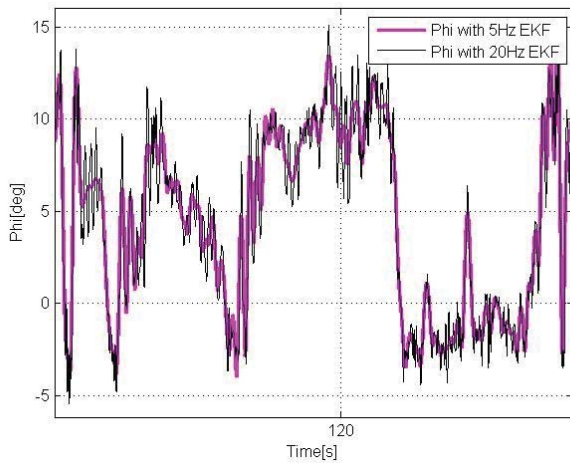
Figure 6. ϕ angle value after low pass filtering with cut-off frequencies from 5 to 20Hz

The results of the Euler angle estimation are shown at Figure 7.
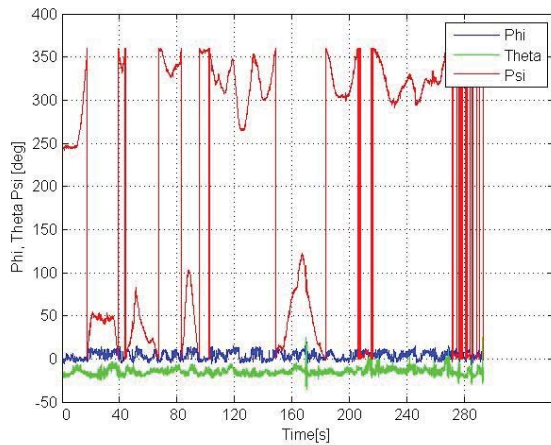
Figure 7. Estimated Euler angles

The ψ Euler angle values varies from 0 to 360 degrees due to the road curves and the travelling direction, which is constantly changed. We are interested in the ϕ Euler angle because it represents the rotation angle of the kid car seat, which movement is provoked by the lateral accelerations in the curves. The frequency of rotation of the kid car seat is also point of interest because the values represent the travelling comfort [15]. The estimated values of ϕ Euler angle are shown at Figure 8.
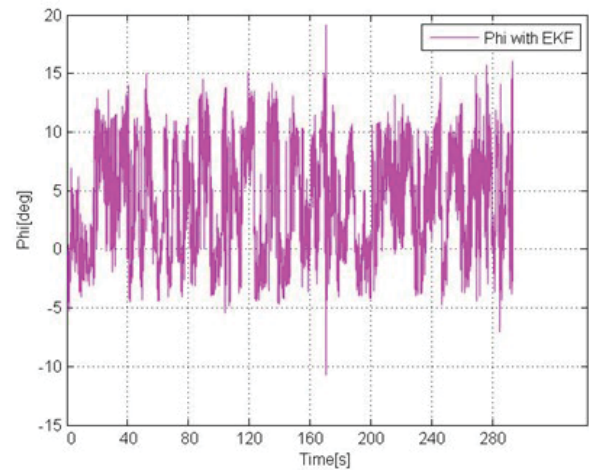
Figure 8. Estimated ϕ Euler angle values (rotation according to X kid car seat axis)

The Euler angles are calculated upon the quaternions model and the results are compared with the EKF ones and can be seen represented at Figure 9.
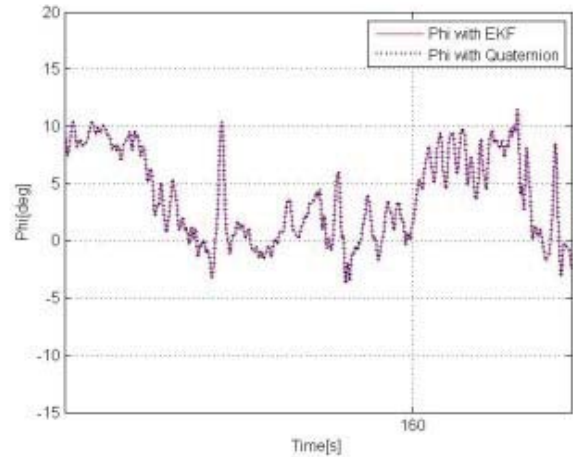
Figure 9. Comparison of the calculated ϕ angle with EKF and quaternions

It shows that there is no apparent difference in the angle values obtained using the two different models.
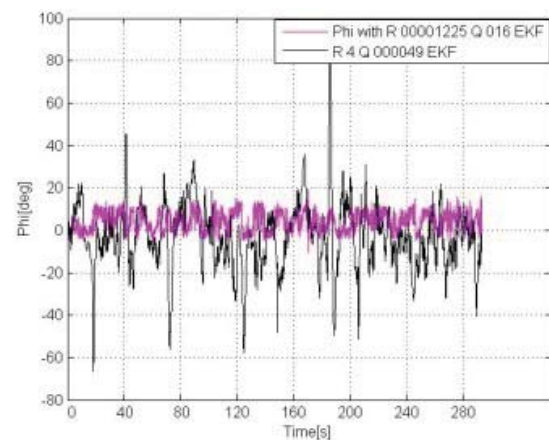
Figure 10. Euler ϕ angle depends from Q and R values

The model parameters are also studied for the Euler angle calculation. The first test (Figure 10, R=0.0001225, Q=0.16) used the noise parameters, which are estimated in our previous paper [16]. When R decreases and Q increases, the contribution of the measurement to the estimate increases. In this case the results match the actual deviation of the kid car seat.

IV. CONCLUSION

The obtained results show that the nonlinear Kalman filter model and the quaternion one represent the similar angle values. The proposed algorithm for the magnetometer calibration and data filtering are adequate. The chosen R and Q values, based on the sensor noise estimation, are included in the model and the results are compared with the real angle values. R and Q values are also studied to the model performance. Better results are obtained when R and Q values correspond to the values of the sensor noise estimation. The cut-off frequency of the low pass filter, which is applied to the accelerometer data, is also studied and it is shown that the lower cut-off frequency lead to the increased latency of the Kalman filter output towards the acceleration data. In the same time, higher cut-off frequency lead to the increased noise levels. Therefore the cut-off frequency of 20Hz is chosen.

REFERENCES

[1] Guochang Xu • Yan Xu, GPS Theory, Algorithms and Applications, Springer, ISBN 978-3-662-50365-2, 2016

[2] Roetenberg Daniel, Inertial and Magnetic Sensing of Human Motion, ISBN-13: 978-90-9020620-2

[3] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems, ASME Journal of Basic Engineering, series D: 35–45, 1960

[4] Manon Kok, Jeroen D. Hol and Thomas B. Sch¨on (2017), "Using Inertial Sensors for Position and Orientation Estimation", Foundations and Trends in Signal Processing: Vol. 11: No. 1-2, pp 1-153. http://dx.doi.org/10.1561/2000000094

[5] Walter T. Higgins, Jr., A Comparison of Complementary and Kalman Filtering, IEEE Thans. Aerospace and Electronic Systems, Vol. AES-li, no. 3, May 1975

[6] Pedley Mark, Tilt Sensing Using a Three-Axis Accelerometer, Freescale Semiconductor, Document Number: AN3461, Application Note Rev. 6, 03/2013

[7] Lewis Frank L., Shuzhi Sam Ge, Autonomous Mobile Robots Sensing, Control, Decision Making and Applications, Taylor & Francis Group, LLC, 2006, pp 99-149

[8] Jekeli Christopher, Jong-Ki Lee, Inertial/GPS Integrated Geolocation System for Detection and Recovery of Buried Munitions, SERDP Project MR-1565, 2011

[9] Kim Phil, Kalman Filter for Beginners with Matlab Examples, A-JIN Publishing company, 2011, ISBN-13: 978-1463648350

[10] Ozyagcilar Talat, Implementing a Tilt-Compensated eCompass using Accelerometer and Magnetometer Sensors, Freescale Semiconductor, Application Note Document Number: AN4248, Rev. 4.0, 11/2015

[11] Ozyagcilar Talat, Calibrating an eCompass in the Presence of Hard- and Soft-Iron Interference, Freescale Semiconductor, Application Note Document Number: AN4246, Rev. 4.0, 11/2015

[12]http://www.analog.com/static/imported-files/data_sheets/ADIS16400_16405.pdf - datasheet of ADIS16400/ADIS16405 inertial sensor family

[13] Nikolay Pavlov, Evgeni Sokolov, Mihail Peychev, Diana Dacova, Design and test of a tilting seat for improving children's comfort during traveling, Second International Scientific Conference ITEMA 2018 – Conference Proceedings, ISBN 978-86-80194-13-4, pp 305-312

[14] Pavlov N. L., E. E. Sokolov, Tilting Child Safety Seat for Reducing the Lateral Acceleration Acting on Children when Vehicle Cornering, ECCOMAS Thematic Conference on Multibody Dynamics, June 19 - 22, 2017, Prague, Czech Republic, pp 779-783

[15] Pavlov N. L., E. E. Sokolov, M. H. Peychev, D. I. Dacova and L. P. Kunchev, Comfort Improvement and Kinetosis Reduction by the Use of a Pendulum Type Child Travel Seat, International Journal of Mechanical Engineering and Technology (IJMET), Volume 10, Issue 03, March 2019, pp. 277-286. Article ID: IJMET_10_03_028, ISSN Print: 0976-6340 and ISSN Online: 0976-6359

[16] Iontchev E., Simeonov Iv., Miletiev R., Analysis of the Inertial MEMS Sensor Parameters for Navigation Applications, XV International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2011, Nish, Serbia, June 29 - July 1, 2011, pp 399, ISBN: 978-86-6125-032-3

Wavelet algorithm for denoising MEMS sensor data

Lachezar Hristov[1], Emil Iontchev[2], Rosen Miletiev[3] and Petar Kapanakov[4]

Abstract – **In this paper we propose a wavelet-based algorithm for denoising data acquired from MEMS based inertial sensors, in order to achieve accurate navigational information in terms of position, speed, acceleration and direction. It is well known that the low-cost MEMS inertial sensors output signals require a proper treatment in order to reduce the different random errors, consisting in the signal. The proposed algorithm is tested and the results are compared with the GPS data.**

Keywords – **Wavelet transform, denoising, inertial sensors.**

I. INTRODUCTION

MEMS integrated system, inertial sensor data includes large signal noise and sensor bias. The main error sources of the MEMS inertial sensors are recognized as: null offset error (bias) b_a, gain error - K, integral and differential nonlinearity and misalignment $- T^p$, specific force - F and output noise - e_n. [1] The influence of these errors may be written by the Eq. (1)

$$A = K(T^p)^{-1}F + b_a + e_n \qquad (1)$$

Among them, sensor biases make more significant error on position result because of integration. Sensor biases are well compensated on good observable trajectory in general integrated navigation system. But sensor biases cannot exactly be compensated on low cost system. If the noise component could be removed, the overall inertial navigation accuracy is expected to improve considerably. The resulting position errors are proportional to the existing sensor bias and sensor noise.

In this paper, the wavelet denoising technique is implemented to eliminate the sensor noise for improving the performance of inertial sensor signals [2], [3]. The wavelet analysis has the advantage over other signal processing techniques in the capability of performing local analysis. It can decompose the signal to frequency component in local time. By using this characteristic, the wavelet denoising method shrinks the signal noise by eliminating the frequency

[1]Lachezar Hristov is with the Faculty of Telecommunications and Electrical Equipment in Transport "Todor Kableshkov" University of Transport, 158 Geo Milev Str, Sofia 1574, Bulgaria, E-mail: Lachezar.Hristov@outlook.com.

[2]Emil Iontchev is with the Faculty of Telecommunications and Electrical Equipment in Transport "Todor Kableshkov" University of Transport, 158 Geo Milev Str, Sofia 1574, Bulgaria, E-mail: e_iontchev@yahoo.com.

[3]Rosen Miletiev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: miletiev@tu-sofia.bg

[4]Petar Kapanakov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: peshoteslata2@abv.bg

component which contains only noise. Furthermore, the wavelet denoising method can also remove the signal bias by decomposing the signal and eliminating the low frequency component.

The wavelet thresholding method is verified using the collected real data from the field test. It is implemented to the MEMS integrated system, later shown in the experimental data section.

II. WAVELET DENOISING ALGORITHM

A. Wavelet Transform

Wavelet transform is a signal transform technique popularly used in image and audio signal processing [4], [5], [6]. Compared by nature of the processing type, the traditional Fast Fourier Transform (FFT) has a fixed relationship between time and frequency while the wavelet does not have a fixed relationship between time and frequency. The wavelet analysis is based on a windowing technique with variable-sized windows shown in Fig. 1. The wavelet transform applies the wide window (long time intervals) to low frequency and the narrow window (short time intervals) to high frequency.
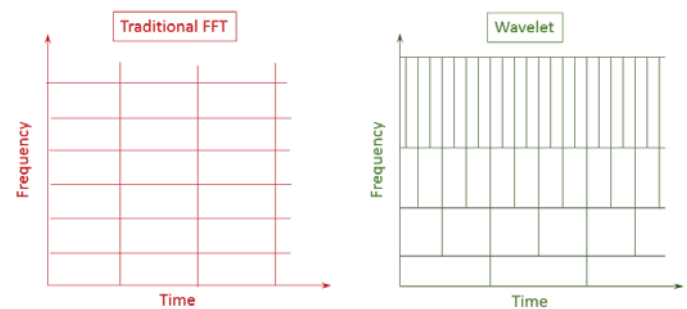
Fig. 1. Time-Frequency Domain Sampling

Discrete time wavelet transform is executed as Eqs. (1), (2) [7].

$$\phi(t) = \sum_k g_0[k]\phi(2t - k) \qquad (2)$$

$$\Psi(t) = \sum_k g_1[k]\phi(2t - k) \qquad (3)$$

Where ϕ is called the scale function and Ψ is called the wavelet function. Each g_0 and g_1 refers to the wavelet coefficient.

Wavelet function can be any function that satisfies the relationship of Eqs. (1) and (2). The scale function of upper level can be expressed as the convolution of the scale function and the wavelet function of lower level. It means that the low-frequency area can be decomposed to the high-frequency area and low-frequency area. Such relationship is shown in Fig. 2

Fig. 2. Wavelet Transform

B. Wavelet Thresholding Technique

Wavelet thresholding technique is a signal estimation technique that exploits the capabilities of wavelet transform for signal denoising. It removes the noise by eliminating coefficients that are insignificant relative to some threshold. Researchers have developed various techniques for choosing denoising parameters none of which is best universal threshold determination technique. Wavelet thresholding technique assumes that the magnitude of the actual signal is greater than the noise level, and in general the noise is white noise.

The wavelet thresholding technique reduces the noise level with almost no distortion for the sudden change in signal. The result of the thresholding technique has almost no distortion and accurate, so that it sits right on the actual signal almost indistinguishably. Therefore, it can be used by preprocessing filter for the inertial sensor signal and overcomes the shortages of the existing low-pass filter.

The thresholding technique is classified into the hard thresholding and soft thresholding operation.

$$T_\lambda^{hard} = \begin{cases} u & if|u| \geq \lambda \\ 0 & otherwise \end{cases} \quad (3)$$

$$T_\lambda^{soft} = \begin{cases} (u - sign(u)\lambda) & if|u| \geq \lambda \\ 0 & otherwise \end{cases} \quad (4)$$

The Eq. (3) shows the hard thresholding function and the Eq. (4) shows the soft thresholding function. The wavelet coefficient means that the magnitude of a certain frequency component. Thus, wavelet coefficients in the band of noise become zero. The soft thresholding technique is known to have better performance than the hard thresholding technique, so the soft thresholding technique was used for the experiment [8], [9].

Important element influencing the performance in thresholding technique is how the standard value of λ is set [2], [8]. Generally, it is determined by the Eq. (5). Here, σ is the standard deviance of the signal, and *n* is the number of signal samples

$$\lambda = \sqrt{2 \log n} \, \sigma \quad (5)$$

The value determined by the formula, will not be the optimal result, so the appropriate λ must be determined through experimentation.

D. Wavelet Denoising Algorithm

The proposed denoising algorithm is executed as follows. Select wavelet based on predefined list. For the input signal, using interval-dependent thresholding method - obtain the maximum level of decomposition, define the intervals and interval-dependent thresholds, decompose the signal, replace original wavelet coefficients by the coefficients resulting from the thresholding operation and perform wavelet transform - apply the thresholds and reconstruct the signal. Integrate the wavelet denoised signal to get the speed and distance. Calculate the SNR and endpoint accuracy against the noisy signal speed and distance. If all predefined wavelets are used, compare the results and draw graphics. Estimate the most appropriate wavelet for denoising the MEMS based accelerometer data.

For the purpose of the experiment the denoising algorithm is implemented using MATLAB.

The process is shown in the flowchart on Fig. 3
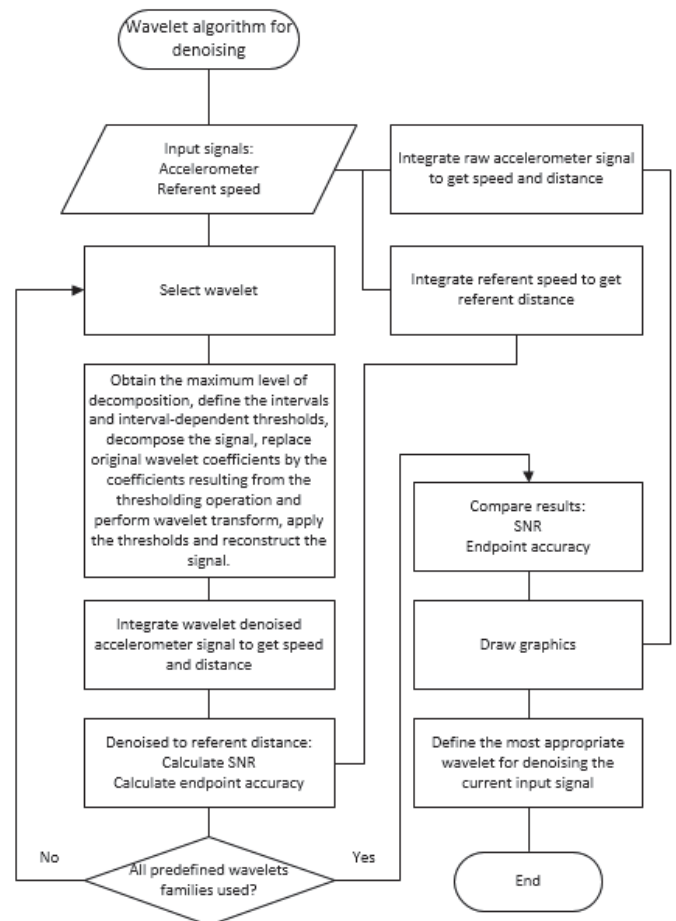
Fig. 3. Block diagram of the denoising

III. EXPERIMENTAL RESULTS

Collected MEMS sensor signal data includes the actual vehicle motion dynamics and the sensor noise as well as some other undesirable noise such as vehicle engine vibration.

The wavelet transform was carried out using predefined list of wavelets by function cmddenoise. Table I shows the experimental data results for each wavelet used – name, calculated maximum level of decompositions, calculated best number of intervals, calculated SNR of denoised signal compared to noise from signal data, deviation between calculated denoised signal last point distance and reference signal last point distance. Results are sorted based on the best SNR.

TABLE I
EXPERIMENTAL DATA

Signal	Wavelet	Max Lvl Decomp	Best Nb of Int	SNR path [dB]	dev last point dist [%]
Accel.	db2	10	3	13.6674	2.8222
Accel.	sym2	10	3	13.6674	2.8222
Accel.	coif1	9	5	13.5842	3.1055
Accel.	db3	9	3	13.5464	3.2062
Accel.	sym3	9	3	13.5464	3.2062
Accel.	sym4	8	1	13.4629	3.4765
Accel.	sym6	8	1	13.4335	3.5684
Accel.	coif2	8	1	13.4139	3.6311
Accel.	db4	8	2	13.3311	3.8938
Accel.	db6	8	1	13.3110	3.9543
Accel.	sym8	7	1	13.2793	4.0551
Accel.	db7	8	2	13.2777	4.0643
Accel.	sym7	8	3	13.2765	4.0700
Accel.	coif3	7	1	13.2726	4.0765
Accel.	haar	11	1	13.2648	3.9585
Accel.	coif4	7	1	13.2502	4.1492
Accel.	sym5	8	1	13.2439	4.1706
Accel.	db8	7	1	13.2402	4.1836
Accel.	coif5	6	1	13.2126	4.2740
Accel.	db9	7	1	13.1922	4.3415
Accel.	db5	8	1	13.1892	4.3506
Accel.	db10	7	1	13.1862	4.3579

It is noticeable that "db2", "sym2" and "coif1" wavelets have the best SNR ratio. The calculation of the last point deviation from reference distance confirm that the wavelets being on the top of the table is the most appropriate as they show less deviance from the last distance point and it is more accurate than the other wavelets in the table.

Figs. 4 and 5 shows a comparison between noisy signal and the top three best SNR ratio denoised signals. On the detailed Figure 5 we can see that "db2" and "sym2" wavelet results are almost identical and their lines are one over another.
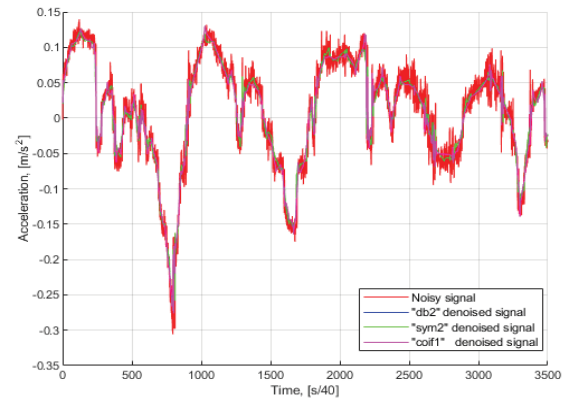
Fig. 4. Comparison between noisy signal and top three most effective wavelet denoised signals
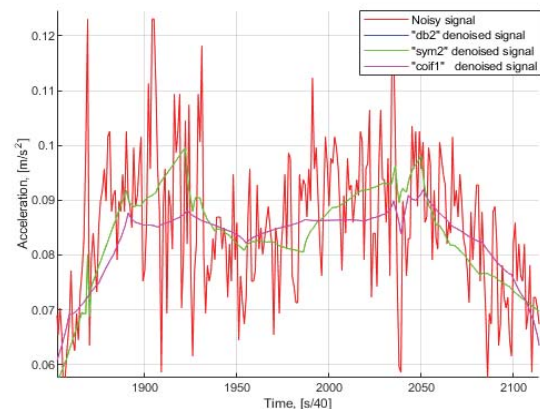
Fig. 5. Detailed comparison between noisy signal and top three most effective wavelet denoised signals

On Fig. 6 and more detailed on Fig. 7 we can see that denoised signal is following the reference speed curve most accurate.
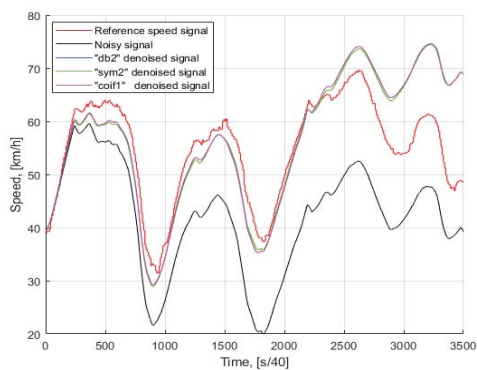
Fig. 6. Comparison between speed calculation results - reference speed signal, noisy signal and top three most effective wavelet denoised signals
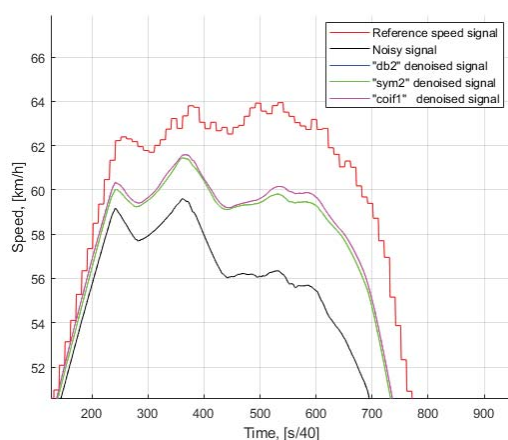
Fig. 7. Detailed comparison between speed calculation results - reference speed signal, noisy signal and top three most effective wavelet denoised signals
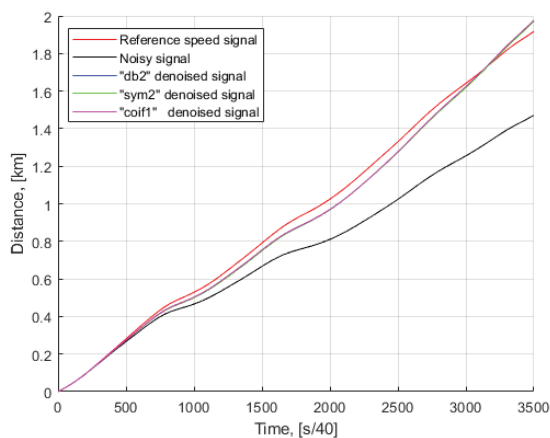
Fig. 8. Comparison between accumulated distance results - reference signal, noisy signal and top three most effective wavelet denoised signals

IV. CONCLUSION

In order to achieve accurate navigational information in terms of position, speed, acceleration and direction it is necessary to use proper denoising. The proposed algorithm will help us calculate the most accurate position, speed or distance by using raw signal acquired by MEMS Inertial Navigation Systems during the blind zones where GPS signal is weak or missing.

In the article we propose an algorithm for selecting the most appropriate wavelet for denoising MEMS sensor data, based on input noisy signal analysis, while searching for the best calculated SNR for each of the applied wavelet denoising.

Calculated the best SNR shows the most appropriate wavelet to be applied to the signal being processed. SNR is calculated by only using data from input noisy signal. Referent signal is used to show and confirm that the algorithm proposes the best wavelet which will approve the results on later calculations by having the smallest last point deviance as well.

REFERENCES

[1] E. Iontchev, Algorithm for denoising parameters selection with wavelets, Elektrotechnica & Elektronica E+E, Vol. 47 No 3-4/2012, pp37.

[2] C.W.Kang, C.G.Park, "Improvement of INS-GPS Integrated Navigation System using Wavelet Thresholding", Journal of The Korean Society for Aeromautical and Space Sciences, Vol.37, No.8, 2009.

[3] AHMED M. HASAN, "Comparative study on Wavelet Filter and Thresholding Selection For GPS/INS Data Fusion," International Journal of Wavelets, Multiresolution and Information Processing, Vol 8, No.3, 2010.

[4] Vetova, S., I. Draganov, I. Ivanov, "CBIR with Dual Tree Complex Wavelet Transform using Maximally Flat All-pass Filter", Journal Electrotechnics and Electronics, "E+E", vol. 53, 11-12, 2018, pp. 314-320, ISSN 0861-4717.

[5] Vetova, S., Ivanov I., "Image Features Extraction Using The Dual-Tree Complex Wavelet Transform", 2nd International Conference on Mathematical, Computational And Statistical Sciences, Gdansk, Poland, 2014, pp. 277 – 282, ISBN: 978-960-474-380-3.

[6] Vetova, S., Ivanov I., "Content–Based Image Retrieval Algorithm Based On The Dual-Tree Complex Wavelet Transform: Efficiency Analysis", 5th European Conference of Computer Science, Geneva, Switzerland, 2014, pp. 71 – 77, ISBN: 978-1-61804-264-4.

[7] Jaideva C. Goswami, Andrew K. Chan, "Fundamentals of Wavelets: Theory, Algorithms, and Applications," Wiley, 1999.

[8] Crouse, M. S., Nowak, R. D., Baraniuk, R. G., "Waveletbased signal processing using hidden Markov models", IEEE Trans. Signal Proc., vol. 46, pp. 886-902, Apr. 1998.

[9] Gao, H.-Y., and Bruce, A. G., "WaveShrink with firm shrinkage", Technical Report 39, StatSci Division of MathSoft, Inc. 1996.

Test methodology for mains interference frequency measurement

Georgy Mihov [1] and Dimiter Badarov [2]

Abstract – **A test procedure for mains interference frequency measuring is represented in this work. The procedure is based on measurement of the transfer coefficient of an averaging digital filter applied to the interference signal. The procedure is implemented in MatLab environment. It includes filters for elimination of the DC offset, for suppression of the second and third harmonic and for measurement of the power-line interference frequency. Additionally, a procedure for measurement of the amplitude of the interference is developed.**

Keywords – **Mains interference, Frequency measurement, Harmonic suppression.**

I. INTRODUCTION

A method for measurement of a mains interference (hum) frequency in electrocardiographic (ECG) signal is represented in [1] that is further developed for mains frequency measurement [2, 3]. It determines the deviation dF of the mains frequency F, calculating the change of the transfer coefficient $K(f)$ of an averaging filter – see Fig. 1.
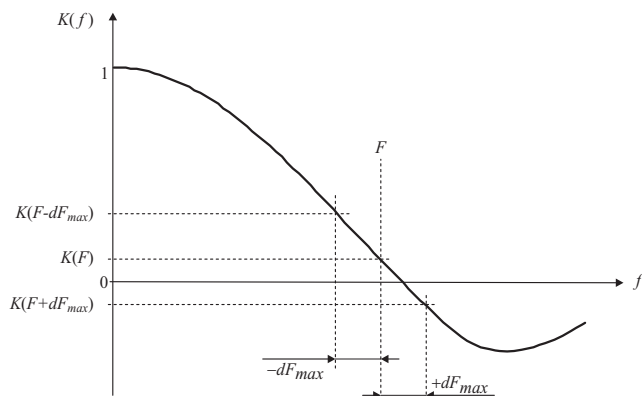
Fig. 1. An averaging filter transfer coefficient changing due to the mains frequency deviation.

The method has advantages in front of other existing methods [4, 5], especially when existing recordings with relatively low sampling frequency are used. The errors induced by harmonic content of the mains interference and by its amplitude changing are analyzed in [6].

In the present work a computer based [7, 8] test methodology allowing detailed investigation of the method is proposed.

II. HARMONIC SUPPRESSION FILTERS

A. Third harmonic suppression

For the purpose of third harmonic filtration, a digital filter Y is synthesized by summation of two filters by the methodology proposed in [9]. Two "three-point" filters are used $Y1$ and $Y2$ which have frequencies of their first zeroes on the both sides of the third harmonic of the mains interference F_3. The two filters are summed by multiplication with complementary to unity coefficients $(1 - k_y)$ и k_y.

$$Y_i = \left(1 - k_y\right).Y1_i + k_y.Y2_i; \quad \left| \begin{array}{l} Y1_i = \dfrac{B_{i-m_3} + 2B_i + B_{i+m_3}}{4} \\ Y2_i = \dfrac{B_{i-m_3-1} + 2B_i + B_{i+m_3+1}}{4} \end{array} \right., (1)$$

where m_3 is the rounded to the lower or equal integer number of samples in the half period of the third harmonic F_3 of the mains frequency.

The transfer coefficient of the resultant filter is also expressed by summation of the transfer coefficients of the two filters multiplied by the complementary to unity coefficients $(1 - k_y)$ и k_y.

$$K(f) = \left(1 - k_y\right).K1(f) + k_y.K2(f), \quad \left| \begin{array}{l} K1(f) = \cos^2 \dfrac{m_3 \pi f}{Q} \\ K2(f) = \cos^2 \dfrac{(m_3+1)\pi f}{Q} \end{array} \right., (2)$$

The coefficient k_y is defined by applying the condition the first derivative of the transfer function of the summed filter $K^I(f)\big|_{f=F_3} = \left(1 - k_y\right).K1^I(f)\big|_{f=F_3} + k_y.K2^I(f)\big|_{f=F_3} = 0$ to have a zero value for the third harmonic of the mains interference $f = F_3$, from where:

$$k_y = \frac{K1^I\left(F_3\right)}{K1^I\left(F_3\right) - K2^I\left(F_3\right)};$$

$$\left| \begin{array}{l} K1^I\left(F_3\right) = -\dfrac{\pi m_3 F_3}{Q}\sin\dfrac{2\pi m_3 F_3}{Q} \\ K2^I\left(F_3\right) = -\dfrac{\pi(m_3+1)F_3}{Q}\sin\dfrac{2\pi(m_3+1)F_3}{Q} \end{array} \right.; \quad (3)$$

and when $f = F_3$ the parameter k_y have a value:

[1] Georgy Mihov is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gsm@tu-sofia.bg .

[2] 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

$$k_y = \frac{\sin\dfrac{2\pi m_3 F_3}{Q}}{\sin\dfrac{2\pi m_3 F_3}{Q} - \dfrac{m_3+1}{m_3}\sin\dfrac{2\pi(m_3+1)F_3}{Q}}, \qquad (4)$$

This way the summed filter Y by equation (1) has a minimum of the transfer coefficient for the frequency of the third harmonic of the mains interference F_3 but it is not zero. For that purpose, it is recurrently modified by the methodology proposed in [1]:

$$Y*_i = Y_i - (X_i - Y_i)\frac{K(F_3)}{1 - K(F_3)}, \qquad (5)$$

where $K(F_3)$ is the transfer coefficient of the summed filter calculated by (2)

$$K(F_3) = (1 - k_y).K1(F_3) + k_y.K2(F_3), \quad \begin{vmatrix} K1(F_3) = \cos^2\dfrac{m_3\pi F_3}{Q} \\ K2(F_3) = \cos^2\dfrac{(m_3+1)\pi F_3}{Q} \end{vmatrix}.(6)$$

The modified filter $Y*$ acquires transfer coefficient of zero for the third harmonic of the mains frequency F_3 keeping tangential character to the frequency axis. The synthesis of digital filter for suppression of the third harmonic of the mains frequency $F_3 = 150\ Hz$ is represented on Fig. 2. The plot contains the first $Y1$ (blue curve - b) and the second $Y2$ (red curve - r) from the initial "three-point" filters, the summed filter Y (green curve - g) and the modified filter $Y*$ (black curve - k). $\mathbf{H}$ is the impulse vector of the summed filter $K(F_3)$ containing the weight coefficients of the impulse response.
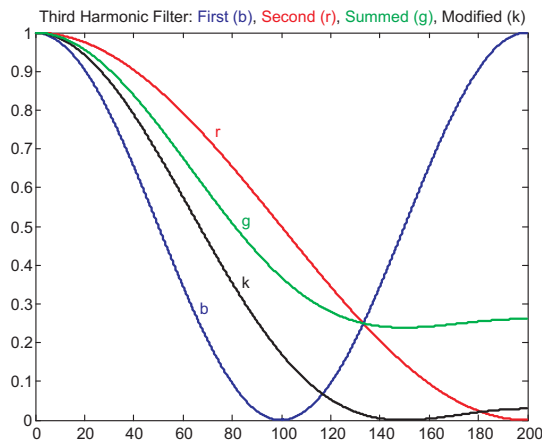
Fig. 2. Synthesis of filter for third harmonic of the mains frequency $F_3 = 150\ Hz$ for sampling frequency $Q = 400$ Hz, $\mathbf{H} = [0.0858\ 0.2426\ 0.3431\ 0.2426\ 0.0858]$.

The filter for the third harmonic reduces the amplitude of the mains frequency with about 0.7. The conducted experiments showed that the amplitude of the steady value F of the mains frequency does not affect the accuracy of the measurement.

B. Second harmonic suppression

In similar way we can build digital filter for suppressing the other harmonics of the extracted mains signal. The synthesis of digital filter for second harmonic of the mains frequency $F_2 = 100\ Hz$ is represented on Fig. 3. The plots are the same as the ones shown on Fig. 2. The frequency of the second harmonic is substituted in equations (1-6) along with the number of samples in the half period of the second harmonic m_2.
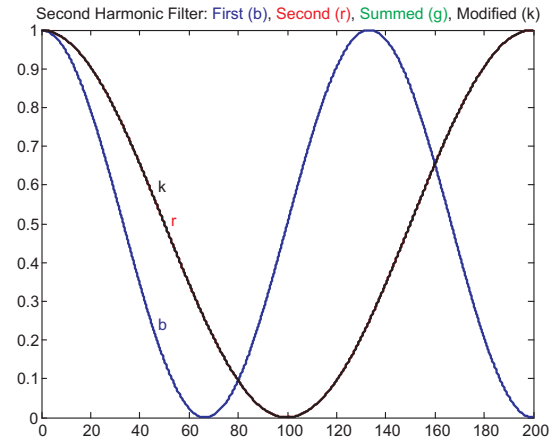
Fig. 3. Synthesis of filter for second harmonic of the mains frequency $F_2 = 100$ Hz for sampling frequency $Q = 400$ Hz, $\mathbf{H} = [0.25\ 0\ 0.5\ 0\ 0.25]$.

The filter for the second harmonic suppression reduces the amplitude of the mains frequency with about 0.5 but this does not affect the accuracy of the measurement of the mains interference frequency.

A very important property can be seen in Fig. 3 with sampling frequency $Q = 400\ Hz$. The coefficient k_y equals one and due to that the complementary to unity coefficient $(1 - k_y)$ for the other filter equals zero. This means that the summed "three-point" filter Y is identical with the second $Y2$ and the modified filter $Y*$ - see equation (1-6). On the figure the three filters are represented one over another. For that reason, only the modified filter is visualized. This can be seen also from the $\mathbf{H}$ vector of the impulse response which contains only one of the summed filters. This property allows easy construction of digital filter for second harmonic suppression when the sampling frequency is $400\ Hz$.

C. Elimination of the DC offset of the signal

The measured mains interference can contain a DC offset. It can be generated by the analog part of the measurement unit or from the output of the analog-to-digital conversion process when we are not using a four-quadrant analog-to-digital converter (ADC). For the purpose of mains frequency extraction from the input signal a simple "three-point" filter is used.

$$Y_i = \frac{-B_{i-m_2} + 2B_i - B_{i+m_2}}{4} \qquad (7)$$
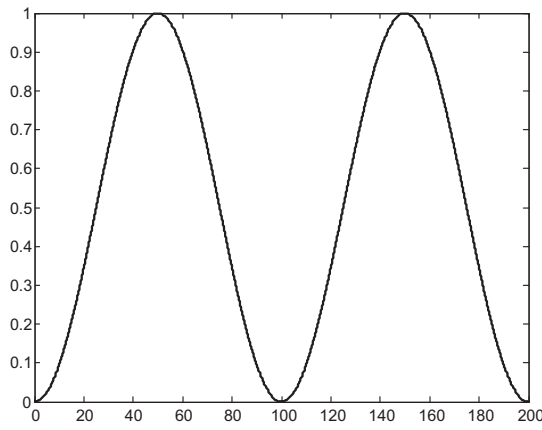
Fig. 4. Filter for DC offset elimination for sampling frequency $Q = 400$ Hz.

From Fig. 4 we can see that with sampling frequency of 400 *Hz* the DC offset filter is suppressing all even harmonics.

III. ALGORITHM AND PROGRAM IMPLEMENTATION OF THE METHOD

The algorithm of the test methodology is represented on Fig. 5. First the input signal is subjected to mains interference extracting using the filter shown on Fig. 4.
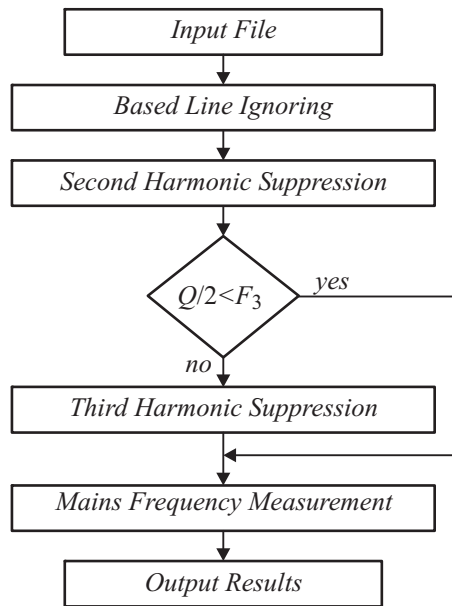
Fig. 5. Algorithm of the test methodology for frequency measurement.

The next step is to suppress the second harmonic of the mains interference signal. If the sampling frequency is multiple of the frequency of the second harmonic this step can be skipped because the mains interference extraction procedure suppresses all even harmonics of the interference.

Then it is checked if the frequency of the third harmonic of the interference F_3 is higher than $Q/2$. If it is not true, the third harmonic suppression procedure is applied.

The calculation of the mains frequency deviation is accomplished by the application of a "two-point" filter using the procedure given in [3].

The results from the application of the different parts of the test program are shown on Fig. 6. The experiment is carried out with a synthesized input signal containing mains interference with changing amplitude by a sinusoidal law with frequency of 0.1 *Hz*. The signal duration is 30 *s* which is divided on three parts. In the first 10 *s* the frequency of the mains interference is 50.25 *Hz*. In the second 10 *s* the frequency is 50 *Hz* and in the last 10 *s* the frequency is 49.75 *Hz*.

The first subplot of Fig. 6a shows the input file, the second subplot shows the file after the procedures for extraction and suppression of the harmonics of the mains interference. The third subplot shows the calculated mains frequency. The figures **b**, **c** and **d** contain the spectrum of the signals calculated by Fourier transform: **b** – of the input signal; **c** – after the extraction and suppression of the second harmonic and **d** – after the suppression of the third harmonic.

a)

b)

c)

Fig. 6. Test with a synthesized signal and sampling rate $Q = 400$ *Hz*: **a)** – visualization of the result; **b)** – spectrum of the input signal; **c)** – spectrum after mains interference extraction and second harmonic suppression; **d)** – spectrum after third harmonic suppression.

The next demonstrations are done with real interference signals extracted from old electrocardiographic signals. For the signal BN039_400.adc the examined epoch is 8 *s* and the calculated frequency values are averaged for period of 1 *s*. For the signal BR024_200.adc the examined epoch is 160 *s* and the calculated frequency is averaged for a period of 10 *s*.

Fig. 7. Experiments with real signals with $Q = 400$ *Hz*.

Fig. 8. Experiments with real signals with $Q = 200$ *Hz*.

IV. CONCLUSION

A test methodology is developed for examination of the method for mains interference frequency calculation using the transfer coefficient of a digital moving averaging filter.

For the cases in which the sampling frequency is higher than the doubled frequency of the third harmonic a suppression filter is developed. The filter is based on the summation and modification of two digital filters having transfer characteristics tangential to the frequency axis.

By the same methodology a filter for suppressing of the second harmonic is synthesized. The procedure keeps its accuracy without increase of the measurement error.

A filter for elimination of the DC component of the input signal is synthesized based on a "three-point" filter. It is found that when the sampling frequency is two or four times higher than the frequency of the second harmonic, the filter also suppresses the frequencies of all even harmonics of the mains interference.

The methodology is developed and tested in the MatLab environment. Different functions are developed for mains interference extraction and suppression of the second and third harmonic.

REFERENCES

[1] G. Mihov, R. Ivanov, C. Levkov, "Subtraction Method for Removing Powerline Interference from ECG in Case of Frequency Deviation", Proceedings of the Technical University – Sofia, Vol. 56, b. 2, 2006, pp. 212-217.

[2] G. Mihov, "Power-Line Interference Frequency and Amplitude Measurement Using the Subtraction Procedure", Proceedings of the Technical University of Sofia, Vol. 62, (4), 2012, pp. 189-198, (in bulgarian).

[3] D. Badarov, G. Mihov, "Mains Frequency Deviation Measurement by Using Elements of the Subtraction Procedure Based on Xilinx FPGA", Proceedings of the 24-th International Conference ELECTRONICS – ET'2015, Vol. 9, September 2015, pp. 112-115.

[4] B. Jovanovic, M. Damnjanovic, "Digital System for Power Line Frequency Measurement", Proceedings of the 48-th ETRAN Conference, Vol. 1, June 2004, pp. 29-32.

[5] S. Ovcharov, N. Tyuliev, P. Yakimov, "A system for mains frequency observing", ELECTRONICS ET'94. Vol. I, Sozopol, Bulgaria, September 28-30, 1994, pp. 98-103.

[6] D. Badarov, G. Mihov, (2017). "Analysis of the Spectrum and Amplitude Error of the Frequency Measurement Using Elements of the Subtraction Procedure", IEEE 40-th International Spring Seminar on Electronics Technology ISSE2017, Sofia, Bulgaria, May 10-14, 2017, 1-6 p.

[7] R. Romansky, "A Formal Approach for Modelling and Evaluation in the Field of Computing", International Transaction on Electrical, Electronics & Communication Engineering, UK, No 4 (vol. 2) , 2012, pp.1-7.

[8] G. Mihov, D. Badarov, "Testing of digital filters for power-line interference removal from ECG signals", ELECTRONICS ET2017, Sozopol, September 13-15, 2017, pp. 1-6.

[9] G. Mihov, "Investigation of the FIR Filters Usage in the Subtraction Method for Power-Line Interference Removing from ECG", Proceedings of the Technical University – Sofia, Vol. 57, b. 2, 2007, pp. 84-93.

AC Current Transducer as an Element of the Electrical Energy Consumption Control System

Viša Tasić[1], Marijana Pavlov-Kagadejev[2], Radoš Jeremijić[3], Vladimir Despotović[4],

Olivera Tasić[5], Ivana Stojković[6]

Abstract – **A paper presents the elements of the realized AC current transducer which solves the problem of converting the input signal (AC current) into a DC voltage signal suitable for further processing in the distributed control system. Such devices are heavily applied in distributed control systems for controlling the electricity consumption, for monitoring of industrial processes, as feedback elements, in telemetry and etc. The proposed AC current transducer is further connected to Arduino microcontroller to form a basic unit of the distributed control system for the control of electrical energy consumption in households.**

Keywords – **AC current, transducer, electricity, consumption**

I. INTRODUCTION

The principle of measuring electrical current by digital instruments is based on voltage measurement. The measured current passes through a resistor of known electrical resistance. By measuring the voltage at the ends of the resistors, the measured current is indirectly determined [1].

AC current converters and transducers are necessary elements in the realization of the Distributed Control System (DCS) because the control and management of the electrical energy consumption requires continuous measurement of electricity/energy in real time [2]. DCS systems accept standard current and voltage signals (4-20 mA, 0-5 V DC) at their input. For this reason, it is necessary to adjust/convert all input signals to a format suitable for further processing in

[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-Kagadejev 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

DCS system.

AC current transducers have wide application in DCS systems for controlling electrical energy consumption, for monitoring of technical processes, as feedback elements, in telemetry and etc. The choice of the appropriate transducer depends on a number of factors, such as: measuring range, accuracy of the measurement, commercial and other conditions. AC current transducers mainly represent a combination of current transformer, precision rectifier, and stage for adapting the output signal level (0-20 mA, 4-20 mA, 0-5 V DC or 0-10 V DC). The following examples indicate that there is a need for this type of device on the market.

The CRT4100AC [3] is designed to measure AC currents in the range 0-150 A, with galvanic isolation of the primary circuit. These devices can be ordered with the desired current or voltage output: 4-20 mA, 0-5V DC or 0-10V DC.

The Swiss company, LEM manufactures, produces AC Current Transducer AP-B10 for measuring AC currents in the range 0-400 A, with galvanic isolation of the primary circuit [4]. The output range can be selected either 0-5V DC or 0-10V DC. YHDC Company manufactures AC current transducers (ACT series) that contain a current transformer and electronics for conditioning of signals in one device [5]. The ACT devices can be made for input AC currents in the range 0-2 A up to the range 0-2000 A. Output signal can be DC current 4-20 mA or DC voltage 0-10 V. The Indian company, Adept Fluidyne, has developed Adept AC Current Transducer 2010 [6]. The output DC signal is proportional to the input AC current and is calibrated to the RMS value.

In the Republic of Serbia, the Company MINEL AUTOMATIKA A.D. Belgrade has developed MPI/L device for measuring the AC current (AC input 0-1 A or 0-5 A, output 0-20 mA or 0-10V) [7].

II. AC CURRENT TRANSDUCER HARDWARE

A. General Characteristics of AC Current Transducer

The realized AC current transducer is primarily intended for measuring AC current in the range 0-10 A, at the frequency 50-60 Hz. If necessary the range of the input AC current can be expanded to 0-50 A, with the appropriate transformer installed in transducer. The input AC current inside the transducer is transformed into a DC voltage 0-5 V signal, which is galvanically separated from the input signal. The wiring diagram of the proposed AC current transducer is shown in Fig 1.
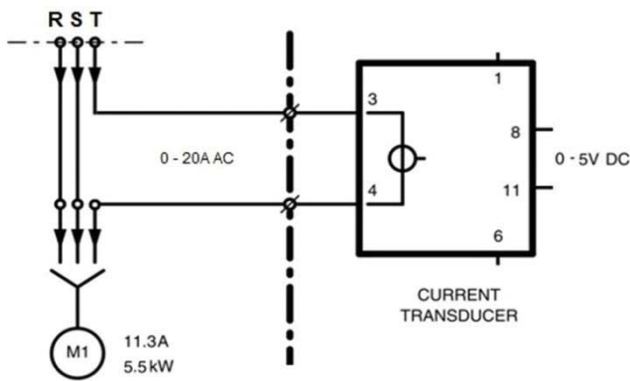
Fig. 1. The wiring diagram of the proposed AC current transducer

The realized AC current transducer is placed in a standard plastic housing with the dimension 50x78x108 mm for installation on a standard 35 mm wide rail (according to EN 60715 standard) with IP20 protection degree. On the front of the transducer housing, there are 12 terminal clamps arranged in two rows of 6 terminals, as shown in Fig 2. Terminals 1 and 6 are intended for connecting the supply voltage of the transducer, 220 V AC. Terminals 3 and 4 are intended for connection to the electric grid (AC input, 0-10 A).

Fig. 2. The transducer's front panel and housing

All transducer's components are located on a single-sided printed circuit board shown in Fig 3. The electrical scheme of the transducer (shown in Fig 5.) consists of the following components: power supply, current transformer/sensor, inverting amplifier, precision two-sided rectifier and output stage.

B. Power Supply

The transducer's power supply supplies electronic components with ±12 V DC. It consists of voltage transformer TR1 (220 V AC, 2x12 V AC, 2 VA), Graetz bridge 1.5A, voltage regulators IC1 (78L12) and IC2 (79L12), and electrolytic capacitors for filtering the supply voltage referred as CF1-CF6 as shown in Fig 5.
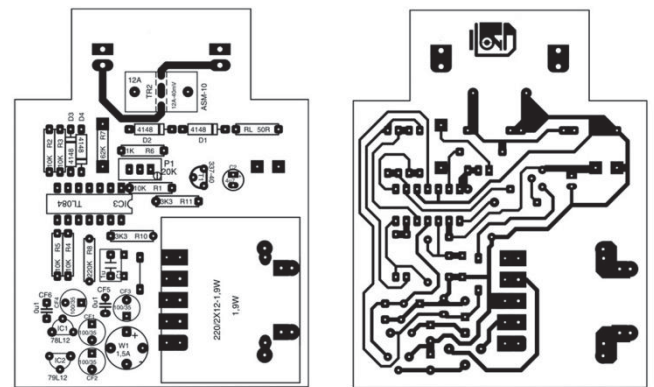
Fig. 3. The transducer's PCB layout

C. Current Transformer/Sensor

A current transformer is an instrument transformer in which the secondary current is substantially proportional to the primary current. Current transformers are switched on to the primary circuit regularly because they need to reduce the current being measured; they work in the short circuit regime. The transducer uses a miniature current sensor type: ASM-010/TALEMA. A typical characteristic of ASM current sensor is shown in Fig 4 [8].
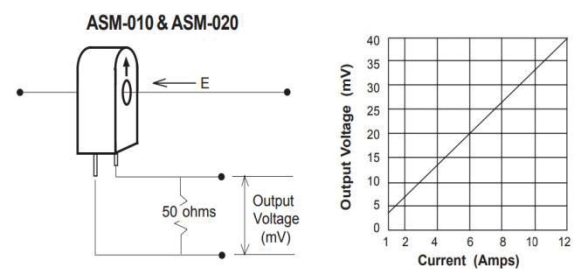
Fig. 4. The ASM-010 current sensor characteristics [8]

D. Inverting Amplifier

Operational amplifier (OP) IC3:A with resistors R6, R7 and potentiometer P1 forms inverting amplifier arrangement as shown in Fig 5. No inverting input of the IC3:A is bound to ground (zero potential). Since the input voltage of an ideal OP is zero, the inverting input of the OP IC3:A is virtually on the potential of the ground. Since the input current of OP is zero, all input current is closed over R7 and P1. The amplification A of this amplifier is negative and equal to:

$$A = -\frac{R7+P1}{R6} \qquad (1)$$

E. Precise Double-sided Rectifier

OP amplifiers IC3:B and IC3:C, diodes D3-D4 and resistors R1-R5 form a precise double-sided rectifier as shown in Fig 5.
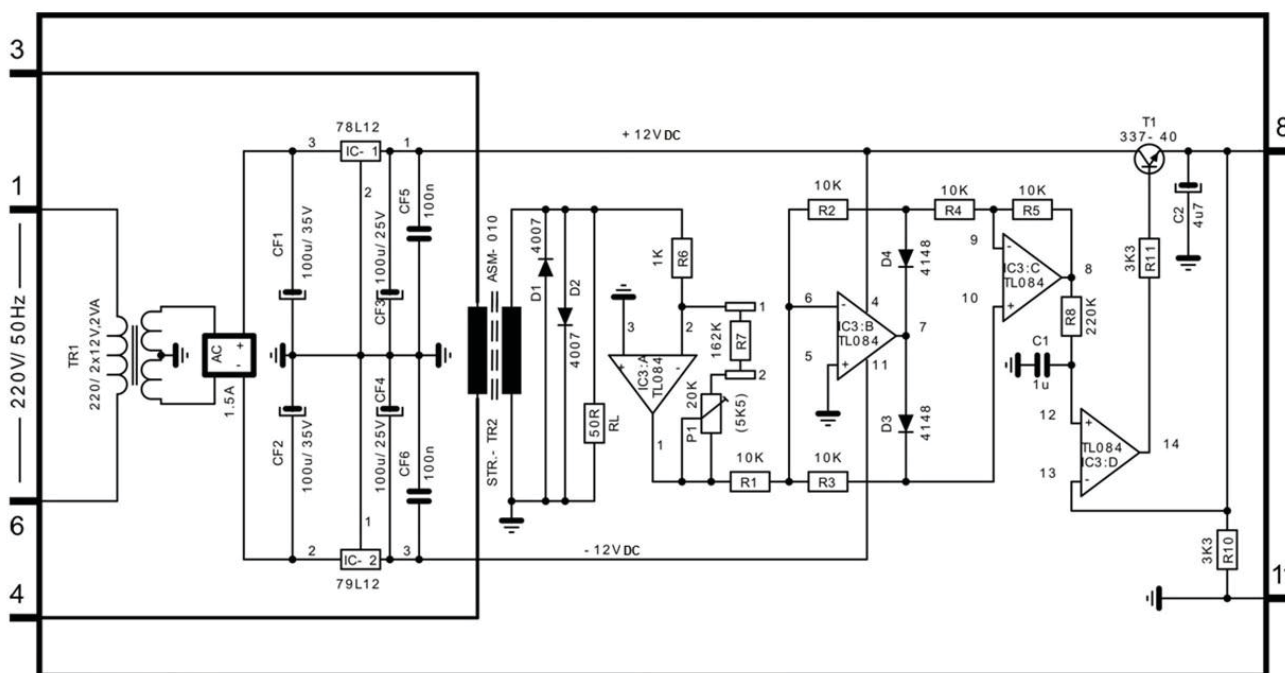
Fig. 5. The electric scheme of the AC current transducer

When the input voltage is positive, the diode D4 leads. On its anode, the voltage is equal to the input one, but with the opposite polarity. In the second OP (IC3:C), this voltage is only inverted and a positive voltage with a single amplification is obtained at the output of the rectifier. Both OP amplifiers work as inverting amplifiers with unity gain. When the input voltage is negative the feedback circuit of the first OP (IC3:B) is realized via diode D3. The input current is negative and it is divided in the ratio of 1:2, with one third going upstream, and over R2 and R4 entering the second OP, and two-thirds closing via D3. The gain is, therefore, ±1 depending on the polarity. The output voltage is not affected by voltage drops on the diode; hence, the circuit acts as a precision rectifier (absolute value detector). The circuit is simple because resistors of equal resistance are used. The disadvantages of the circuit are a unit gain and a small input resistance [9].

F. Output Stage

OP amplifier IC3:D, transistor T1 (BC337), resistors R8, R10, R11 and capacitors C1 and C2 form the voltage controlled current source, as shown in Fig 5. The OP amplifier IC3:D acts as a voltage comparator. No inverting input of OP IC3:D is connected to the output of precise double-sided rectifier through the integrator, formed by the R8 resistor and the capacitor C2. The output stage provides greater current availability due to the use of transistor at the output instead of using the direct output from OP. The OP IC3:D and transistor T1 ensure that the voltage over resistor R10 is kept equal to the one at the non-inverting input of OP IC3:D. The voltage on the resistor R10 is proportional to the AC current at the transducer input. Adjustment of the transmission characteristics of the transducer is done using the potentiometer P1.

III. TESTING THE CHARACTERISTICS OF THE AC CURRENT TRANSDUCER

Adjusting the I/O characteristics of the transducer (slope of the working curve) was done using the potentiometer P1 (shown in Fig 5.). Recording of the I/O characteristics was done by using the PHILIPS FLUKE PM2525 [10]. The measurement of DC voltage in the range 0-20 V with these instruments has an accuracy class of 0.02. The I/O characteristic of the transducer when it uses the current transformer CT010P-A/B/C [11] is shown in Fig 6.
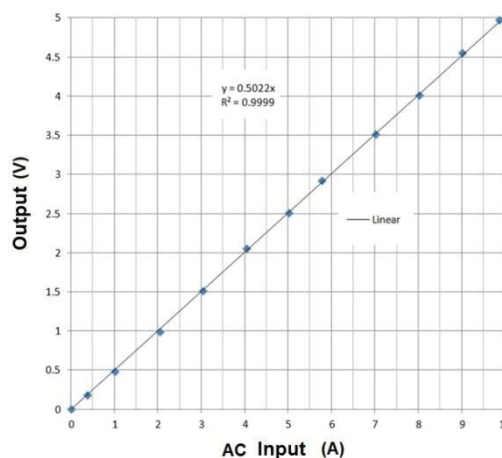
Fig. 6. The I/O characteristic of the transducer

Simulation of DC, AC and transient working regime of the transducer was carried out using the TINA-TI V9 program [12]. The response time of the transducer is approximately 1s, as shown in Fig 7. The transducer's response time can be adjusted as needed by choosing the appropriate value of capacitor C1 (shown in Fig 5.).
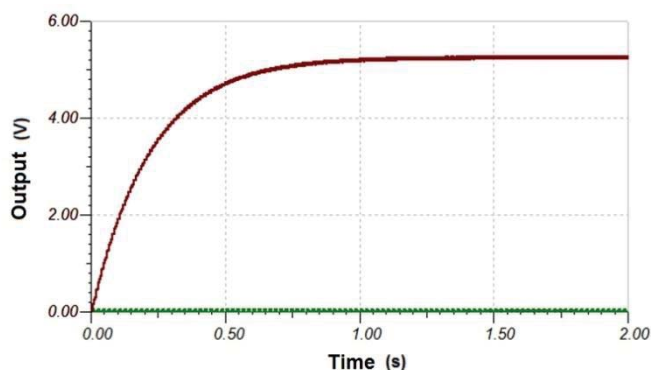
Fig. 7. Transducer response time for the maximum current on the input

IV. THE EXAMPLE OF APPLICATION

A circuit diagram of the electric motor overcurrent load protection system is shown in Fig 8. The protective element compares the preset maximal current load with the actual current load measured by the transducer. When the actual current exceeds the preset maximum, the protection system activates the sound and the visual alarm. Hence, the information about the actual electrical energy consumption of the electric motor is forwarded from the Arduino board [13], that serves as DCS node, via the wireless network to a DCS control level (PC workstation).
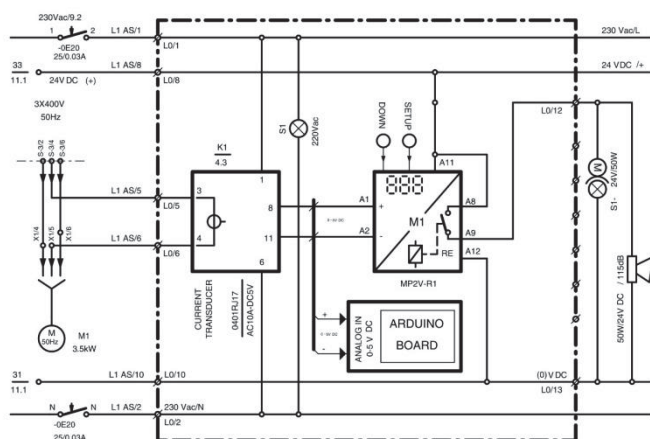
Fig. 8. The electric motor overcurrent load protection system [14]

V. CONCLUSIONS

In this paper, the main characteristics of the recently realized AC current transducer are presented. The application of such devices enables measurement of the electricity consumption of a wide variety of electrical consumers in the industrial facilities, as well as in the households. The realized device is easy to manufacture, install, calibrate and maintain. In the practical application in overcurrent protection systems, it has shown good stability and reliability. We expect that, in the near future, it will be extensively applied in the electrical energy consumption control systems.

ACKNOWLEDGEMENT

REFERENCES

[1] H.N.Norton, *Handbook of Transducers*, Prentice Hall, Inc. Englewood Cliffs, NJ 07632, 1998.
[2] G.C.Barney, *Intelligent Instrumentation*, Prentice Hall International, 66 Wood Lane End, Hemel Hempstead, HP2 4RG (UK), 1988.
[3] http://www.crmagnetics.com/Assets/ProductPDFs/CR4100_Series.pdf (accessed 15.03.2019)
[4] https://www.lem.com/sites/default/files/products_datasheets/ap-b10.pdf (accessed 15.03.2019)
[5] https://cdn.automationdirect.com/static/specs/acuampact.pdf (accessed 15.03.2019)
[6] http://www.adeptfluidyne.com/product/ac-current-transducer/ (accessed 15.03.2019)
[7] http://www.minelautomatika.rs/pdf/06/minel_automatika_merpretvaraci.pdf (accessed 15.03.2019)
[8] https://www.gotronic.fr/pj-39.pdf (accessed 15.03.2019)
[9] https://www.researchgate.net/profile/Radojle_Radetic2/publication/313853383_Operacioni_pojacavaci_sa_zbirkom_sema_-_odabrana_poglavlja/links/58ab4422a6fdcc0e079bb850/Operacioni-pojacavaci-sa-zbirkom-sema-odabrana-poglavlja.pdf (in Serbian, accessed 18.03.2019)
[10] https://doc.xdevs.com/doc/PHILIPS/PHILIPS%20PM%202525%20Operation.pdf (accessed 15.03.2019)
[11] http://cxem.net/izmer/files/izmer183_current_trans.pdf (accessed 18.03.2019)
[12] https://www.tina.com/ (accessed 15.03.2019)
[13] https://www.arduino.cc/en/Main/Software
[14] Tasić V., Pavlov-Kagadejev M., Jeremijić R., Despotović V., Tasić O., and Stojković I., "The Elements of Low-Cost DCS for Electricity Consumption Control," ICEST 2018, Conference Proceedings, pp. 303-307, Sozopol, Bulgaria, 2018.

Performance Assessment of IEEE 802.11a 54 Mbps WPA Laboratory Links

José A. R. Pacheco de Carvalho[1],

Cláudia F. F. P. Ribeiro Pacheco[2], Hugo Veiga[3]

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 WPA 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 WPA 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 enabled 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 and ac standards [1]. Nominal transfer rates up to 11 (802.11b), 54 Mbps (802.11 a, g), 600 Mbps (802.11n) and 3.5 Gbps (802.11ac) 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 led to substantial 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.

Studies are published 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 Open and 54 Mbps 802.11a WEP links [15],[16].

The motivation of this work is to evaluate and compare performance in laboratory measurements of WPA 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 WPA 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

[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.

TCP connections) versus TCP packet size, and UDP jitter and percentage datagram loss (from UDP communications) versus UDP datagram size.

The remaining of the paper is organized as follows: Section II treats 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

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 chosen 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/ac adapter, running Acrylic WiFi software [20]. WPA encryption was activated in the AP and the wireless adapters of the PCs, with a key composed of twenty six hexadecimal characters. The experiments were made under far-field conditions. 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, in the range from 0.25k to 64k bytes.

The Wi-Fi network was as follows. 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 set 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 IP APTEL (Applied Physics and Telecommunications) Research Group network, via switch.

III. RESULTS AND DISCUSSION

The wireless network adapters of the PCs were manually configured for a nominal rate of 54 Mbps. WPA 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 WPA 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 WPA 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 23.5%. 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 axis in Fig. 5 has a log 10 scale. It was found that, on average, jitter performances are not significantly diferent for both link types. 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, the best performance was found for PTP links (Table I). 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 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 for maintaining 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, TCP throughput did not show significantly sensitive to WPA within the experimental error. In passing from Open to WPA links, where data length increases due to encryption, jitter and percentage datagram loss have shown visible 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.

TABLE I
AVERAGE WI-FI (IEEE 802.11A) RESULTS
WPA PTP AND 4N-PTMP LINKS

Parameter/Link type	PTP	4N-PTMP
TCP throughput (Mbps)	23.0 +- 0.7	5.4 +- 0.2
UDP-jitter (ms)	4.0 +- 0.5	3.7 +- 0.8
UDP-% datagram loss	2.5 +- 0.3	6.0 +- 0.5
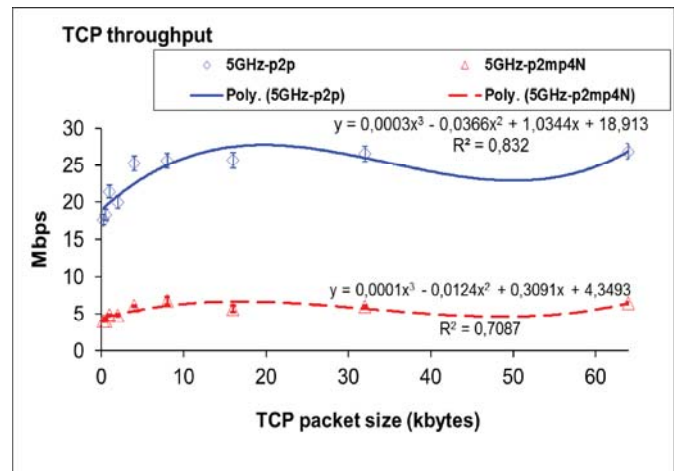
Fig. 3. TCP throughput (y) versus TCP packet size (x). WPA links.

Fig. 1. Wi-Fi laboratory arrangement.
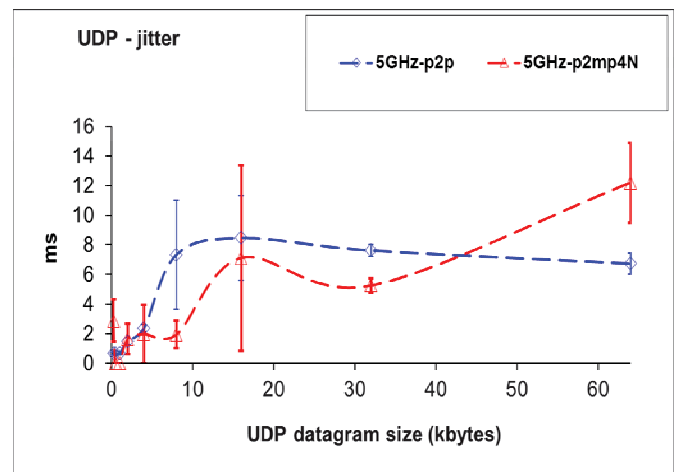
Fig. 4. UDP - jitter versus UDP datagram size. WPA links.

Fig. 2. Typical SNR (dB) and N (dBm).
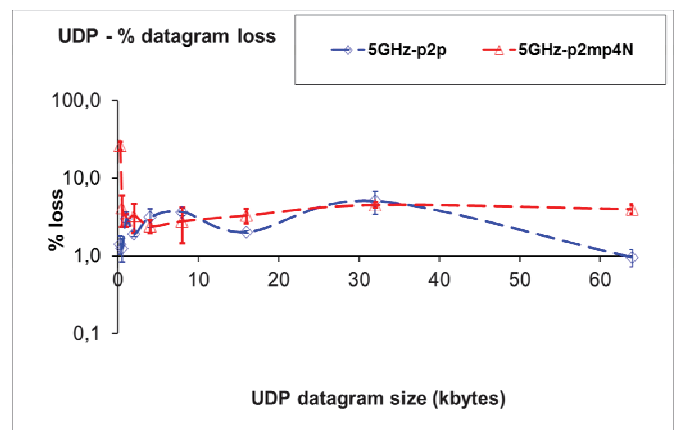
Fig. 5. UDP – percentage datagram loss versus UDP datagram size. WPA links.

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 WPA 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 WPA PTP and 4N-PTMP links. TCP throughput was found to increase with packet size. As for jitter, for small sized datagrams, it was 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 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 WPA. 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

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

[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.

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.

[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, "Performance Assessment of IEEE 802.11a 54 Mbps WEP Laboratory Links", Proc. ICEST 2018 – LIII International Scientific Conference on Information, Communication and Energy Systems and Technologies, P.39, pp. 247-250, Sozopol, Bulgaria, 28-30 June, 2018.

[17] Web site http://www.hp.com; HP V-M200 802.11n access point management and configuration guide; 2010; accessed 10 Jan 2019.

[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 2019.

[21] Iperf software; 2016; http://iperf.fr; accessed 16 Feb 2019.

[22] Network Working Group. "RFC 1889-RTP: A Transport Protocol for Real Time Applications", http://www.rfc-archive.org; 1996; accessed 13 Apr 2019.

[23] P. R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, Mc Graw-Hill Book Company, 1969.

Impact of noise in printing process

Svetlana Mijakovska[1], Roberto Pasic[1], Ivo Kuzmanov[1], Filip Popovski[1]

Abstract – In this paper are given the impact of noise in printing process. Each printing process tends to reduce the disadvantage of the core material. The basic material in the printing houses is paper. Microclimate has a strong influence on the the paper disadvantage. Also noise is factor that has great influence in printing process. We made experimental measurements on two kind of printing machine in order to get conclusions about controlling then noise in printing processes.

Keywords – Printing, Paper damage, Microclimate, Noise.

I. INTRODUCTION

During the printing process, the impact of the microclimate is importance for the reduction of the paper damage. In the microclimate we will look at internal and external temperature, brightness and humidity [1].

Traditional printing processes can be inherently very noisy [2]. Although modernization has introduced quieter processes into the industry, high noise levels and noise exposures remain a health risk; 93% of print workers assessed in this study had noise exposure estimates above the 80 dB.

Another factor that has great impact in printing process is noise. The noise is a mixture of various sounds with different number of flashes at a certain time and can be defined as an occurrence that causes negative environmental impacts. Exposure to a person under the influence of sound presents a danger to his health, both from the sounds of less intensity than from the bullies. The research showed that the noise in the man causes the trembling of the bloodstream of the middle ear, the function of the official organs decreases, with the partially or periodically shifting of the probability of probability that can lead to full-blown. The disease negatively affects both the nervous and the cardiovascular system (the blood pressure rises, it reduces the work of the heart, the lying of the pulses and the elderly creases, destroys the work of the nasal cavity, and to nonsense and tediousness and the like).

The arousal of a man can also affect the psychological state, a neurosis occurs, depression, aggression prevents him in the artificial formation, interrupts the rest - it disturbs the son and the like. Therefore, the company should be more concerned with the exposure of noise to employees to improve their work obligations. The highest source of noise in the furnaces is the machines with mechanical parts which, when rotating, discharge various sounds, compressors, motors, the movement of paper through the machine, the air conditioning systems and ventilation and the like.

[1]Svetlana Mijakovska, Roberto Pasic, Ivo Kuzmanov and Filip Popovski are with the Faculty of Technical sciences at University of Bitola, st. Makedonska Falanga 33, Bitola 7000, Macedonia.

Table 1. Marginal time of exposure in relation to the level of noise

Daily exposure in hours	Noise level in (db)
8	90
6	92
4	95
3	97
2	100
1.5	102
1	105
1/2	110
1/4	115

II. NOISE IN PRINTING PROCESS

In the printing houses there are many sources of noise from the installation of the machines and from the ventilation and air conditioning systems. We will do the measurements and will ask for the answer as to how the noise works in increasing the quality of printing process [2].

The measurements are made with a digital measuring sound instrument Digital Sound Level Meter IEC651, the principle of which is based on OM meter and measurement resistance. It receives the volume of the sound through a graphite microphone whose resistance is changed in the presence of sound or higher sound exposure contributes to the convergence of the graphite beads into the microphone that gives greater resistance, which then turns out to be digital display.

Specifications;
- measuring range (30 - 130) db
- power supply 9 batteries
- working temperature (0 - 40) ℃
- the ability to display the maximum value
- measure two times a second
- dimensions (57 x 26 x 149) mm
- the weight is 144 grams

Machines on which the measurement is carried out are:
- Four-shot offset machine KBA Rapida 72 [5]

Figure 1. Four-shot offset machine KBA Rapida 72

- Unique offset machine Adast dominant 816 [5]

Figure 2. Unique offset machine Adast dominant 816

III. RESULTS OF MEASUREMENTS ON FOUR-SHOT OFFSET MACHINE KBA RAPIDA 72

The measurements are made on four-shot offset machine KBA Rapida 72 for lowest and largets value of noise. The results gives answer how the noise works in increasing the card.

Table 2. Measured noise is consistent with the card at KBA Rapida 72(lowest values)

Noise level in (db)	paper damage (%)
80.5	1.78
80.7	0.8
80.7	0.93
80.8	0.73
81	0.279
81.05	2.14
81.06	0.22
81.1	1.45
81.13	0.67
81.2	0.58
81.22	0.83
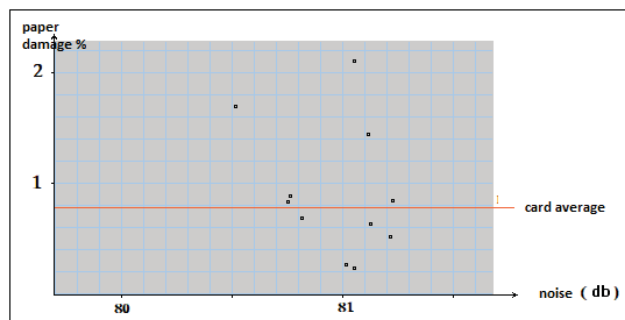
Figure 3. The graphic representation of the relation of noise from noise was the lowest value

From the measured values presented on the graph, it can be concluded that the increased measured points are located around the average line up and down and has values that are higher than the average and lower below the average, so they can not to bring to some conclusion that the lower humiliation should result in a palliation, even though we expect it.

Table 3. Measured noise is consistent with the card at KBA Rapida 72(largest values)

Noise level in (db)	paper damage (%)
83.02	0.88
83.06	0.35
83.06	0.38
83.08	0.65
83.17	0.42
83.38	0.6
83.46	0.51
83.5	0.189
83.68	0.33
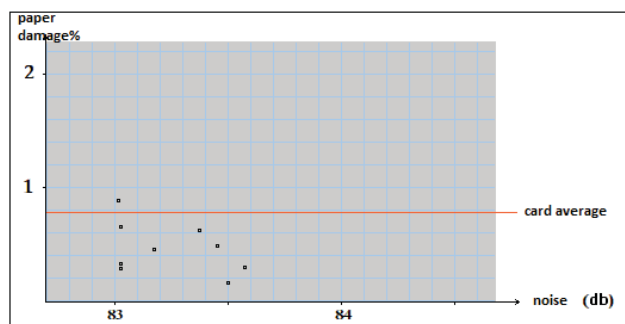
Figure 4. The graphic representation of the relation of noise from noise was the largest value

Table 3 shows the measured points at the highest noise level and the card moves below the average card, which in one way is contradictory, since we do not have high noise when waiting for a high card this is not confirmed in this case. Comparing the figure 3 and figure 4, we cannot conclude that the high noise does not carry high cards, perhaps because the noise in the magazine and in the smaller ones and in the larger values still go within the permitted limits for an average working time of 8 hours as shown in Table 1.

IV. RESULTS OF MEASUREMENTS ON UNIQUE OFFSET MACHINE ADAST DOMINANT 816

The measurements are made on unique offset machine Adast dominant 816 for lowest and largets value of noise.

Table 4. Measured noise is consistent with the card at Adast dominant 816(lowest values)

Noise level in (db)	paper damage (%)
79.08	2.96
79.86	10.28
80.02	5.6
80.1	4.07
80.15	3.6
80.38	3.06
81.37	7.2
81.45	4.06
81.55	4.31
81.57	7.75
81.67	4.05

Table 5. Measured noise is consistent with the card at Adast dominant 816(largest values)

Noise level in (db)	paper damage (%)
82.01	4.98
82.06	6.19
82.17	2.82
82.3	1.57
82.31	6.57
82.37	4.68
82.4	1.6
82.425	6.02
82.77	4.07
82.81	4.73
82.9	5.14
83.16	4.95
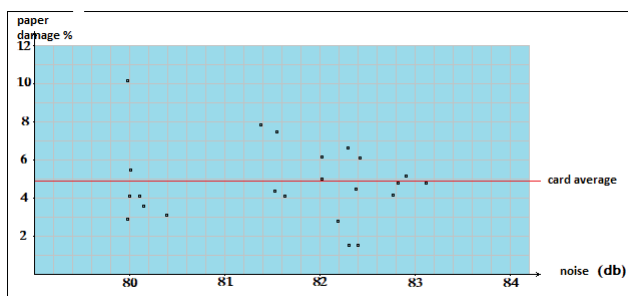
Figure 5. The graphic representation of the relation of noise from noise on Adast dominant 816

From the measurement of the noise of the Adast Dominant 816 in figure 5, we cannot conclude that the card is smaller in smaller noise than it is larger because the points in the measurements are found, due to the unexpected positions may be due to the lower noise

values and the increased permissibility within the permitted limits for an average working time of 8 hours shown in Table 1.

V. USE OF HEARING PROTECTION

The printing industry is a high noise industry and it is likely that hearing protection will be needed [6]. At more modern printing houses, where technical and organisational means of reducing noise were already in place, hearing protection use was only required for activities identified as noisy, meaning it was only used when necessary. At printing houses where older machinery and technology was in use, hearing protection was relied on as the primary control measure. Using hearing protection in this way is not acceptable as a long-term solution and must not be used an alternative to controlling noise by technical and organisational means. For hearing protection to be effective it must be worn all of the time when in the noisy environment. At some printing houses it was evident that routine use of hearing protection was an accepted and integral part of the safety culture. At other printing houses, it was less clear if the safety culture supported and managed the use of hearing protection or if it was left to workers to decide if they wanted or needed hearing protection.

VI. CONCLUSION

Noise is one of factor that has a huge impact in printing processes. In both cases on different printing machines we cannot define the lowest values of noise. But we can recommended usage of hearing protection. The inherently noisy nature of the industry, even in more modern print works where quieter machinery is used, means that there is likely to be an ongoing requirement for the use of hearing protection. Although hearing protection was observed to be widely provided and used, failure to correctly fit plug-type protection was commonly observed.

REFERENCES

[1]. Svetlana Mijakovska, Roberto Pasic, Ivo Kuzmaniv, Filip Popovski, Borce Jakimovski *"Impact of the microclimate in printing process"*,; International Journal of Advances in Engineering & Technology, Volume 12, Issue 1, pages 11-17, ISSN 22311963, February 2019.

[2]. R. Pasic; I. Kuzmanov; S. Mijakovska*;"Print Quality Control Management for Papers Containing Optical Brightening Agents*; International Journal of Scientific and Engineering Research, Volume 7, Issue 2, Feb 2016, pages 271-274, ISSN 2229-5518

[3]. Control of noise risk in the printing industry,Prepared by the Health and Safety Executive, 2017.

[4]. Emma Shaks, *Noise in the United Kingdom printing industry: then and now*, Health & Safety Laboratory, United Kingdom

[5] https://www.machineryeurope.com/machine/

[6] P.Brereton, J.Patel,"*Noise risk as described in instructions for printing machinery supplied in Europe*", Euronoise 2015.

ENGINEERING EDUCATION & OTHER

Teaching Digital Filters using NI LabVIEW and USRP

Aleksandar Atanasković, Biljana Stošić, and Nataša Maleš-Ilić

Abstract – **Computer based learning has become an important part of education systems. It provides animation and interactive processes that are not possible with textbooks. Filtering of digital signals is a fundamental concept in digital signal processing. In this paper, a framework for application of NI LabVIEW and USRP in computer based teaching and learning, with a few examples in the domain of digital filtering is described. Real-time communication system demonstrations for the classroom are discussed.**

Keywords – **Computer-based learning, NI LabVIEW, NI USRP, Digital filters, Real-time visualization.**

I. INTRODUCTION

One of the most important parts of engineering education is the application of theoretical knowledge in practice, which can be realized in different ways. Standard experimental set-up for teaching software defined radio (SDR) and related communication system/signal processing topics is expensive, takes a significant amount of time for development, etc. Because of that, the universal software radio peripheral (USRP) based system can be great solution for different topics in real-time.

Digital communications is taught in most electronic degree curricula worldwide, where the main focus is normally the theory of communications and understanding of different building blocks of a communication system. Learning theory is essential for students, but it is not enough to prepare them for marketplace in industry. Hence, this paper introduces the application of Laboratory Virtual Instrument Engineering Workbench (LabVIEW) virtual instruments in electrical engineering education. After explaining the advantages of using LabVIEW [1-2], a few selected examples, typical for digital signal processing [3-4], will be discussed and illustrated. Highlights include the ability to work with signals and different filters to facilitate real-world filter testing.

The USRP family of products has become a popular platform for hardware-based research and test bed validations conducted by universities in SDR field [2], [5-7].

The LabVIEW/USRP combination presents an opportunity to enhance communications education by enabling a low cost, hands-on approach with live signals for realistic, real-world demonstrations, laboratory exercises, capstone design projects, and cutting-edge research.

Among the many software products, LabVIEW as one of the leading products of National Instruments (NI) is the most worthy to solve engineering problems. LabVIEW virtual instrument concept allows to create complex system of information processing which using the isolated blocks, some

The authors are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mails: aleksandar.atanaskovic@elfak.ni.ac.rs, biljana.stosic@elfak.ni.ac.rs, natasa.males.ilic@elfak.ni.ac.rs

unique subprogrammes, virtual instruments (VIs), working with the flow of information. With educational and practical point of view, the main advantages of LabVIEW are the possibility of combining the computer simulation technology, virtual management, and natural data collection. All of these possibilities provide a good reason for inclusion of LabVIEW to student training of the natural sciences and technical specialties.

The rest of the paper is structured as follows. URSP and LabVIEW are shortly introduced in Sections II and III, respectively. Filtering of digital signals by using LabVIEW/USRP combination is presented in Section IV along with the discussion. The paper is concluded in Section V.

II. USRP INTRODUCTION

The USRP [2], [5-7] is a SDR reconfigurable radio frequency (RF) hardware designed to build and test digital communication systems. The USRP is commonly used by researchers as a wireless prototyping platform and by universities as a teaching aid. With support from NI LabVIEW, which is a system-design platform and development environment, real-time processing is made possible. Hence, the availability of both USRP and LabVIEW has benefited various researchers and educators.

Generally speaking in an academic setting, USRP's can be used in courses, laboratories, and researches involving cognitive radio, digital signal processing (DSP), multi-rate DSP, communication systems, Field-Programmable Gate Array (FPGA) system design and implementation, a wide variety of other areas (e.g. homeland security applications, satellite systems, radio astronomy, wildlife tracking, radio frequency identification, medical imaging, sonar systems, and customizable test equipment development).
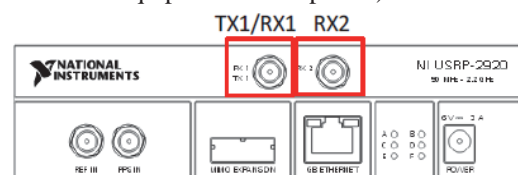
Fig. 1. USRP front panel
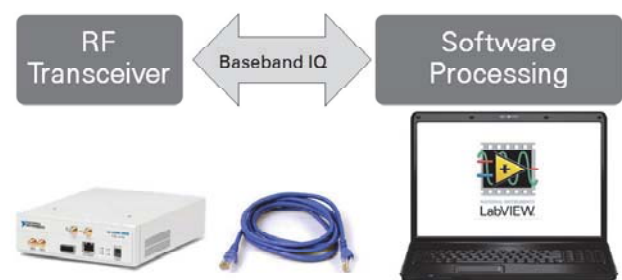
Fig. 2. Simplified overview of a SDR setup built around a NI USRP [2]

USRP front panel is shown in Fig. 1. The NI USRP connected to a host PC acts as a software-defined radio. Simplified overview of a SDR setup built around a NI USRP is depicted in Fig. 2 [2].

III. LabView Introduction

LabVIEW [1-2] is developed by Jeff Kodosky of NI Cooperation in the mid year of 1986. Among other advantages, the three reasons listed below are the most important ones for applying LabVIEW: - LabVIEW is a powerful, platform-independent, graphical programming development system which is ideally suited for data acquisition, storage, analysis, and presentation; - LabVIEW is a programming environment which is widely used for measurement and automation; - LabVIEW helps to blend existing educational hardware inventory with virtual instrumentation in an economical way and with reasonable funds.

LabVlEW is used to teach digital signal processing. The graphical programming approach is based on data flow theory which is an ideal platform for learning how signals flow from one function to another such as from an acquisition function through a filter to a spectrum analysis and finally to a graph. Each function is an icon that is wired to other icons. The wires are the signals flowing from one icon to the next.

The basic building block of LabVIEW is the virtual instrument (VI). Conceptually, a VI is analogous to a procedure or function in conventional programming languages. Each VI consists of a block diagram and a front panel. The block diagram describes the functionality of the VI, while the front panel is a top level interface to the VI.

The NI LabVIEW Digital Filter Design Toolkit (DFDT) is complete filter design and analysis software that one can use to design filters to meet required specifications. With the DFDT, users work within the LabVIEW development environment to design, analyze, and implement a variety of IIR and FIR filters. This work has a number of advantages, such as the ability to do filter testing with simulated signals (generated algorithmically) or live signals. One can view the time waveforms or the spectra of both the input signal and the filtered output signal to show how the current design performs on real-world signals.
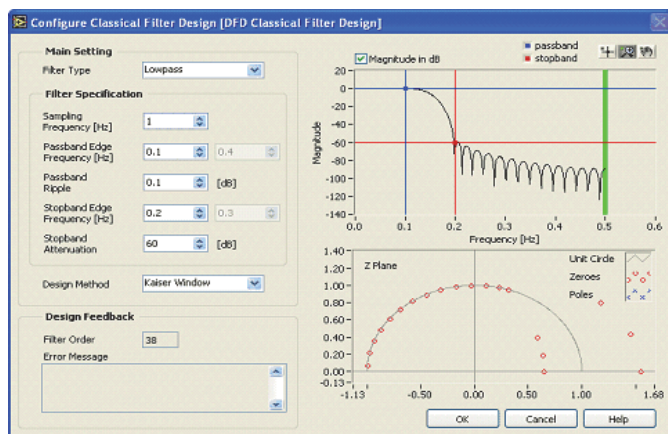
Fig. 3 presents a classical design express VI of DFDT which allows users to specify filters by typing in passband/stopband frequencies and other parameters. Results of performed design are shown immediately as magnitude response and pole-zero plots.

IV. Digital Filter Examples

In order to present concept of digital filtering for the purpose of a computer based tutoring/learning system that uses LabVIEW and USRP, a system consisting of a transmitting and receiving parts has been realized. For correct operation of the system, it is necessary to properly set-up the USRP parameters (device name, carrier frequency, IQ rate, gain, active antenna) for both the transmitting and the receiving part. The transmitter is used to generate a multitone signal; one has to set the start frequency, the frequency spacing between the tones and the number of tones. The signal is transmitted as AM (Amplitude Modulation) signal with a carrier frequency corresponding to the carrier frequency of the USRP. In the receiving part, AM signal demodulation is performed in order to obtain original multitone signal generated in the transmitter, and then digital filter is used to observe filter's influence on the signal. One can select lowpass, highpass, bandpass or bandstop filter with Chebyshev or Butterworth approximation. The filter order as well as the cut-off frequency can be adjusted. In the receiving part, signal can be viewed in time and frequency domains. In order to observe the effect of signal filtering in time domain, it is recommended to use multitone signal generated with a lower number of tones. To observe the effect of filtering in frequency domain, signal with a higher number of tones with close frequencies can be used. In that case, a study of filter transfer function can be done.

This system can serve as an excellent complement to classroom teaching and learning. One can apply different filter types to test different input signals. A set-up presented here uses a signal composed of three sinusoidal components (such as one generated of three tones with frequencies 1000 Hz, 2000 Hz and 3000 Hz), as shown in Fig. 4.

Fig. 4. Front panel of transmitter (Signal generated with three tones on 1000 Hz, 2000 Hz and 3000 Hz)

Fig. 3. The Classical Design Express VI of DFDT

The data passes through the currently designed filter, and analysis such as plots of the input and output waveforms and spectra can be generated further.

To interactively specify the desired filter, the parameters can be entered directly on user interface on receiver. The design method is set to Chebyshev approximation. Once the filter is designed, several alternatives for testing the design are possible. One can now select different filter types and get an immediate feel of the power spectrum of the signal at the receiver point, as well as the influence of the filter cut-off frequencies. Figs. 5-10 show graphs of the results of such analysis, i.e. the action of the filter on a test signal. In these cases, filter type is changed and cut-off frequencies are interactively varied for selected filter type.

Figs. 11 and 12 show block diagrams of the receiver and transmitter behind the VI [8].

Fig. 5. Front panel of receiver for case of lowpass Chebyshev filter with 1500 Hz cut-off frequency

Fig. 6. Front panel of receiver for case of lowpass Chebyshev filter with 2500 Hz cut-off frequency
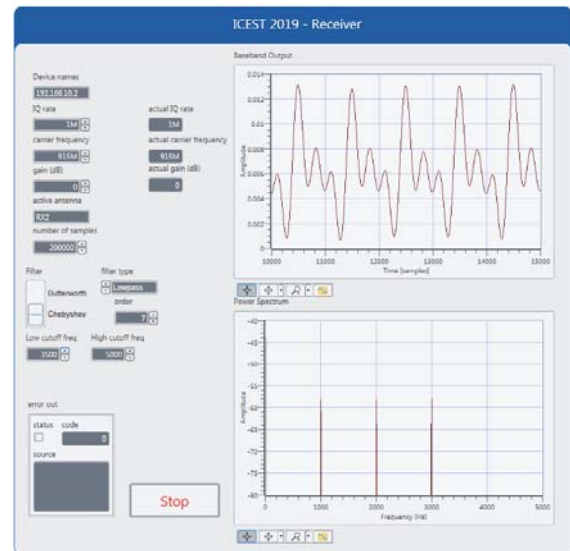
Fig. 7. Front panel of receiver for case of lowpass Chebyshev filter with 3500 Hz cut-off frequency
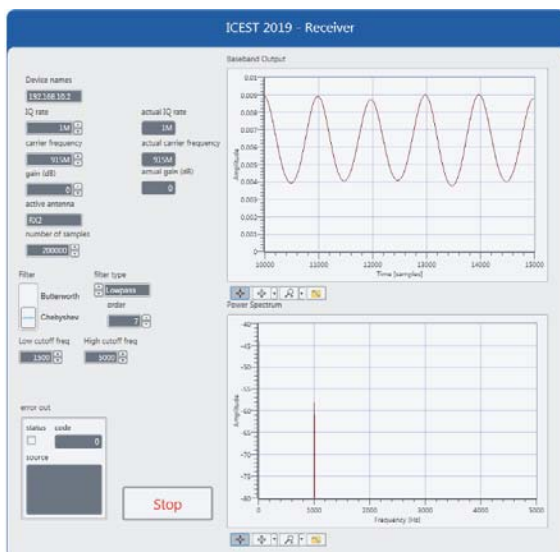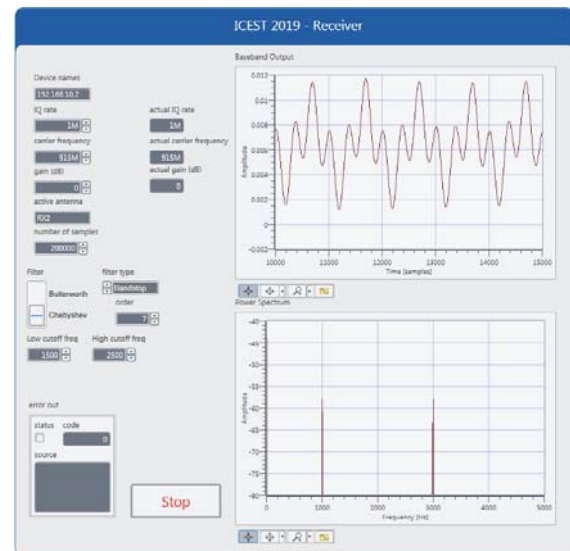
Fig. 8. Front panel of receiver for case of bandstop Chebyshev filter with 1500 Hz and 2500 Hz cut-off frequencies
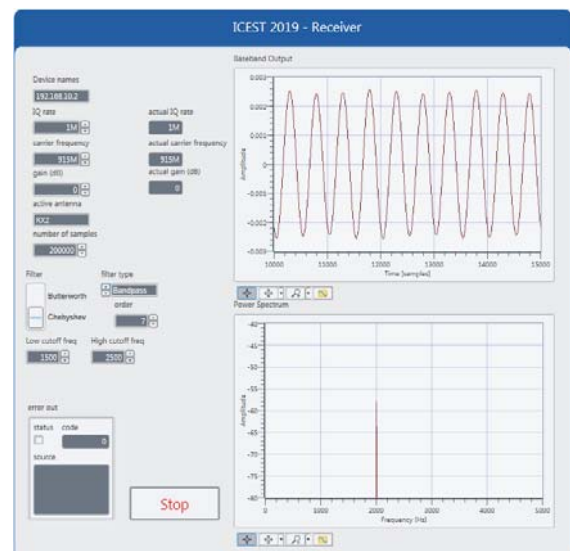
Fig. 9. Front panel of receiver for case of bandpass Chebyshev filter with 1500 Hz and 2500 Hz cut-off frequencies
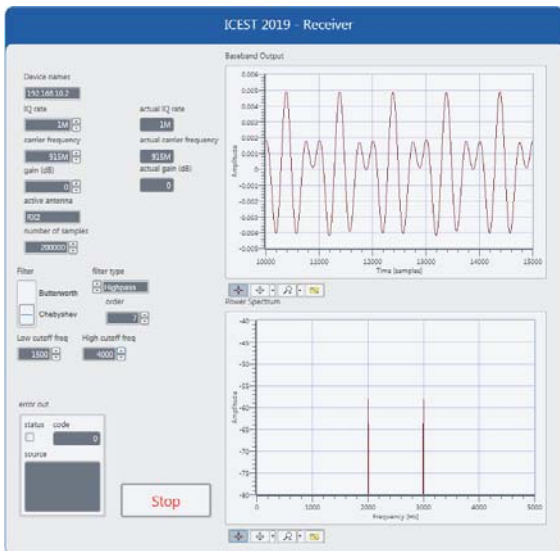
Fig. 10. Front panel of receiver for case of highpass Chebyshev filter with 1500 Hz cut-off frequency

V. CONCLUSION

One of the major requirements in engineering education is to provide students with convenient environments to practice what they have been taught or what they have learned conceptually.

In this article, it is presented how the LabVIEW and USRP can be used to give students hands-on experiences and teach them how to link theory with practical implementation instead of only using theoretical analysis and simulations. Additionally, it is important to teach students the importance of testing their VIs as they are building their systems instead of building the whole system and then finding that it does not work and is very difficult to debug.

The educational benefits, i.e. learning outcomes of this application are the following: design and implementation of digital filters and understanding VI concepts.

ACKNOWLEDGEMENT

REFERENCES

[1] National Instruments, *LabVIEW Digital Filter Design Toolkit User Manual*, February 2005.
[2] National Instruments, *Une introduction à la radio définie par logiciel avec NI LabVIEW et NI USRP*, Version 1.1 – Q4 2013.
[3] N. Kehtarnavay and N. Kim, *Digital Signal Processing System-Level Design Using LabVIEW*, Elsevier, 2005.
[4] S.K. Mitra and J. Kaiser, *Handbook for Digital Signal Processing*, 1993 John Wiley and Sons, Inc.
[5] A. Abrol and E. Hamke, "Introduction to Communication Systems Using National Instruments Universal Software Radio Peripheral - Lab Manual", The University of New Mexico, September 11, 2014.
[6] T.B. Welch and S. Shearman, "AC 2011-2086: LabVIEW, the USRP, and their Implications on Software Defined Radio", American Society for Engineering Education, 2011.
[7] T.B. Welch and S. Shearman, "Teaching Software Defined Radio Using the USRP and LabVIEW", 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 25-30 March 2012, pp. 2789-2792.
[8] B.A. Black, "Introduction to Communication Systems - Lab Based Learning with NI USRP and LabVIEW Communications - Student Lab Manual", National Instruments, 2014.
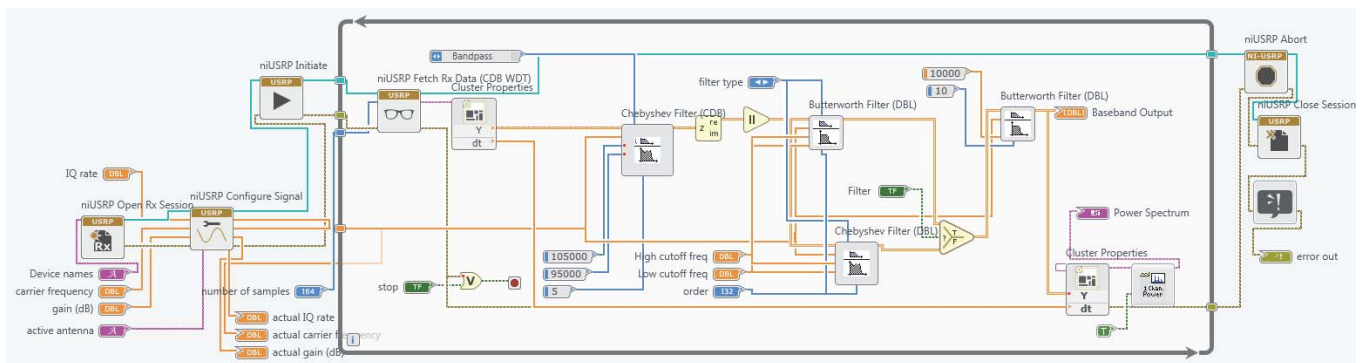
Fig. 11. LabVIEW VI interface of receiver - block diagram of the system behind the VI
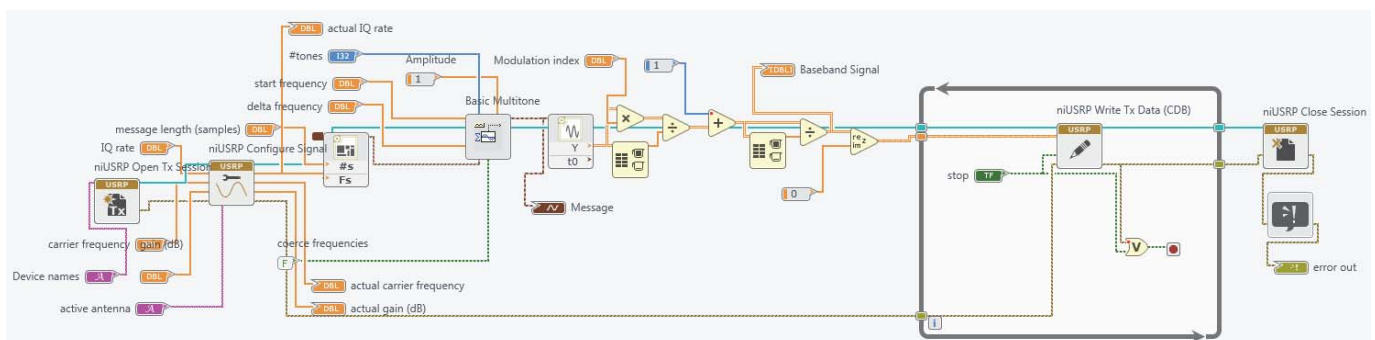
Fig. 12. LabVIEW VI interface of transmitter - block diagram of the system behind the VI

Bi-isotropic Cylinder Placed in Homogeneous Electric Field Generated Using Plan-parallel Electrodes

Žaklina J. Mančić[1] and Zlata Ž. Cvetković[2]

Abstract – **In this paper, the influence of bi-isotropic body on the homogeneity of electric field, generated using plan-parallel cell of the first order (primary cell), is analyzed and a 2D view of its impact on the achieved homogeneity of the field is given.**

Keywords – **electrostatics, homogeneous field, image theorem, bi-isotropic material of Tellegen type.**

I. INTRODUCTION

The problem of generating homogeneous fields is relatively old [1,2] but still attractive research area [3-10]. For homogeneous field generation, various approaches may be used. For instance, rings are used for homogeneous field generation in [1, 3, 4, 6], plan-parallel systems are exploited in [2,5, 8-10] whereas conical electrodes are considered in [4, 7]. Also, the influence of conducting body placed in homogeneous electric field is analyzed in the case of ring electrodes [6], biconical electrodes [7] and plan-parallel electrodes [8, 9]. Complex plan-parallel systems are modelled and the primary *N*-th order cell, whose electrodes are positioned on an imaginary cylindrical surface, was defined in [8]. Furthermore, the stability and efficiency of such systems in the case of small deviations of some geometric parameters is analyzed in [9].

One of the major issues in the analysis of generated homogeneous electric fields is to determine to what extent an external body, placed into the field, affects the achieved homogeneity. There are several types of external bodies that are of interest. Classification is made depending on their material, so that conducting body, bodies made of homogeneous dielectric, anisotropic dielectric and bi-isotropic material of Tellegen type can be considered as the most significant. In the case when plan-parallel electrodes are used for field generation, the influence of all aforementioned types of external bodies is analyzed in [10].

In the recent decades, bi-isotropic materials have gained increasing attention [11-20]. This paper is extension of the research presented in [10] and its aim is to illustrate how a bi-isotropic body influences the generated homogeneous field and, above all, to give a 2D view of the bi-isotropic body's impact on the achieved homogeneity.

[1]Žaklina Mančić is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: zaklina.mancic@elfak.ni.ac.rs.
[2]Zlata Cvetković is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000Niš, Serbia. E-mail: zlata.cvetkovic@elfak.ni.ac.rs.

II. EXTERNAL BODY IN THE HOMOGENEOUS ELECTRIC FIELD

In Fig. 1 is shown a primary cell [10] for generating homogeneous electrostatic field in the case of plan-parallel systems. The cell consists of two pairs of plan-parallel, straight linear lines of a circular cross-section. In the case of the coordinate system used in Fig. 1, symmetry plane is $y = 0$, i.e. $\theta = \dfrac{\pi}{2}$ for the zero potential. It should be noted that conductors are on the same absolute potential. An external cylindrical body of radius a is placed into the origin of coordinate system.
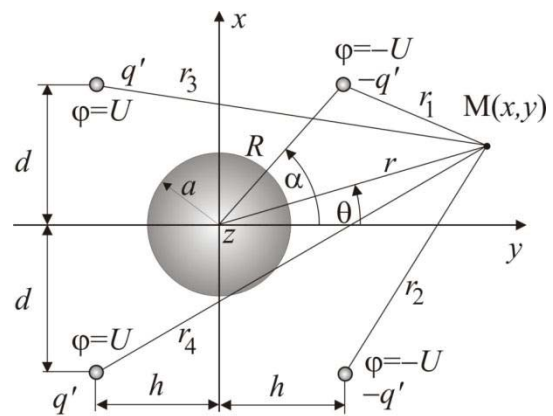
Fig. 1. Primary cell and external body.

Charges densities per unit length q' and $-q'$ are on conductors. Potential in the point $M(x, y)$ of the adopted coordinate system, generated by charges on conductors in the presence of air is:

$$\varphi(\theta, r) = \frac{q'}{2\pi\varepsilon_0} \ln\left(\frac{r_1 r_2}{r_3 r_4}\right) \tag{1}$$

A detailed analysis of plan-parallel systems of *N*-th order is presented in [10] and it is demonstrated that a primary cell of the first order have dimensions $h = 0.5R$ and $d = 0.8660254R$, which we adopt here for the homogeneous field generation. It has been also shown that due to geometry of the system, the usage of cylindrical coordinate system would be an advantage for the purpose of calculating field in the case of cylindrical external body [10].

Image theorem can be used to determine how and to what extent the material from which the body is made, "spoils" the homogeneity of the field. This theorem significantly simplifies calculation process and it provides low complex

solutions in mathematical terms. It has been shown that the images on the basis of which is determined the electric scalar potential for charges per unit length outside the external body are positioned along the cylindrical surface of the radius $D = a^2/R$ in the case of perfectly conducting external body [8]. In previous equation for calculating radius D, parameter a represents the radius of the conducting body, whereas R is the radius of the imaginary cylinder, upon which are placed the plan-parallel electrodes for generating a homogeneous electrostatic field.

III. IMAGE THEOREM FOR BI-ISOTROPIC MATERIALS

Constitutive relations between vectors **E**, **D**, **B** and **H** for bi-isotropic Tellegen medium are defined as [11]:

$$\mathbf{D} = \varepsilon\mathbf{E} + \xi\mathbf{H}, \quad \mathbf{B} = \mu\mathbf{H} + \xi\mathbf{E} \qquad (1)$$

where $\varepsilon = \varepsilon_r\varepsilon_0$ (ε_r is relative permittivity and $\varepsilon_0 = 8.8541878\cdot10^{-12}$ F/m is vacuum permittivity) and $\mu = \mu_r\mu_0$ (μ_r is relative magnetic permeability and magnetic permeability of free space is $\mu_0 = 4\pi\cdot10^{-7}$ H/m).

According to Tellegen, such materials are polarized and magnetized simultaneously if condition $\varepsilon\mu \neq \xi^2$ is satisfied and if they are placed into electric or magnetic field [11-14]. Starting from Maxwell's equations and constitutive relation (1), Poisson's equation for electric scalar potential φ can be obtained:

$$\Delta\varphi = -\rho/\varepsilon_e \qquad (2)$$

where is $\varepsilon_e = \varepsilon\left(1 - \xi^2/(\varepsilon\mu)\right)$.

Poisson's equation for magnetic scalar potential φ_m is:

$$\Delta\varphi_m = \frac{\xi}{\varepsilon_e\mu}\rho. \qquad (3)$$

Once determined φ and φ_m, electric and magnetic field are calculated as:

$$\mathbf{E} = -\operatorname{grad}(\varphi) \text{ and } \mathbf{H} = -\operatorname{grad}(\varphi_m) \qquad (4)$$

whereas remaining fields (**B** and **D**) are calculated from the constituent relation (1).

In the case when the external body is bi-isotropic cylinder of Tellegen type (Fig. 2), unknown constants are [10]:

$$A = -B = \frac{(\varepsilon - \varepsilon_0)(\mu + \mu_0) + \xi^2}{(\varepsilon + \varepsilon_0)(\mu + \mu0) - \xi^2}, \qquad (5)$$
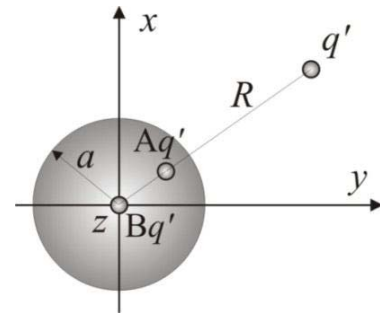
Fig. 2. Charges per unit length and its images in the bi-isotropic cylindrical mirror.

IV. BI-ISOTROPIC EXTERNAL BODY PLACED INTO A HOMOGENEOUS ELECTROSTATIC FIELD

Fig. 3 shows normalized electric field in the case of bi-isotropic external body presence with the weak ($\xi^2/(\varepsilon\mu) = 0.1$) and strong bi-isotropy ($\xi^2/(\varepsilon\mu) = 0.9$). By observing Fig. 3, it can be noticed that bi-isotropic body with a strong bi-isotropy slightly distrupts the homogeneity of the field. Furthermore, it can be seen that bi-isotropic body with a weak bi-isotropy disrupts homogeneity to the larger extent than the body with a strong bi-isotropy, whereas conducting body disrupts the homogeneity of the field most of all. It has been demonstrated in [10] that a bi-isotropic body increasingly disrupts achieved homogeneity by increasing ε_r for a given $\xi^2/(\varepsilon\mu)$, so that for $\varepsilon_r > 40$ the body begins to behave similarly to the conductive body of the same radius a.
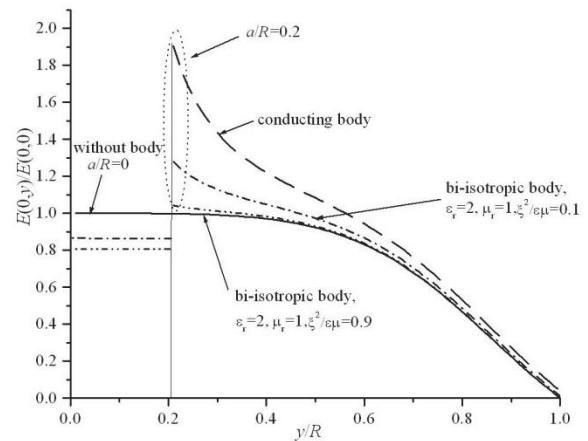
Fig. 3. Normalized electric field along y-axis using one primary cell with and without external bi-isotropic and conducting body,[10].

V. 2D PRESENTATION OF THE UNIFORM ELECTRIC FIELD FOR PRIMARY CELL OF THE FIRST ORDER WITH AND WITHOUT EXTERNAL OBJECTS

In Fig. 4 it is qualitatively presented electric field of the primary cell without external cylindrical body (Fig. 4.a), with external cylindrical conducting body of radius $a = 0.2R$, Fig. 4.b), and bi-isotropic cylinder inserted as an external body, Fig. 4.c) − Fig. 4.f), by using the *Wolfram's Mathematica* ContourPlot function.
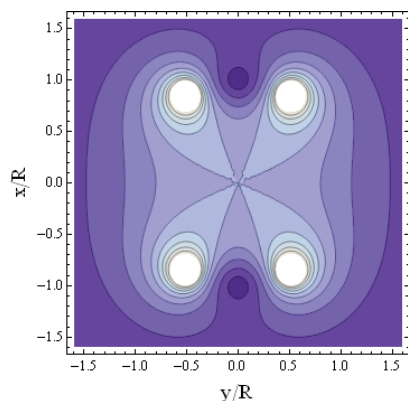
a)   Without external body
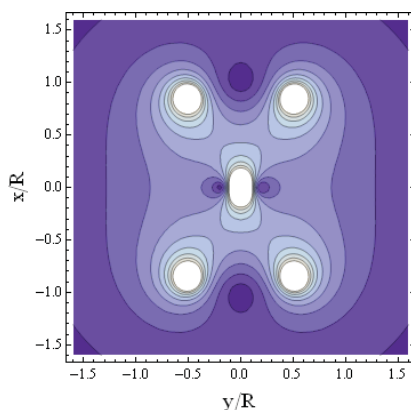
b)   With conducting external body
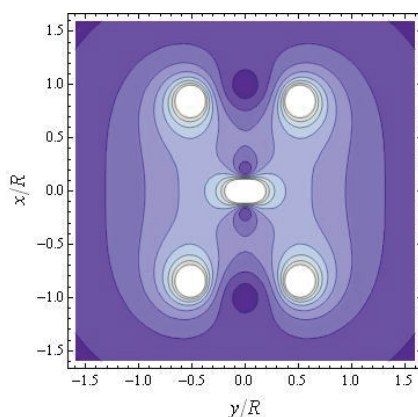
c) Bi-isotropic external body,weak bi-isotropy,

$$\varepsilon_r = 2, \ \mu_r = 1, \xi^2/\varepsilon\mu = 0.1$$
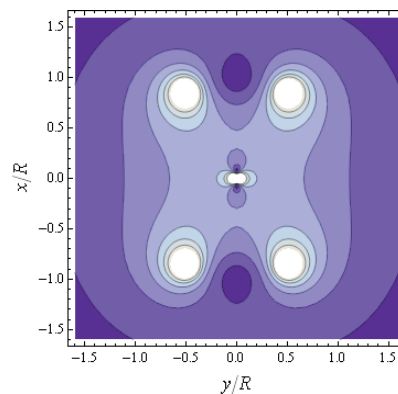
c)   Bi-isotropic external body, bi-isotropy of medium

strength, $\varepsilon_r = 2, \ \mu_r = 1, \xi^2/\varepsilon\mu = 0.5$
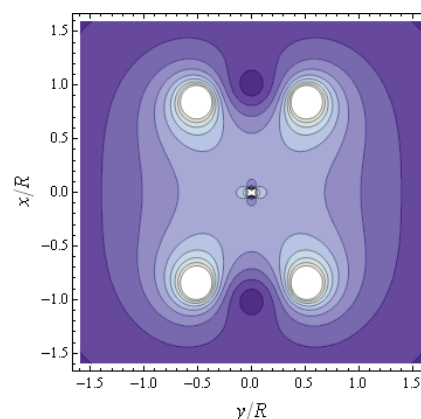
e) Bi-isotropic external body, strong bi-isotropy

$$\varepsilon_r = 2, \ \mu_r = 1, \xi^2/\varepsilon\mu = 0.9$$
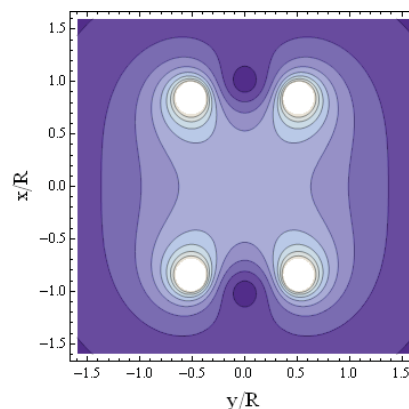
f) Bi-isotropic external body, extremely strong bi-

isotropy, $\varepsilon_r = 2, \ \mu_r = 2, \xi^2/\varepsilon\mu = 0.99$

Fig. 4. 2D presentation of the uniform electric field for a primary cell a) without body, b) with conducting body and with c) bi-isotropic external body with weak bi-isotropy; e) bi-isotropy of medium strength, e) strong bi-isotropy and f) extremely strong bi-isotropy.

It can be noticed that the external body with stronger bi-isotropy $(\varepsilon_r = 2, \mu_r = 1, \xi^2/\varepsilon\mu = 0.9)$ less disrupts the homogeneity of the field than the body with weak bi-isotropy $(\varepsilon_r = 2, \mu_r = 1, \xi^2/\varepsilon\mu = 0.1)$ whereas in the case of extremely strong bi-isotropy, Fig. 4.f), the body practically does not disrupt the homogeneity of the field.

VI. Conclusion

In this paper it is presented the influence of bi-isotropic external body on the achieved homogeneity of electric field generated by using plan-parallel system of the first order (primary cell). The body is modelled using a cylinder whose axis coincides with the axial axis of the system. Field calculations are performed using the Image theorem and we have observed bodies made by bi-isotropic materials of various strengths, conducting body as well as the case without a body.

The results show that the bi-isotropic body with a high degree of bi-isotropy $\xi^2/\varepsilon\mu > 0.8$ disrupts the homogeneity of the field in its vicinity to the least extent, comparing to the bi-isotropic body with a low degree of bi-isotropy $\xi^2/\varepsilon\mu < 0.2$. Furthermore, we have demonstrated that in the case of very strong bi-isotropy, external body does not disrupt homogeneity of the field whereas a conducting body disrupts the homogeneity to the largest extend comparing to the all other observed cases.

Acknowledgement

References

[1] M. W. Garrett, "Axially Symmetric Systems for Generating and Measuring Magnetic Fields", Prt I. J. Appl Phys, vol. 22. Pp. 1091-1107, 1951.

[2] C. E. Baum, "Impedances and Field Distributions for Symmetrical Two Wire and Four Wire Transmission Line Simulators", Sensor and Simulation, Note 27, Oktober 1966.

[3] J. A. Jungerman, "Fourth-order Uniform Electric Field from Two Charged Rings", Rev. Sci. Instrum., vol 55, pp. 1479-1482, 1984.

[4] D. M. Veličković, Z. Ž. Cvetković, "Systems for Generating of Homogeneous Electrical Field", Facta Universitatis, series Electronics and Energetics, Vol.14, no. 1, pp. 93-106, 2001.

[5] Z. Ž. Cvetković, "Homogeneous electric and magnetic field generation", PhD thesis, Faculty of Electronic Engineering, Niš, 2001. (in Serbian)

[6] Z. Ž. Cvetković, B. R. Petković, M. T. Perić, "Systems for Homogeneous Electrical Fields Generation and Effects of External Bodies on Field Homogeneity", Applied Computational Electromagnetics Society Journal, vol. 26, No. 1, pp. 56-63, January 2011.

[7] Z. Ž. Cvetković, S. R Aleksić, M. T. Perić, B. R. Petković, "Influence of the Conducting Body on the Biconical Electrode Field", The second European Conference on Antennas and Propagation - EuCAP 2007, CD-Proceedings (ThPA.025), Edinburg, United Kingdom, November 11-16, 2007.

[8] Z. Z. Cvetkovic, M. M. Potrebic, "Influence of Conducting Body on the Plan-Parallel Electrode Field", Procedings of 12th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, (TELSIKS), Serbia, Nis, ECIT.8, pp. 350-353, 14-17 October 2015.

[9] Z. Ž. Cvetković, M. M. Potrebić, "Stability analysis of plan-parallel systems for uniform electrostatic field generation", XII International SAUM Conference on Systems, Automatic Control and Measurements, pp. 308–311, Niš, Serbia, November 12–14, 2014.

[10] Z. Cvetković, Ž. Mančić, M. Potrebić, S. Ilić, Effects of external bodies made of diferent materials on plan-parallel system field homogeneity, Revue Romaine des sciences techniques, Series electrotechnique and energetique, Romanian Academy, Publishing House of the Romanian Academy, Vol. 64, no. 1, pp. 3-8, 2019.

[11] A. H. Sihvola, "Electromagnetic modeling in bi-isotropic media", Progres in electromagnetic research, PIER 9, pp. 45-86, 1994

[12] Jin Au Kong, "Theorems of bianisotropic media", Proceedings of the IEEE, Vol. 60, Issue 9, pp. 1036 – 1046, Sept. 1972.

[13] I. V. Lindel, "Static image method for layered isotropic and Bi-isotropic cylinders", Microwave and Optical Technology Letters, Vol. 6, No. 6, pp. 383-387, 1993.

[14] D. M. Veličković, Z. J. Mančić.: "Method of images for bi-isotropic media", Proceedings of full papers ПЕС`96, pp. 59-63, Niš, 22-23 may 1996.

[15] D. Veličković, B. Milovanović, Ž. Mančić, "New approach for equalising phase velocities in coupled microstrip lines", Electronics Letters, Vol. 33, no. 17, pp. 1444 – 1445, 14 Aug 1997.

[16] Ž. J. Mančić, S. Ilić, S. Aleksić, One type of strip lines on the surface of biisotropic cylindrical dielectric, PES 2007, pp. 59-60, Niš, 2007.

[17] J. Ž. 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), pp. 359−362, Niš, Serbia, October 18−20, 2017.

[18] Ž. J. Mančić, Z. Ž. Cvetković, S. S. Ilić, "Contribution to the electrostatics of bi-isotropic materials for the case of charged ring placed near to a sphere of a bi-isotropic material", Procedings of 7th Small Systems Simulation Symposium 2018, CD P10, pp. 59-62, Niš, Serbia, 12 - 14 Februari 2018.

[19] Ž. J. Mančić, Z. Ž. Cvetković, S. S. Ilić, "Point charge in the air spherical hollow inside bi-isotropic material", LIII International Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST 2018, Sozopol, Bulgaria, June 28.-30, 2018, pp. 71-74, 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 (Online).

[20] Ž. J. Mančić, Z. Ž. Cvetković, "Kružni linečni obruč u sfernoj vazdušnoj šupljini unutar bi-izotropnog materijala Telegenovog tipa", Zbornik radova 62 konferencije ETRAN 2018, Palić, Srbija, Jun 11-14, 2018, str. MT1.8-1-4, ISBN 978-86-7466-752-1, pp. 338-341. Izdavač: Društvo za ETRAN, Srbija

The Impact of Informative Technology in the Decision Process in Higher Education, with Special Studies in the University of Prishtina

Osman Sertolli[1] and Shpetim Lajqi[2*]

Abstract – **This paper gives some information about an information system used in higher education and in University in Pristina "HASAN PRISHTINA".**

Project as the impact of information systems at the University of Pristina actually is very important for the digitalization of our University.

Keywords – **Management, investment, project, risk, capital.**

I. INTRODUCTION

There are several goals which aim to achieve in this paper, we consider that some of the main goals are to provide a good basis to meet the theoretical knowledge on the use of computing systems and their impact on decision making.

To determine the level of fulfilling needs of managers with information processed by the computing system should:

1. Knowing the process of six stages of decision making,
2. Identifying the types of decisions,
3. Difference of qualitative and quantitative factors of decision making, and
4. Identifying potential problems of decision.

Century XX, especially the last two decades recorded an exponential growth in the use of information technology and information systems.

Information systems are tools without which there can be thought of the business of enterprises. This technology is being applied in the field of education and making decisions today can't be assumed and is not successful without the support of information systems.

Use of information systems in order to support the process of making management decisions will be our main objective, we focus particularly on Higher Education with special emphasis on the University of Pristina.

In the first part of the paper an introduction to paper is provided and research goals and hypotheses are presented.

[1]Osman Sertolli, Faculty of Economy at University of Prishtina "HASAN PRISHTINA", Rr. "Agim Ramadani" pn, 10 000 Prishtin, Kosovo. E-mail: osman.sertolli@uni-pr.edu

[2*]Shpetim Lajqi is corresponding author, Faculty of Mechanical Engineering at University of Prishtina, Bregu i Diellit pn, 10 000 Prishtina, Kosovo. E-mail: shpetim.lajqi@uni-pr.edu

Also, access is provided within the methodology used in preparing this paper.

In the second part of paper we will discussed about using information system and its importance. It provides basic notions about computing system, the concept of computing system, computing system components, types of computing systems and it will present the role and importance of information system as a processor of information

II. PROBLEM

A. Advantages of using computing system in Higher Education and the University of Pristine.

This section specifies the main features that higher education system should have in order to realize the vision and mission and to meet the challenges presented in introduction.

Are defined six key aspects which will be subject to transformation, reform and development, and six main objectives?

It is believed that transformation, reform and development provided in these areas will bring a more quality, more efficient and that will be comparable and harmonized with European standards.

The issue of quality assurance and its monitoring has become an important element in higher education policy at the present time.

Quality issue in higher education is dictated by national needs and the trend to internationalization of higher education.

This interest and concern about quality assurance stems from a broader context of social, political, economic.

B. Higher education institutions are responsible before to society and must demonstrate that they are making possible efforts to better use resources effectively entrusted.

In this context quality is foreseen as one of the key principles that will support the development of higher education in Kosovo in the next decade.

The ultimate goal of the proposed measures for this area is to achieve a quality comparable with European standards.

Investments in physical and academic infrastructure by international donors and financial means in the postwar period resulted primarily offer enriching academic programs in quantitative terms, but not in increasing the quality of work in higher education institutions.

Considered that such a disagreement between the investments made and results achieved is largely caused by the evident weaknesses in the management of higher education institutions and of the higher education system.

Nowadays, the operation of a changing environment and requirements, in terms of competition at national and regional level, it is necessary that higher education to have an entrepreneurial and market oriented approach.

Without an effective coordination and management, the development of a higher education system that functions to increase the welfare of society can't be imagine.

Implementation of an Integrated Management Information System on the campuses of universities and higher education has become very important in the XXI century, for contributing to sustainable development and affects the institution to increase the quality levels of all campuses the university.

Integrated Management Information System not only affects the constant improvement of the institution, but also increases public awareness about the new management system which would be a good model to use effective rules and efficient management.

So, Integrated Management Information System is a management system which integrates all the institution's systems and processes them into a comprehensive institution giving such an opportunity to work as a cohesive unit objectives.

C. The integrated system makes a completely unified organization, with each function listed after a single goal, thus improving the presentation of the whole organization.

Integrated Management Information System is designed in a way which provides a clear overall picture, of all aspects of the university, as these aspects affect each other and their associated risks.

Integrated Management for University Computing System is designed in order to bring solutions and measurable results by developing appropriate and state of the art applications in flexible collaborative approach to students, teaching staff, and workers in general and all persons involved in the university.

Integrated Management Information System as the university has developed into a way to integrate a suite of applications and databases that allow the university to perform administrative and educational duties in a very effective, efficient and very organized way.

Integrated Management Information System can modernize the university system, may enable to reduce the burdens of daily day to improve the quality of services for students, to bring new digital services that are relevant to the university, and last but great importance brings a full integration with all applicable systems.

Integrated Management Information System also offers data and systems with very fast connection possibilities, also has an automation of the administrative work of students, academic and administrative things, simplifying the form of parameters and finally the system provides support to technical things.

The Campus management system is considered to be the perfect solution for creating an administrative system management platform for students and academics to communicate and express their thoughts.

The Campus management system is designed in such a way that the university provide solutions portfolio also provides a system which integrates all the functions and activities of a university in a single platform.

III. RESULTS

A. The current situation of using computing system at the University of Prishtina

The Office of Information Technology at the University of Pristina has key role to set standards and to provide electronic services, arranges and coordinates the process of regeneration and investment in information technology in all academic units and administration central.

However, the use of computing systems at the University of Prishtina was accompanied by several problems and difficulties which are usually present in the initial stages of information system development.

Not forgetting the previous period in which we have been in the past we can't say that these obstacles were overcome after each period has its own challenges.

Especially, we should mention that in the field of Information Technology there are problems inherited from previous periods.

B. Empiric research results on the impact of information technology in the decision making process in higher education, with particular emphasis at the University of Pristina

In order to research the impact of Managing Information System in Higher Education with special emphasis on the University of Pristina for the purpose of this paper is a survey conducted by different respondents of different levels at the University of Pristina. Of the total 10 respondents interviewed all respondents answered positively.

1. When asked, the users of computing system in work processes in the University of Pristina, the 10 questionnaires that were made with 70% of respondents have answered yes, while 30% have answered NO.

2. When asked, do you know about the impact of information processed by the computing system in making decisions and how you use them at the University of Pristina, 90% of respondents answered regarding the importance of using the information processed by the computing system the decision making process, while a small percentage (10%) resulted to have had no knowledge regarding the impact of information processed by the Managing Information System in making decisions.

3. On the question, What is the nature of the works in which the use computing system, provided these answers: 50% of respondents answered that the computing system technology used in daily activities, 20% of them said they used the routine work, 30% of respondents answered that the computing system used in finance and accounting, and 10% in decision making.

4. When asked, What are your acquaintances on the application of information technology in University of Prishtina, 80% of respondents answered that possess good knowledge in terms of technology use informatics while 20% of them responded well know that on average use of this technology, since there has been no response to their acquaintances are weak.

5. When asked, By what year you started to use computer system at the University of Pristina?
Research shows that: 20% of respondent's answer that the use of Information Systems in University of Pristina has begun since 2006, and 80% of respondents answered that the use of Information System at the University of Pristina began in 2010.

6. The question, Do you think that the computing system allows performing more effectively and making quality decisions, the respondents ranked this way: 90% of respondents think that the computing system enables the effective execution of the works, while 10% of respondents answered no, not sure about this attitude of the respondents.

7. On the question, What use management structure computing system at the University Pristina, 10% of respondents answered senior management, 40% of them responded middle management. And 50% consider how low management use this system shows that the use of computing at the University of Pristina little used by senior management

while the other two management levels such as middle management 50% and low 40% use more of this technology.

IV. CONCLUSION

This paper is based on scientific and professional literature in the field of Management Information System and decision making are also used methods of analysis.

In the first, paper provides an introduction of the treated problem, purpose of research and methodology used.

Based on interviews and survey results we can conclude that the practice of using computing system at the University of Pristina in terms of 10 years has begun a major application of information system. While the level of computing system development in recent time steps begin development soon.

In general browsing interviews conducted in the relevant respondent have observed that the University of Pristina after the application of computing systems have been:

a. Management easier.
b. Facilitation of work and
c. Internal organization of work.

University funding should be clear, consistent and transparent.

Universities should be informed of their budget at least eight months before to create certainty and to enable forward planning.

It is not possible to increase the number of students without expanding the human and physical capacities of institutions of higher education. While some sources may be raised by student contributions, additional resources must come from the other state higher education system will be seriously threatened.

Universities should be encouraged to develop alternative sources of funding and should have full autonomy in the use of these funds. Faculties should have autonomy in the use of funds that they generate.

ACKNOWLEDGEMENT

REFERENCES

[1] M. Berisha-Namani, "Sistemi informatik i investimeve", Disertacioni i doktoraturës, Prishtinë, 2004.

[2] M. Berisha-Namani, "Informatika e biznesit", Ushtrime praktike në Excel, Prishtinë, 2003.

[3] M. Berisha–Namani, "Informatika e biznesit", Ligjerata, Prishtinë, 2004.

[4] B. Gjorgjijovski, "Menaxhment informacion sistemi", Fakulteti Ekonomik, Shkup, 1989.

[5] Koontz, H, Heinz, Weihrich, "Essentials of Management", 5th Edition, Mc Graw- Hill Publishing Company, New York, USA, 1990.

[6] L. Kenneth & K. Jane, "Management Information Systems", 8th Edition, Prentice Hall, USA 1990.

[7] Ph. Kotler, "Marketing Management" 11th Edition, Pearson Education LTD, New Jersey, USA, 2003.

[8] RG. Murdock & Munson J.C, "MIS-concepts and design", New Jersey, 1986.

[9] Ash. Mehmeti, "Fjalorë enciklopedik nga teknologjia e informacionit", Prishtinë, Kosovë, 2006.

[10] C.S. Parker, "Management Information System: Strategy and Action", New York, USA, 1989.

SYNTHESIS OF SOLUTIONS IN TRANSPORT TESTING IN MATLAB SOFTWARE ENVIRONMENT

Plamen Dqnkov, Stefan Kazakov

"Konstantin Preslavsky" University of Shumen

Republic of Bulgaria

Abstract: The task is a special type of linear optimization task. It refers to the allocation of quantities between a source group and a group of destinations in such a way as to minimize the total cost of this allocation.

Keyword: linear, optimization, transport, costs

I. INTRODUCTION

The solution for the transport task is based on a modified Simplex method for solving the Line Optimization Problem [1,2], called the transport simplicity method. The development of computer technology and the growth of its computing power, especially in the last decade, allows the study of complex objects and phenomena in real-time mode. The Matlab programming environment is a typical representative of this field [6]. In the MATLAB program environment [3,4], commands are available to solve the classical linear limiting task with limitations: linprog, bintprog, but in this case for a transport task, there is no such command. Here is considered a method for solving a Transport task by reducing it to a task of linear optimization.

Formulation of the transport task

The introduction of engineering design techniques into the engineering practice allows to move from the traditional method of creating a model of the developed equipment to modeling with the help of personal computers. [7]

Let m be given by sources Ai, which offer quantities a_i $i = 1...m$, and the destinations Bj are n in number, need for quantities bj, $j = l...n$. Let matrix C be given for the transport cost of one distribution unit from the i source to j destination:

$$C = \begin{cases} c_{11}\, c_{12} \,.... \,c_{1n} \\ c_{21}\, c_{22} \,.... \,c_{2n} \\ c_{m1}\, c_{m2} \,.... \,c_{mn} \end{cases} \quad (1)$$

Let x_{ij} be unknown units of distribution from i source to j, $i = 1.m, j = 1...n$.

The goal is to minimize the total cost of allocating these quantities.

Transport task:

Minimize the function of total transport costs:

$$\min Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \quad (2)$$

subject to restrictions:

$$\sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1 ..... m, \quad (3)$$
$$\sum_{i=1}^{m} x_{ij} = b_j, \quad j = 1 ..... n \quad (4)$$

and limitations for the non-negative of the unknown variables $x_{ij} > 0$.

The Matrix X of the Unknown is called a transport matrix.

$$X = \begin{pmatrix} x_{11}\ x_{12}\, ...\, x_{1n} \\ x_{21}\ x_{22}\, ...\, x_{2n} \\ ...\ ...\ ...\ ...\ ...\ . \\ x_{m1}\ x_{m2}\, ...\, x_{mn} \end{pmatrix} \quad (5)$$

A transport matrix satisfying the system constraints (2) - (3) is called an acceptable solution to the transport task. For convenience, the prices, availability and needs of the transport task are given in Table 1:

	B_1	B_2	B_n	a_i
A_1	c_{11}	c_{12} ……………........	c_{1n}	a_1
A_2	c_{21}	c_{22} ……………….	c_{2n}	a_2
.				
.				
A_m	c_{m1}	c_{m2} ………………….	c_{mn}	a_m
b_j	b_1	b_2 ………………….	b_n	

Specific features of the transport task

If the unknown variables are treated as a pillar vector derived from the X matrix rows, then the Transport task is recorded as a Line Optimization task:

$$min\ Z = c_{11}x_{11} + c_{12}x_{12} +....c_{1n}x_{1n} + c_{21}x_{21} + c_{22}x_{22} +...+ c_{m1}x_{m1} + c_{m2}x_{m2} +...c_{mn}x_{mn} \quad (6)$$

subject to restrictions:

$$\begin{pmatrix} 1\ 1...1\ 0\ 0\ ...\ 0\ 0\ 0\ ...\ 0 \\ 0\ 0...0\ 1\ 1\ ...1\ 0\ 0\ ...\ 0 \\ .... \\ 0\ 0\ ...0\ 0\ 0\ ...\ 0\ 1\ 1\ ...\ 1 \\ 1\ 0\ ...0\ 1\ 0\ ...\ 0\ 1\ 0\ ...\ 0 \\ 0\ 1\ ...0\ 0\ 1\ ...\ 0\ 0\ 1\ ...\ 0 \\ .... \\ 0\ 0\ ...0\ 0\ 0\ ...\ 0\ 0\ 0\ ...\ 0 \\ 0\ 0\ ...1\ 0\ 0\ ...\ 1\ 0\ 0\ ...\ 1 \end{pmatrix} \begin{pmatrix} x_{11} \\ . \\ x_{1n} \\ . \\ x_{m1} \\ . \\ x_{mn} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \\ b_1 \\ b_2 \\ b_n \end{pmatrix} \quad (7)$$

and limitations of non-negativity $h_{ij} \geq 0$.

The transport task has the following specific features:

• The task is to minimize the target function;

• All restrictions are of the "equality" type;

• The matrix elements of the coefficients before the unknowns in the limitations are zeros or units;

• In each pillar of the matrix of coefficients in front of the unknown in the limits there are two units, ie each variable is involved only in two equations with a coefficient unit.

Theorem 1: The transport task (1) - (3) has an acceptable solution then and only when the class is fulfilled:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \qquad (8)$$

Theorem 2: The rank of the equation system (2) equals $m + n-1$.

Consequence: Any acceptable solution X of the transport task contains at most $m + n-1$ positive components x_{ij}. The remaining components are zero.

Definition 1: An acceptable solution to the transport task, in which there are exactly $m + n -1$ positive components h_{ij}, is called unconditionally.

Definition 2: An acceptable solution to the transport task in which there are less than $m + n-1$ positive components h_{ij} is called degenerate.

Types of transport tasks
• Balanced and unbalanced transport tasks

According to whether the condition of Theorem 1 is met, transport tasks are two types: balanced and unbalanced. If the condition (4) is met, the transport task is called a balanced one, or in the opposite case:

$$\sum_{i=1}^{m} a_i \neq \sum_{j=1}^{n} b_j \qquad (9)$$

the task is unbalanced. There are two possible scenarios:

1. Offering is greater than demand. In this case, a fictitious destination (user) is entered, with zeros for transport costs and the required amount - the difference between total stocks and needs:

$$\sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j \qquad (10)$$

2. Demand is greater than supply. In this case, a fictitious source (manufacturer) is introduced, the transport costs are zero and the quantity produced - the difference between total needs and stocks:

$$\sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i \qquad (11)$$

• A blocked transport task (with bans)

When there is no transport link between a source and destination, the transport task is "blocked shipments". In this case, a very positive number $M >> 0$ (large M method) is assumed for transport expense.

Another case where the use of the M-method is required is when the Transport task is unbalanced and the user B_κ is required to obtain the requested amount in full. In this case, after introducing a fictitious source A_s, M $(M >> 0)$ is applied to transport expense c_{sk}, ensuring that B_k obtains the necessary quantities only from real sources.

An analogous situation occurs when the task is unbalanced, and in addition a producer wants to distribute all of his available quantity. Then, after introducing a dummy user (destination) B_K, M $(M >> 0)$ is applied for transport expense c_{sk}, thus ensuring that A_s can distribute its available quantities only to real users (destinations).

Program implementation

The Transport Troubleshooting Program is a MATLAB function with the following arguments:

Entry arguments:
• Matrix C of transport costs;
• Vector a - quantities offered;
• Vector b - Queries;
Output arguments:
• Z - value of the target function;
• X - matrix of the solution of the Transport task.
After input values are entered, the command is set:
$[Z,X]=transport(C,a,b)$

In the event of an unbalanced transport task, the program automatically balances it. When solving a blocked transport task(Fig. 1, Fig. 2), the corresponding element in transport matrix C is assigned a large enough number, of the order of $1e+009$.

```
function [Z,X]=transport(C,a,b)
[m,n]=size(C);
[ma,na]=size(a);
[mb,nb]=size(b);
if ma<na
    a=a';
    [ma,na]=size(a);
end
if mb<nb
    b=b';
    [mb,nb]=size(b);
end
if (ma~=m)|(mb~=n)
    error('некоректен брой елемемти на a или b')
end
suma=sum(a);
sumb=sum(b);
Iflag=0;
if sumb>suma
```

Fig. 1. Blocked transport task

```
    Iflag=1;
    nulb=zeros(1,length(b));
    C=[C;nulb];
    a=[a;sumb-suma];
end
if sumb<suma
    Iflag=2;
    nula=zeros(length(a),1);
    C=[C,nula];
    b=[b;suma-sumb];
end
B=[a;b];
[m,n]=size(C);
f=reshape(C',m*n,1);
A1=zeros(m,m*n);
A2=eye(n);
A3=eye(n);
for k=1:m
    A1(k,((k-1)*n+1):(k*n))=1;
end
for k=1:(m-1)
    A2=[A2,A3];
end
A=[A1;A2];
lb=zeros(m*n,1);
options = optimset('LargeScale', 'off', 'Simplex', 'on');
[X,Z]=linprog(f,[],[],A,B,lb,[],[],options);
X=reshape(X,n,m);X=X';
if Iflag==1
    X(end,:)=[];
elseif Iflag==2
    X(:,end)=[];
end
```

Fig. 2. Blocked transport task

Numerical experiments

The program for solving a transport task has been tested multiple times with different numerical data. Here are three examples of balanced, unbalanced tasks and a Transport task with blocked shipments (Fig. 3, Fig. 4).

No.1	B_1	B_2	B_3	a_i
A_1	4	4	2	30
A_2	5	3	1	20
A_3	3	6	4	40
b_j	45	35	10	

No.2	B_1	B_2	B_3	a_i
A_1	1	2	4	80
A_2	4	2	6	70
A_3	6	1	3	50
b_j	50	80	50	

No.3	B_1	B_2	B_3	a_i
A_1	20	45	40	60
A_2	–	30	25	60
A_3	–	15	20	30
b_j	20	10	30	

```
>> Пример 1: балансирана транспортна задача
>> C=[4 4 2;5 3 1;3 6 4];
>>a=[30 20 40];
>> b=[45 35 10];
>> [Z,X]=transport(C,a,b)
Z =
    280

X =
    5    25    0
```

Fig. 3. Transport task with blocked shipments

```
    0    10    10
    40    0     0

>> Пример 2: небалансирана транспортна задача
>> C=[1 2 4;4 2 6;6 1 3];
>> a=[80 70 50];
>> b=[50 80 50];
>> [Z,X]=transport(C,a,b)
Z =
    360
X =
    50    10    0
    0     70    0
    0     0    50

>> Пример 3: небалансирана транспортна задача с блокирани превози
>> C=[20 45 40;1e9 30 25;1e9 15 20];
>>a=[60 60 30];
>> b=[20 10 30];
>> [Z,X]=transport(C,a,b)
Z =
    1200
X =
    20    0     0
    0     0    10
    0    10    20
```

Fig. 4. Transport task with blocked shipments

II. CONCLUSION

The Transport Problem Solving Program is widely applied in science and practice and is successfully applied to the study of Optimization Methods.

REFERENCES

[1] Караколева, Ст. Investigation of Operations, Part 1, Angel Kanchev University, 2002.

[2] Pavlov, V. Mathematical Methods in Economics, "Angel Kanchev" University, 2006.

[3] Tonchev, Y., Matlab 7, Part 3, Technique, 2009.

[4] Matlab Optimization Toolbox documentation: www.mathworks.com/access/helpdesk/help/toolbox/optim/

[5] Образовательнь математический сайт http://matlab.exponenta.ru/optimiz/index.phply applied to the study of Optimization Methods.

[6] Lalev H., Zhelev S., Tsankov Ts. Analyzing and monitoring the frequency response of the circuit. Scientific Session of the "Vasil Levski" National Military University, "Artillery, Air Defense and CIS" Faculty, Shumen, ISSN 1314-1953 (in Bulgarian).

[7] Lalev H., Tsankov Ts. Synthesis of simulink modules. Scientific Conference, dedicated to the 105th anniversary of the birth of the pioneers of computing John Atanasoff and John von Neumann, Shumen, 2008, ISBN 978-954-577-540-6. (in Bulgarian).

[8] Антон Антонов, Транспортна логистична система за екологичен морфологичен анализ на отпадъци, Университетско издателство "Епископ Константин Преславски", гр. Шумен 2016

[9] Антон Антонов, Логистични методи и транспортна схема в преработката на био-разградими отпадъци. Third international scientific conference science, education, innovation, shumen, Bulgaria 21 - 23 MAY 2014. Vol. 2, 2014

Effectiveness of Using GeoGebra in Mathematics

Sindize Jusufi[1] and Stojan Kitanov[2]

Abstract – **The rapid growth of Information and Communication Technology results with the software used in education. This paper explores the effectiveness of GeoGebra software in learning the 7th grade elementary school mathematics among 50 students. The results show that students have a positive perception of learning mathematics and have better learning outcomes using GeoGebra.**

Keywords – **GeoGebra, digital educational content, educational software, educational technology, learning performance, mathematics.**

I. INTRODUCTION

The rapid growth of Information and Communication Technology (ICT) resulted with a powerful learning resources in the education. One such tool that can help teachers to design effective learning lessons in mathematics is GeoGebra [1]. GeoGebra is a dynamic software for teaching and learning mathematics. It is a computer program (software) for mathematics, especially for the learning of geometry and algebra [2], [3]. Although GeoGebra has not yet been widely used in teaching mathematics in Macedonia, it improves the learning efficiency, because it allows easy access to information which facilitates the learning.

This paper explores the different methods and activities that would positively influence the student to solve the problems and mathematical tasks, by using the GeoGebra software. It also aims to prove the extent to which technology tools can influence their teaching and learning in the subject of mathematics.

The rest of the paper is organized as follows. Section II states the problem motivation and related work. Section III provides information on the research methodology. Section IV performs analysis of the results. Firstly, it makes a comparison of the learning performances of the students who studied mathematics classically and the learning performances of the students who studied mathematics with GeoGebra. The results show better learning performances of the students that used GeoGebra. At the end, an analysis of the results from the question survey about the effectiveness of using GeoGebra in learning and understanding the mathematical topics was performed. The analysis of the results confirm the fact that the use of GeoGebra in teaching positively influences students' understanding and knowledge. Finally, Section V concludes the paper, it gives some recommendations, as well as some future work directions.

[1]Sindize Jusufi is with the Faculty of Informatics at Mother Teresa University, Skopje, Macedonia, E-mail: sindize.jusufi@students.unt.edu.mk.

[2]Stojan Kitanov is with the Faculty of Informatics at Mother Teresa University, Skopje, Macedonia, E-mail: stojan.kitanov@unt.edu.mk.

II. PROBLEM MOTIVATION AND RELATED WORK

Some of the factors that affect the students' attitude towards efficient learning of mathematics are the learning materials used by teachers, classroom management, teaching methods, teacher's knowledge and personality of the content, as well as similar and related topics with the real-life situation. The learning materials and methods that are used by the teachers are either conventional classic educational materials or methods, or can be digital educational software together with online learning management system such as Moodle, Google Classroom, etc. Although the teaching methods in Macedonia are still classical, our motivation was to improve the learning outcomes of the students in mathematics, by using the GeoGebra software, because nowadays technology plays an important role in the development of the educational process.

Several studies have been conducted about using GeoGebra in secondary and elementary schools in the world [3]. To our best knowledge no similar research has been performed so far in Macedonia. One approach, which is very close to our research methodology, is given in [4]. Here are shown the performances of students' achievements by using GeoGebra particularly in statistics. In this research study, firstly pre-achievement tests were analyzed. Then, the students learned statistics with GeoGebra and after that post-achievement tests were analyzed. At the end a survey questions were answered by the students about the effects of learning statistics with GeoGebra. However, this approach does not consider the classical learning and it does not measure the effectiveness by the classical learning methods. Our research methodology includes this.

III. RESEARCH METHODOLOGY

The research methodology is given in Fig. 1. It consists of 5 phases.
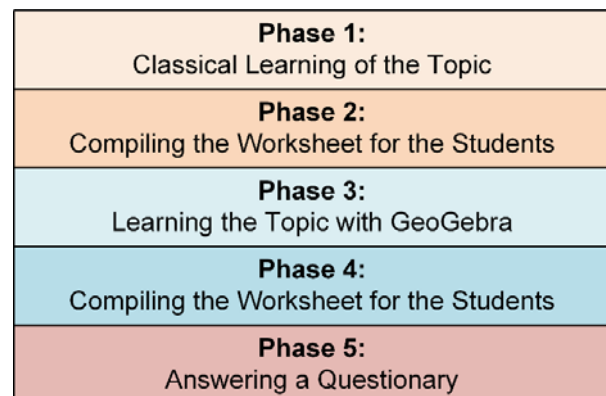
Fig. 1. Phases of the Research Methodology

In the **first phase** the mathematical teaching unit topic is classically explained. In the **second phase** a worksheet for the students is compiled, in order to understand how much the students have learned the mathematical unit topic with the classical teaching method. During the **third phase** the teaching unit is explained the with the application program GeoGebra. In the **fourth phase**, a worksheet for the students is compiled again in order to understand how much they understood the methodology unit with GeoGebra. Finally, in the **fifth phase** a questionnaire survey was answered by the students in order to see the evaluate the impressions about the explaining teaching units with GeoGebra. The worksheets were used to compare what the students knew when the lesson was classically explained and what they knew after the lesson was explained with the GeoGebra program. The questionnaire contains statements that reflect students' perception of using GeoGebra software. Table I summarizes the questions included in the question survey, as well as its possible answers.

TABLE I
QUESTION SURVEY

Question	Possible Answers
I want to learn Mathematics	Always Sometimes Never
Mathematics is difficult	Always Sometimes Never
Students understand the tasks of mathematics	Always Sometimes Never
Mathematical tasks are done by the students themselves	Always Sometimes Never
How do the student understand mathematics with GeoGebra	Nothing Slightly Somewhat More Much More
Student easily understand the mathematical tasks and Problems with GeoGebra	Always Sometimes Never
Do you want to continue to study mathematics with GeoGebra	Yes No

The research methodology was conducted in the primary school "Adem Jashari" in Skopje from 17 April to 30 April 2019, in a 7th grade class of 50 students. The topic that was taught was 3D formats. The research was done after the explanation of the "3-D format" teaching. Some techniques are used for data collection such as worksheets and questionnaires for students.

IV. ANALYSIS OF THE RESULTS

The results of the research methodology, are shown in the following two subsections. Subsection A analyzes

performances of the students for the planned work, i.e. the results of the students when the topic is classically delivered to the students and the results when the topic is explained to the students with GeoGebra. Subsection B analyzes the results of the question survey.

A. Results of the Students for the Planned Work

The results based on the first job sheet (Phase 2 of the research methodology), i.e. when the mathematical classes are explained to the students with the classical and conventional teaching methods are provided in Fig. 2. The students' test results show that 42% of students have given the correct answer, while 58% of the students have given incorrect answers.
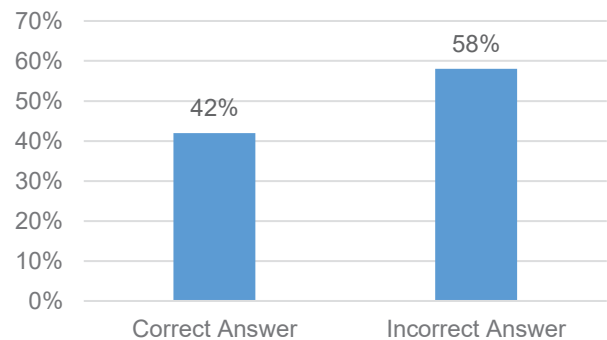
Fig. 2. Results of the Assessment when Mathematics is taught with Classical Methods

The results based on second working paper (Phase 4 of the research methodology), i.e. when the mathematical classes are explained to the students with the GeoGebra are provided in Fig. 3. It can be noticed that that 78% of the students have answered correctly, while the remaining 22% didn't provide correct answers.
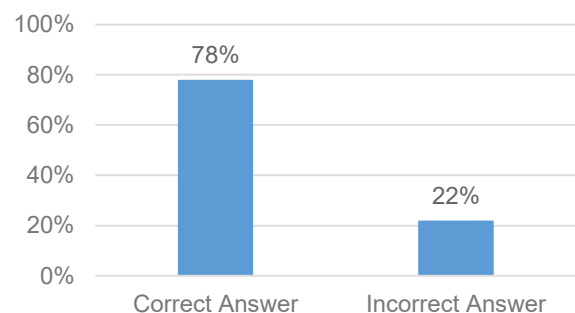
Fig. 3. Results of the Assessment when Mathematics is taught with GeoGebra

It can be concluded that when GeoGebra is not used in teaching mathematics students provide more incorrect answers. If GeoGebra is used for teaching mathematics students provide much more corrects answers.

B. Results of the Question Survey with the Students

The results of the question survey with Students (Phase 5 of the research methodology) are provided from Figs 4 - 10.

The results of how much the students want mathematics are displayed in Fig. 4. It can be noticed that about one third of the students love to learn mathematics, almost two thirds students sometimes love to learn mathematics and very small number of students never want to learn mathematics.
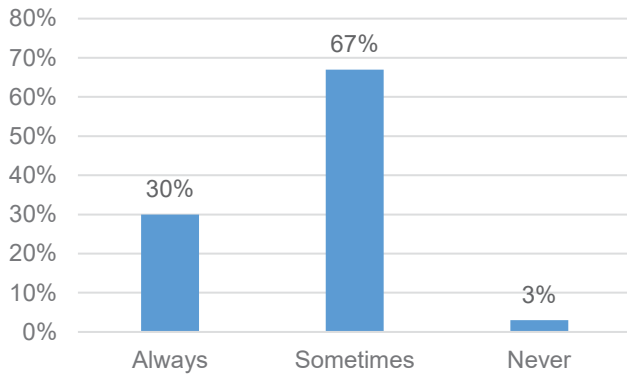
Fig. 4. Students Want to Learn Mathematics
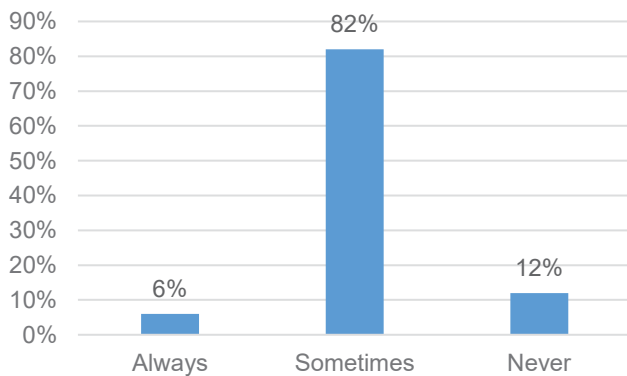
Fig. 5. Mathematics is Difficult for the Students

The results of how much mathematics is difficult for the students are displayed in Fig. 5. More than 80 % have stated that mathematics sometimes is difficult to learn, about 12 % have stated that mathematics is never difficult to learn, and about 6 % stated that mathematics is always difficult to learn.

The results of how much students understand the tasks from mathematics are provided in Fig. 6. About 80 % of the students have stated that always understand the tasks of mathematics, 16 % of the students have stated that sometimes understand the tasks of mathematics and very small number of students (about 4 %) do not understand the mathematical tasks.
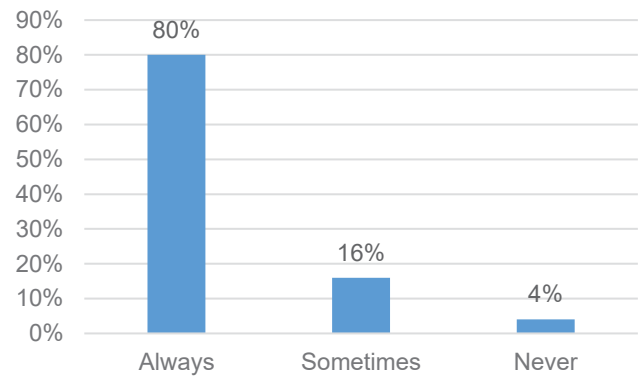
Fig. 6. Students Understand the Tasks of Mathematics
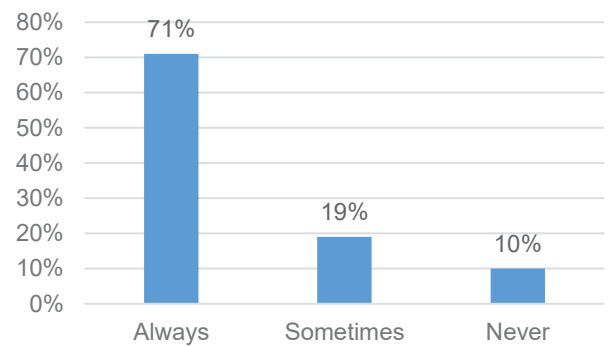
Fig. 7. Mathematical Tasks are Done by the Students Themselves – Without Any Help

The results of how much the mathematical tasks are done only by the students themselves are given in Fig. 7. More than 70 % of the students always do the mathematical tasks by themselves, almost 20 % of the students sometimes do the task of mathematics with the assistance from others, and 10 % of the students never do the mathematical tasks by themselves.
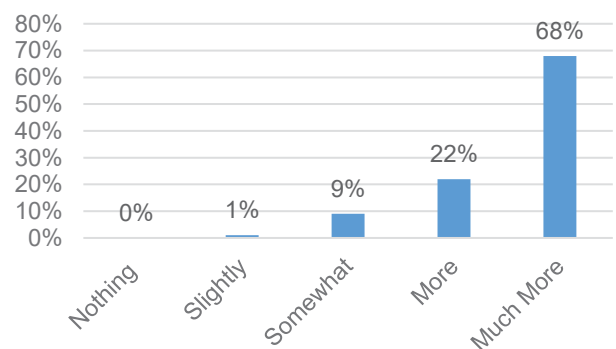
Fig. 8. How much the Student understands Mathematics with GeoGebra

The results of how much the students understand the course mathematics with GeoGebra are given in Fig. 8. Almost 70 % of the students have stated that much more understand mathematics with GeoGeobra, more than 20 % of the students have declared that understand more mathematics with GeoGebra, almost 10 % of the students have stated that slightly somewhat understand mathematics with GeoGebra, and 1 % of the students have stated slightly understand mathematics with GeoGebra. In general, it can be concluded that most of the students understand better mathematics with GeoGebra.

The results of how each student easily understands the mathematical tasks with GeoGebra are provided in Fig. 9. All students have stated that always understand the mathematical tasks with GeoGeobra, which corresponds to the same conclusion extracted from Fig. 8.

Finally, answers whether the students would like to continue to study mathematics with GeoGebra are provided in Fig. 10. All of them stated that they would like to continue to learn mathematics with GeoGebra software, because most of them have improved the learning efficiency in mathematics, even those who thought that mathematics is not easy to learn.
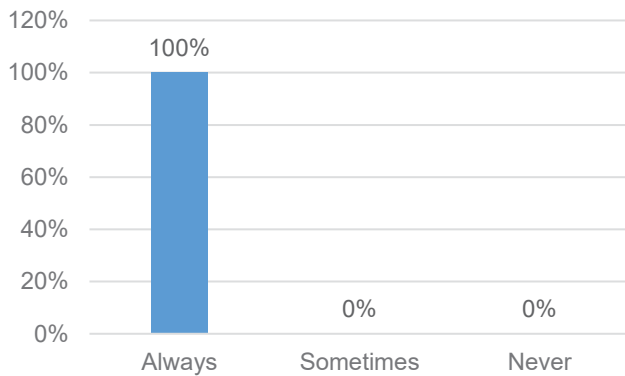
Fig. 9. Students Easily Understand the Mathematical Tasks and Problems with GeoGebra
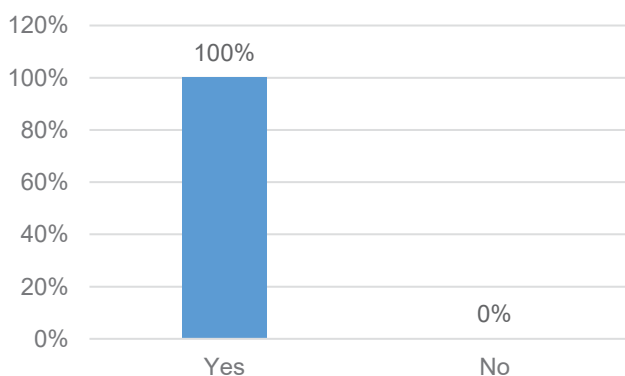
Fig. 10. Students Would Like to Continue to Study Mathematics with GeoGebra

V. Conclusion

This paper was about the effectiveness of using GeoGebra software in learning mathematics. Firstly, it was stated the motivation for this research and the related work. Then the research methodology was provided. Finally, the obtained results were analysed and discussed.

The results from this research proved that mathematics can be learned more effectively and more efficiently by the use of GeoGebra, or even any other software, because it makes the subject topics more interesting, and easy to learn like other subjects. Also, this research made it possible to change the persuasions of students who considered the subject of math very difficult to learn. Students can benefit from the GeoGebra software in learning mathematics, because they can interact with technology. They use the Internet, cell phones, computers, laptops, tablets and other software to communicate with others. The digital environment increases the students' motivation and self-confidence in learning mathematics. Therefore, GeoGebra should be introduced to math educators so that students can explore the world of mathematics in a broader way. Moreover, GeoGebra in the process of learning the mathematics would introduce new innovative teaching and learning methods for the students.

However, this research was performed with a small number of students and classes. The results would be even more compelling if more students were involved. Likewise, the duration of the research may also be a restrictive circumstance for the results to be even more convincing.

In the future our plan is to make a more deeper research with greater number of students in order to obtain more realistic results, as well as the obtained results would be analyzed with other statistical methods and software tools.

Acknowledgement

References

[1] I. Žilinskienė, "Use of GeoGebra in Primary Math Education: a theoretical approach," Proc. of the Lithuanian Mathematical Society, Ser. A, Vol. 55, 2014, pp. 73-78.

[2] Ljubica Diković, "Applications GeoGebra into Teaching Some Topics of Mathematics at the College Level," ComSIS Journal, Vol. 6, No. 2, 2009, pp. 191-203.

[3] I. Žilinskienė, M. Demirbilek "Use of GeoGebra in Primary Math Education in Lithuania: An Exploratory Study from Teachers' Perspective," Journal of Informatics in Education, 2015, Vol. 14, No. 1, pp. 127–142.

[4] N. Arbain, N.A. Shukor, "The Effects of GeoGebra on Students Achievement," Procedia - Social and Behavioral Sciences, Vol. 172, 2015, pp. 208-214.

An Original Approach To The Construction of $(3,2,\rho)$-N-Symmetrizable Hilbert Spaces

Sonja B. Chalamani[1], Elena T Kotevska[2] and Marzanna J. Seweryn-Kumanovska[3]

Abstract: **In this paper we will present an original approach to the construction of a special type of Hilbert space. Assuming that a set M is given with a $(3,2,\rho)$-metric d on it, the construction of $(3,2,\rho)$ -N-symmetrizable Hilbert space will be described in detail. This Hilbert space is characterized by very interesting topological structure and properties.**

Keywords – **$(3,2,\rho)$-metric, $(3,2,\rho)$-metric spaces, $(3,2,\rho)$-N-symmetrizable spaces**

I. INTRODUCTION

The notion of metric space leads to an important topological notion, viz., the notion of a metrizable space. We lay great stress on metric and metrizable spaces because many important topological spaces used in various branches of mathematics are metrizable. Axiomatic classification and the generalization of metric and metrizable spaces have been considered in a lot of papers. We will mention some of them: K. Menger ([14]), V. Nemytzki, P.S. Aleksandrov ([16],[1]), Z. Mamuzic ([13]), S. Gähler ([12]), A. V. Arhangelskii, M. Choban, S. Nedev ([2],[3],[17]), J. Usan ([18]), B. C. Dhage, Z. Mustafa, B. Sims ([8], [15]). The notion of (n,m,ρ)-metric is introduced in [9]. Connections between some of the topologies induced by a $(3,1,\rho)$ -metric d and topologies induced by a pseudo-o-metric, o-metric and symmetric are given in [10]. For a given $(3,j,\rho)$ -metric d on a set M, $j\epsilon\{1,2\}$, seven topologies $\tau(G,d)$, $\tau(H,d)$, $\tau(D,d)$, $\tau(N,d)$, $\tau(W,d)$, $\tau(S,d)$ and $\tau(K,d)$ on M, induced by d, are defined in [4] and several properties of these topologies are shown, such as: $\tau(W,d)\subseteq\tau(N,d)\subseteq\tau(D,d)\subseteq\tau(G,d)$, $\tau(W,d)\subseteq\tau(H,d)\subseteq\tau(G,d)$, $\tau(W,d)\subseteq\tau(S,d)\subseteq\tau(K,d)$. For a $(3,2,\rho)$ -metric d on M, the following inclu-sions are also satisfied $\tau(W,d)\subseteq\tau(N,d) = \tau(S,d) = \tau(K,d)\subseteq\tau(D,d)\subseteq\tau(G,d)$.

1. Sonja B. Chalamani is with the Faculty of Technical Sciences University St. Kliment Ohridski, Makedonska Falanga 33, Bitola 7000 Email: scalamani@yahoo.com
2. Elena T. Kotevska is with the Faculty of Technical Sciences University St. Kliment Ohridski, Makedonska Falanga 33, Bitola 7000 Email: elena.kotevska@tfb.uklo.edu.mk
3. Marzanna J. Seweryn-Kumanovska is with the Pedagogical Faculty University St. Clement of Ohrid St. Kliment Ohridski, General Vasko Karangjelevski B.B., Bitola 7000 Email: marzanna.kuzmanovska@yahoo.com

In this paper we consider only the topologies $\tau(N,d)$ and $\tau(D,d)$ induced by a $(3,2,\rho)$-metric d. We will prove that the set

$$\beta = \{B(a,a,\varepsilon)|a \in M, \varepsilon > 0\}$$

is a base for $\tau(N,d)$ and the Hilbert space H is a $(3,2,\rho)$-N-symmetrizable with topology $\tau = l^2$(on R^3) and is a separable space.

II. SOME PROPERTIES OF $(3,2,\rho)$-N-SYMMETRIZABLE SPACES

In this part we state the notions (defined in [4]) used later.

Let M be a nonempty set and let $d: M^3 \to R_0^+ = [0,\infty)$.

We state five conditions for such a map.

$(\boldsymbol{M0})$ $d(x,x,x) = 0$, for any $x \in M$;

$(\boldsymbol{P})$ $d(x,y,z) = d(x,z,y) = d(y,x,z)$, for any $x,y,z \in M$

$(\boldsymbol{M1})$ $d(x,y,z)\leq d(x,y,a) + d(x,a,z) + d(a,y,z)$, for any $x,y,z,a \in M$;

$(\boldsymbol{M2})$ $d(x,y,z)\leq d(x,a,b) + d(a,y,b) + d(a,b,z)$, for any $x,y,z,a,b \in M$;

$(\boldsymbol{Ms})$ $d(x,x,y) = d(x,y,y)$, for any $x,y \in M$.

For a map d as above let

$\rho = \{(x,y,z)|(x,y,z) \in M^3, d(x,y,z) = 0\}$.

The set ρ is a $(3,j)$ -equivalence on M, as defined and discussed in [9], [4].

The set $\Delta = \{(x,x,x)|x \in M\}$ is (3,1)-equivalence on $M, j = 1,2$, and the set

$\nabla = \{(x,x,y) \mid x,y \in M\}$ is a (3,1)-equivalence, but it is not a (3,2)-equivalence on M. However, the condition $(\boldsymbol{M0})$ implies that $\Delta\subseteq\rho$.

Definition.2.1. Let $d: M^3 \to R_0^+$ and ρ be as above. If d satisfies $(\boldsymbol{M0})$, $(\boldsymbol{P})$ and $(\boldsymbol{Mj})$, $j \in\{1,2\}$, we say that d is a $(3,j,\rho)$**-metric** on M. If d is a $(3,j,\Delta)$-metric on M, we say that d is a $(\boldsymbol{3,j})$**-metric** on M. If d is a $(\boldsymbol{3,j,\rho})$**-metric** and satisfies $(\boldsymbol{Ms})$, we say that d is a $(\boldsymbol{3,j,\rho})$**-symmetric** on M, and if d is a $(\boldsymbol{3,\rho})$**-metric** and satisfies $(\boldsymbol{Ms})$, we say that d is a $(\boldsymbol{3,\rho})$**-symmetric** on M.

Remark 2.1. Any $(3,j,\rho)$ -metric d on M induces a map $D_d: M^2 \to R_0^+$ defined by:

$$D_d(x,y) = d(x,x,y).$$

It is easy to check the following facts.

a) For any $(3,j,\rho)$-metric d, $D_d(x,x) = 0$. (D_d is called a **distance** in [15] and a **pseudo o-metric** in [17].)

b) For any $(3, j)$-metric d, $D_d(x, y) = 0$ if and only if $x = y$. (D_d is called an *o-metric* in [17].)

c) For any $(3, j)$-symmetric d, $D_d(x, y) = D_d(y, x)$. (D_d is called a **symmetric** in [17].)

d) For any $(3, 2, \rho)$-metric d,
$D_d(x, y) \leq 2D_d(z, x) + D_d(z, y)$ and
$D_d(x, y) \leq 2D_d(y, x)$.

e) For any $(3, 2)$-symmetric d,
$D_d(x, y) = D_d(y, x) \leq 3(D_d(x, z) + D_d(z, y))/2$. (In the literature D_d is called a **quasimetric,** a **nearmetrics** or an **inframetrics**.)

Let d be a $(3, 2, \rho)$-metric on M, $x, y \in M$ and $\varepsilon > 0$. As in [4], we consider the following ε-ball, as subset of M
$B(x, y, \varepsilon) = \{z | z \in M, d(x, y, z) < \varepsilon\}$–$\varepsilon$-ball
with center at (x, y) and radius ε.

Remark 2.2. a) For $x = y$, $B(x, x, \varepsilon) = B(x, y, \varepsilon) = \{z \mid z \in M, d(x, x, z) < \varepsilon\}$. For any $a \in M$, $a \in B(a, a, \varepsilon)$, but, it is possible for some $x \neq a$ to have $a \notin B(a, x, \varepsilon)$.

b) For a pseudo o-metric $D: M^2 \rightarrow R_0^+$, there is only one possibility for defining ε-balls, i.e.
$B(x, \varepsilon) = \{z \mid z \in M, D(x, z) < \varepsilon\}$.

Among the others, a $(3, 2, \rho)$-metric d on M induces the following two topologies as in [4]:

1) $\tau(N, d)$–the topology defined by: $U \in \tau(N, d)$ iff $\forall x \in U$, $\exists \varepsilon > 0$ such that $B(x, x, \varepsilon) \subseteq U$;

2) $\tau(D, d)$ –the topology generated by all the ε-balls $B(x, x, \varepsilon)$.

In [4] we proved that $\tau(N, d) \subseteq \tau(D, d)$ for any $(3, 2, \rho)$-metric d. With the next proposition we will show that $\tau(N, d) = \tau(D, d)$ for any $(3, 2, \rho)$-symmetric d.

Proposition 2.1. For any $(3, 2, \rho)$-symmetric d on M, the ball $B(a, a, \varepsilon) \in \tau(N, d)$, for any a on M and $\varepsilon > 0$.

Proof: It is enough to show that for any $x \in B(a, a, \varepsilon)$ there is $\delta > 0$, such that $B(x, x, \delta) \subseteq B(a, a, \varepsilon)$. Let $x \in B(a, a, \varepsilon)$ and $\delta = (\varepsilon - D_d(a, x))/2$. Then, for any $z \in B(x, x, \delta)$ we have that
$D_d(a, z) = D_d(z, a)$
$\leq 2D_d(z, x) + D_d(x, a)$
$< 2\delta + D_d(x, a)$
$= 2(\varepsilon - D_d(a, x))/2 + D_d(x, a) = \varepsilon$.
This implies that $z \in B(a, a, \varepsilon)$, i.e., $z \in B(x, x, \delta) \subseteq B(a, \varepsilon)$.

This proposition shows that the set $\beta = \{B(a, a, \varepsilon) | a \in M, \varepsilon > 0\}$ is a base for $\tau(N, d)$. Moreover, $\tau(N, d) = \tau(D, d)$.

Definition 2.2. We say that a topological space (M, τ) is $(3, 2, \rho)$-N-symmetrizable via a $(3, 2, \rho)$-symmetric d on M, if $\tau = \tau(N, d)$.

III. CONSTRUCTION OF $(3, 2, \rho)$-N-SYMMETRIZABLE HILBERT SPACE

In [4] we proved that any $(3, 2)$-N-D-metrizanle spaces (M, τ) is metrizable, so by this in [6] we proved that any $(3, 2)$-N-metrizable spaces (M, τ) is perfectly normal. The last property of the perfect normality of the $(3, 2)$-N-metrizable

space allows us to make the constructed Hilbert space to be metrizable if $\rho = \Delta$, even more so that it is separable. First, we construct the space.

Proposition 3.1. The Hilbert's space is $(3, 2, \rho)$ -N-symmetrizable space via a $(3, 2, \rho)$-symmetric d.

Proof: Let $M = H$ be the set of all infinite sequences $\{x_i\}_{i=1}^{\infty}$ of real numbers satisfying the condition $\sum_{i=1}^{\infty} x_i^2 < \infty$ and let $d(x, x, y) = D_d(x, y)$, where D_d is defined in Remark 2.2.

We shall show that with
$$D_d(x, y) = \sqrt{\sum_{i=1}^{\infty}(x_i - y_i)^2},$$
for $x = \{x_i\}_{i=1}^{\infty}$, $y = \{y_i\}_{i=1}^{\infty}$, a $(3, 2, \rho)$ -symmetric d is defined.

First of all, we will prove that D_d is well-defined, i.e., that the series in the definition of D_d is convergent. In the proof we shall apply the Cauchy inequality
$$\left|\sum_{i=1}^{k} a_i b_i\right| \leq \sqrt{\sum_{i=1}^{k} a_i^2} \cdot \sqrt{\sum_{i=1}^{k} b_i^2},$$
that holds for all finite sequences $a_1, a_2, \cdots, a_k$ and $b_1, b_2, \cdots, b_k$ of real numbers.

Let us note that for every pair of points $x = \{x_i\}_{i=1}^{\infty}$, $y = \{y_i\}_{i=1}^{\infty}$ in H and any positive integer k we have
$\sum_{i=1}^{k}(x_i - y_i)^2 = \sum_{i=1}^{k} x_i^2 - 2\sum_{i=1}^{k} x_i y_i + \sum_{i=1}^{k} y_i^2$

$\leq \sum_{i=1}^{k} x_i^2 + 2\sqrt{\sum_{i=1}^{k} x_i^2} \cdot \sqrt{\sum_{i=1}^{k} y_i^2} + \sum_{i=1}^{k} y_i^2$

$= \left(\sqrt{\sum_{i=1}^{k} x_i^2} + \sqrt{\sum_{i=1}^{k} y_i^2}\right)^2$

$\leq \left(\sqrt{\sum_{i=1}^{\infty} x_i^2} + \sqrt{\sum_{i=1}^{\infty} y_i^2}\right)^2$.

Since the last inequality holds for any positive integer k, the series in the definition of D_d is convergent and $D_d(x, y)$ is well-defined.

Obviously, D_d satisfies conditions $(M0)$, (P) and (Ms). In addition, we shall show that condition $(M2)$ is also satisfied.

Let $x = \{x_i\}_{i=1}^{\infty}$, $y = \{y_i\}_{i=1}^{\infty}$ and $z = \{z_i\}_{i=1}^{\infty}$ be any points of H, let
$$x^k = \{x_1, x_2, \cdots, x_k, 0, 0, \cdots\},$$
$$y^k = \{y_1, y_2, \cdots, y_k, 0, 0, \cdots\},$$
$$z^k = \{z_1, z_2, \cdots, z_k, 0, 0, \cdots\}$$
and
$$a_i = x_i - y_i, \quad b_i = y_i - z_i, \quad c_i = x_i - z_i.$$
By the Cauchy inequality we have
$$|D_d(x^k, z^k)|^2 = |D_d(z^k, x^k)|^2$$
$$= \sum_{i=1}^{k} c_i^2 = \sum_{i=1}^{k}(a_i + b_i)^2$$
$$= \sum_{i=1}^{k} a_i^2 + 2\sum_{i=1}^{k} a_i b_i + \sum_{i=1}^{k} b_i^2$$

$\leq \sum_{i=1}^{k} a_i^2 + 2\sqrt{\sum_{i=1}^{k} a_i^2} \cdot \sqrt{\sum_{i=1}^{k} b_i^2} + \sum_{i=1}^{k} b_i^2 =$

$\left(\sqrt{\sum_{i=1}^{k} a_i^2} + \sqrt{\sum_{i=1}^{k} b_i^2}\right)^2$

$< 9\left(\sqrt{\sum_{i=1}^{k} a_i^2} + \sqrt{\sum_{i=1}^{k} b_i^2}\right)^2/4$

$= 9|D_d(x^k, y^k) + D_d(y^k, z^k)|^2/4.$

From the last inequality it follows that for $k = 1,2,\ldots$ we have

$$3(D_d(x,y) + D_d(y,z))/2$$
$$\geq 3(D_d(x^k,y^k) + D_d(y^k,z^k))/2$$
$$\geq D_d(x^k,z^k),$$

and this implies that

$$3(D_d(x,y) + D_d(y,z))/2 \geq D_d(x,z).$$

From the last inequality we have that $D_d(x,y) = \sqrt{\sum_{i=1}^{\infty}(x_i - y_i)^2}$, defines $(3,2,\rho)$-symmetric on H. We also have the topology τ, induced by the metric d, in literature known as l^2-topology so the topological space (H,τ) where $\tau = l^2$, is a Hilbert $(3,2,\rho)$-N-symmetrizable space via a $(3,2,\rho)$-symmetric d. Moreover, if $\rho = \Delta$ the space (H,τ) is metrizable. ∎

The set of all sequences $\{x_i\}_{i=1}^{\infty}$, where all the x_i's are rational numbers only finitely many of which are distinct from zero, is dense in (H,τ) and countable, so the Hilbert's $(3,2,\rho)$-N-symmetrizable space is also separable.

IV. Conclusion

Under the assumption that on the set H a $(3,2,\rho)$- metric d is defined, we considered the induced topologies $\tau(N,d)$ and $\tau(D,d)$. Due to the symmetric properties of metric d, we proved that these topologies are equal. In addition, we constructed the $(3,2,\rho)$-N-symmetrizable space Hilbert space, by showing that conditions mentioned in II are satisfied. Moreover, this space is a metrizable and also a separable Hilbert space.

References

[1] П.С. Александров, В.В.Немыцкий, "Условия метризуемости топологических пространств и аксиома симетрии", Мат. сб. 3:3, p.p. 663-672, 1938.

[2] А.В. Архангельский, "О поведении метризуемости при факторных тображениях", ДАН 164, №o 2, p.p. 247-250, 1965.

[3] М. Чобан, "О метрисуемыих пространствах", Вестн.Моск. Ун-та, сер.Матем., мех, №o 3, p.p. 44-50, 1959.

[4] S. Čalamani, D. Dimovski, "Topologies induced by $(3,1,\rho)$-metrics and $(3,2,\rho)$-metrics", Inernational mathematical forum, Vol.9, no.21-24, p.p. 1075-1088, 2014

[5] S. Čalamani, D. Dimovski, "On continuity of a $(3,1,\rho)$-metric", Mathematical Bulletin Vol.38(LXIV) No.1, p.p. 5-1, 2014.

[6] S. Čalamani, T. Dimovski, D. Dimovski, "Separation properties for some topologies induced by $(3,j,\rho)$-metrics, $j \in \{1,2\}$", FMNS, Vol.1, p.p. 24-30, 2015.

[7] S. Čalamani, D. Dimovski, M. Seweryn-Kuzmanovska, "On $(3,2,\rho)$-E-K-metrizable spaces" Mathematical Bulletin Vol 40(LXVI) No3, p.p. 43-49. 2016.

[8] B.C. Dhage, "Generalized metric spaces and topological structure II", Pure Appl.Math. Sci., 40 (1-2), p.p. 37-41, 1994.

[9] D. Dimovski, "Generalized metrics - (n,m,r)-metrics", Mat. Bilten, 16, Skopje, p.p. 73-76, 1992.

[10] D. Dimovski, "$(3,1,\rho)$-metrizable topological spaces", Math. Macedonica, 3, p.p. 59-64, 2005.

[11] R. Engelking, "General Topology", Warsaw, 1977.

[12] S. Gähler, "2-metrische räume und ihre topologische Struktur", Math. Nachr. 26, p.p. 115-148 1963.

[13] Z. Mamuzič, "Abstract distance and neighborhood spaces", Proc. Prague Symp., p.p. 261-266, 1962.

[14] K. Menger, "Untersuchungen über allgemeine Metrik", Math. Ann. 100, p.p. 75-163, 1928.

[15] Z. Mustafa, B. Sims, "A new approach to generalized metric spaces", Journal of Nonlinear and Convex Analysis, Vol. 7, Number 2, p.p. 289-297, 2006.

[16] V. Nemytzki, "On the „third axiom of metric spaces„", Tr. Amer. Math. Soc. 29, p.p. 507-513, 1927.

[17] С. Недев, "o-Метризуемые пространства", Тр. Моск. Мат. Общ. Том 24, p.p. 201-236, 1971.

[18] J. Usan, "<Nm,E>-seti s (n+1)-rastojaniem", Review of Research, PMF, Novi Sad, Ser. Mat. 17, 2, p.p. 65-87, 1989.

On Groups of Cohomologies on a Locally Compact Topological Space

Elena Kotevska[1] and Sonja Chalamani[1]

Abstract – **In this paper we will present an example of a group of cohomologies on a locally compact topological space. Moreover, the properties of this group of cohomologies will be examined and proved in detail.**

Keywords – **Topological space, compact set, $p-$measurable function, $p-$measurable coboundary, group of cohomologies.**

I. INTRODUCTION

Let X be a topological space, and let G be an Abelian group. In [1] we introduced the Abelian group $\Phi^p(X, G)$ of p-functions and we constructed the homeomorphism

$$\delta: \Phi^p(X, G) \to \Phi^{p+1}(X, G).$$

This homomorphism is defined for every $p \geq 0$ such that for $\varphi \in \Phi^p(X, G)$ the corresponding $\delta\varphi \in \Phi^{p+1}(X, G)$ is given by

$$(\delta\varphi)(x_0, x_1, \cdots, x_p, x_{p+1})$$
$$= \sum_{i=0}^{p+1} (-1)^i \varphi(x_0, x_1, \cdots, x_{i-1}, \hat{x}_i, x_{i+1}, \cdots, x_{p+1}),$$

where "^" signs that we dropped the marked coordinate. Moreover, we proved the following properties

(1) $\Phi_0^p(X, G) \subset \Phi_C^p(X, G)$
(2) $\Phi_0^p(X, G), \Phi_C^p(X, G)$
(3) $\delta\left(\Phi_0^p(X, G)\right) \subseteq \Phi_0^{p+1}(X, G)$
(4) $\delta\left(\Phi_C^p(X, G)\right) \subseteq \Phi_C^{p+1}(X, G),$

for the sets

$\Phi_0^p(X, G) = \{\varphi | \varphi \in \Phi^p(X, G), |\varphi| = \emptyset\}$ and
$\Phi_C^p(X, G) = \{\varphi | \varphi \in \Phi^p(X, G), |\varphi| = compact\}$
and

$$\Phi_0^p(X, G) = \{\varphi | \varphi \in \Phi^p(X, G), |\varphi| = \emptyset\},$$

where $|\varphi|$ is the ***support of p-function***.

Lastly, we proved that the quotient spaces

[1]Elena Kotevska and Sonja Chalamani are with the Faculty of Technical Sciences at the University st. Kliment Ohridski, 7000 Bitola, Republic of Macedonia,
E-mail: elena.kotevska@tfb.uklo.edu.mk
        scalamani@yahoo.com

$$C_C^p(X, G) = {\Phi_C^p(X, G)} \Big/ {\Phi_0^p(X, G)} \text{ and }$$
$$C^p(X, G) = {\Phi^p(X, G)} \Big/ {\Phi_0^p(X, G)}$$

inherit the properties of the Abelian mapping that a homeomorphism from the Abelian group $\Phi^p(X, G)$.

II. DEFINITION OF GROUPS OF COHOMOLOGIES

We define the sets

$$Z^p(X, G) = \{u | u \in C_C^p(X, G), \delta u = 0\}$$

$$B^p(X, G) = \delta\left[C_C^{p-1}(X, G)\right].$$

Since δ is a homomorphism between groups, it immediately follows that the sets $Z^p(X, G)$ and $B^p(X, G)$ are subgroups of the group $C_C^p(X, G)$. Moreover, these sets are actually the Kernel and the Image of the corresponding homomorphisms δ. They are called groups of $p-$**measurable cocycles** and $p-$**measurable coboundaries** (of topological space X with coefficients in the group G) correspondingly. It can be shown that for a homomorphism δ, $\delta \circ \delta = 0$ is valid ([1]). By this we immediately have that the group of $p-$measurable coboundaries $B^p(X, G)$ is a subgroup of the group of $p-$measurable cocycles $Z^p(X, G)$. This means that we can define the quotient group

$$H_C^p(X, G) = {Z^p(X, G)} \Big/ {B^p(X, G)}.$$

This group is called $p-$**measurable group of cohomologies** with compact supports on topological space X, with coefficients in the group G. In this paper, we envisage the groups $H_C^p(X, G)$, for a locally compact topological space X. If we consider a compact topological space, then the group of cohomologies is denoted by $H_C^p(X, G)$. In such case, we also write $\Phi^p(X, G) = \Phi_C^p(X, G)$.

It is a general impression, at least to this level of presentation of definitions, that they are somehow unnatural and just scattered without any real meaning. However, as one digs deeper into the material, one discovers the importance and applications of theory of cohomologies, as well as it's not so apparent naturalness of the abovementioned definitions. Groups of cohomologies can be introduced, i.e., defined in lots

of different ways. For compact topological spaces with „good" local behavior, different definitions lead to the same result. For „pathological" spaces however, different definitions can lead to the different results. Considering that each definition is based on a set of different assumptions they all have some advantages or/and drawbacks with respect to the other. In this presentation, we have chosen the approach that has the advantage of being general and yet simple. By this approach, we arrive very quickly to the subject of interest and study. In the following, we will present a very simple example of group of cohomologies.

III. AN EXAMPLE OF GROUPS OF COHOMOLOGIES

We notice that for any topological space X and any $p < 0$

$$H_C^p(X, G) = 0.$$

Proposition 1: Let the topological space X consist of exactly one point, i.e., let $X = \{a\}$. Then

$$H_C^p(X, G) = 0, p > 0$$

and

$$H_C^0(X, G) \cong G.$$

Lemma 1: Let the topological space X consist of exactly one point, i.e., let $X = \{a\}$. Then

1. $\Phi^p(X, G) \cong G$
2. $\Phi_0^p(X, G) = \{0\}$
3. $C_C^p(X, G) = C^p(X, G) = \Phi^p(X, G)$

Proof:

1. Let $p \geq 0$ be any given, but fixed natural number. Then the set of all $p + 1$ −taples $(x_0, x_1, \cdots, x_p)$, where $x_i \in X$ and $i = 0, \cdots, p$ has exactly one element $(a, a, \cdots, a)$ with a showing $p+1$ times. We can conclude that X^p is a one-element set. So, for any $p \geq 0$ and any $p −$ function φ we can consider that

$$\varphi(x_0, x_1, \cdots, x_p) = s \in G,$$

i.e., we can say that every element $\varphi \in \Phi^p(X, G)$ maps the $p + 1$ −tupple $(a, a, \cdots, a)$ to an unique element $s \in G$. This way every element $\varphi \in \Phi^p(X, G)$ for any $p \geq 0$ determines a unique element of the group G. The opposite is also valid, i.e., each element of $s \in G$ determines a unique element $\varphi \in \Phi^p(X, G)$. This means that we have a bijection between groups $\Phi^p(X, G)$ and G given by $f: G \rightarrow \Phi^p(X, G)$

$$f(s) = \varphi \Leftrightarrow \varphi(a, a, \cdots, a) = s$$

We set $\varphi \equiv \varphi_s$. The mapping f is also a homomorphism. Indeed, for any $s_1, s_2 \in G$ we have

$$f(s_1 + s_2)(a, a, \ldots, a) = \varphi_{s_1+s_2}(a, a, \cdots, a) = s_1 + s_2$$
$$= \varphi_{s_1}(a, a, \cdots, a) + \varphi_{s_2}(a, a, \cdots, a)$$
$$= (f(s_1) + f(s_2))(a, a, \ldots, a).$$

This yields that the mapping f is an isomorphism, q.e.d.

2. Let $\varphi \in \Phi_0^p(X, G)$. Then the support $|\varphi| = \emptyset$ and for the element $a \in X$, $a \notin |\varphi|$. This means that there is a neighborhood V of a, $a \in V$, such that for every $(x_0, x_1, \cdots, x_p) \in V$, $\varphi(x_0, x_1, \cdots, x_p) = 0$. Actually, we have $\varphi(a, a, \cdots, a) = 0$ with $X = \{a\}$. It follows that $\varphi = \mathbf{0}$. This shows that $\Phi_0^p(X, G) \subseteq \{0\}$. The opposite is always valid, so it follows that $\Phi_0^p(X, G) = \{0\}$.

3. It follows from 2, that

$$C^p(X, G) = \left. \Phi^p(X, G) \middle/ \{0\} \right. = \Phi^p(X, G) \cong G.$$

On the other hand, $\Phi_C^p(X, G) = \Phi^p(X, G)$, since a topological space of one point is always compact. It follows that

$$C_C^p(X, G) = \left. \Phi_C^p(X, G) \middle/ \{0\} \right. = \Phi^p(X, G) \cong G, \text{q.e.d.}$$

We now proceed with the prove of the Proposition.
Let $p > 0$ be any natural number. Then, for any $p −$ function $\varphi \in \Phi^p(X, G)$, we have

$$(\delta(\varphi))(a, a \ldots, a) = \sum_{i=0}^{p+1} (-1)^i \varphi(a, a, \ldots, a, \hat{a}, a \ldots, a),$$

where $\hat{a}$ is on position with index i.
We will consider two cases:

1. Let $p > 0$ be an even number. Then, $p + 1$ is odd and on the right side we have even number of additions and so the sum will be zero. This means that

$$(\delta(\varphi))(a, a \ldots, a) = 0, \text{i.e.,} \delta(\varphi) = \mathbf{0}.$$

2. Let $p > 0$ be an odd number. Then, $p + 1$ is even and on the right side we have odd number of additions. This means that

$$(\delta(\varphi))(a, a \ldots, a) = -\varphi(a, a, \ldots, a),$$

where on the left we have $p + 1$ −tupples and on the right we have p −tupples. The nature of the groups considered gives us

$$\delta(\varphi) = -\varphi,$$

because $\Phi_0^p(X, G) = \{0\}$ and $C_C^p(X, G) = \Phi^p(X, G)$.
As a result of the above we get:

1. For even $p > 0$
$\delta: C_C^p(X, G) \rightarrow C_C^{p+1}(X, G)$ is such that $\delta(\varphi) = \mathbf{0}$, for every $\varphi \in C_C^p(X, G)$ and so we have that

$$Z^p(X, G) = C_C^p(X, G).$$

On the other side $p-1$ is odd and thus we have $\delta(\varphi) = -\varphi$, for every $\varphi \in C_C^{p+1}(X,G)$. It follows that

$$\delta\big[C_C^{p-1}(X,G)\big] = C_C^p(X,G),$$

i.e.,

$$B^p(X,G) = C_C^p(X,G).$$

Consequently we get

$$H_C^p(X,G) = Z^p(X,G) \Big/ B^p(X,G)$$
$$= C_C^p(X,G) \Big/ C_C^p(X,G) = \{\mathbf{0}\}.$$

2. For odd $p > 0$, we have that for every $\varphi \in C_C^p(X,G)$, $\delta(\varphi) = -\varphi$, i.e., we have

$$Z^p(X,G) = \{\varphi \in C_C^p(X,G), | \, \delta(\varphi) = \mathbf{0}\} = \{\mathbf{0}\}.$$

On the other side $p-1$ is even and thus we have $\delta(\varphi) = 0$, for every $\varphi \in C_C^{p-1}(X,G)$. It follows that

$$B^p(X,G) = \delta\big[C_C^{p-1}(X,G)\big] = \{\mathbf{0}\}.$$

It follows that

$$H_C^p(X,G) = Z^p(X,G) \Big/ B^p(X,G) = \{\mathbf{0}\} \Big/ \{\mathbf{0}\} = \{\mathbf{0}\}.$$

If $p = 0$, then since it is considered an even number, for

$$\delta: C_C^0(X,G) \to C_C^1(X,G)$$

we have

$$Z^0(X,G) = C_C^0(X,G) = G$$

and

$$B^0(X,G) = \{\mathbf{0}\}.$$

It follows that

$$H_C^0(X,G) = Z^0(X,G) \Big/ B^0(X,G) = G \Big/ \{\mathbf{0}\} = G.$$

This proves the Proposition 1.

IV. CONCLUSION

Following the definitions, structure and the general approach in [1], we presented a way to define groups of cohomologies. Then, in order to give life to these definitions, we've presented an example of such a group of cohomology. For our example we chosed a simple compact topological space of only one point. Following the sets and definitions we constructed the group of cohomologies $H_C^p(X,G)$ for this topological space. Moreover, we have given exact calculation of the coboundary operator δ. By doing this we are trying to give "life" to seemingly sterile and unnatural constructions.

REFERENCES

[1] E. Kotevska, S. Calamani, "On some properties of groups of p-cochains", International Scientific Journal HORIZONS Series B 2016, p.7-15.
[2] E. Kotevska "Cohomologies on a Locally Compact Topological Spaces", University St. Cyril and Methodius, Faculty of Natural Sciences, Institute of Mathematics, 1996
[3] S. McLane, Homology, Springer Verlag, 1963
[4] W. Massey, Homology and cohomology theory. an approach based on Alexander-Spanier cochains, M. Dekker,1978
[5] J. Neisendorfer, Primary homotopy theory, Mem. A.M.S. 232.
[6] D. Notbohm, Spaces with polynomialmod–p cohomology, Math. Proc. Camb. Phil. Soc.,1999

Microprocessor Development System Applying a Multifunctional Demonstration Board

Valentina Rankovska[1] and Stanimir Rankovski[2]

Abstract – **In this paper, a microprocessor development system based on a multifunctional demonstration board is presented. It is intended for educational purposes and to be controlled by various development boards with different microcontrollers and minimum size. The goal is to be comparatively "universal". The paper presents some experiments using Nucleo L476RGT6 development board.**

Keywords – **microprocessor development system, multifunctional demonstration board, Nucleo L476RGT6 board, software development tools.**

I. INTRODUCTION

The curriculum of the Bachelor's degree of Electronics at the Technical University of Gabrovo includes two disciplines related to microcontrollers and embedded systems design and application - Microprocessor Cicuits and Embedded Systems. Several microcontroller families based on different cores are being used: 8-bit general purpose microcontrollers with PIC and AVR processor cores and 32-bit microcontrollers with ARM processor cores. This implies the use of different types of hardware development tools, which in turn increases the costs and hampers students.

In general the hardware development tools for educational purposes could be divided into two major groups in relation to their complexity [1]:

- More complex development tools, containing a wide range of peripheral blocks - input devices, displays, interfaces, sensor blocks, etc.;

- Development boards almost without peripheral blocks, adapted to add various additional modules, shields, click-boards, etc.

In short, the advantages and disadvantages of the both approaches have been examined in [1]. Applying the second variant is motivated in the case of using several types of microcontrollers in training. It is supported by the presented in [1] multifunctional demonstration module, which includes commonly used blocks in the microprocessor systems. There is also a breadboard on it, allowing easy addition of various input/ output devices, sensor and actuator blocks, etc.

In the present paper, this approach is further extended by designing a microprocessor development system, including the multifunctional module and additional devices that are

[1]Valentina Rankovska is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, E-mail: rankovska@tugab.bg.

[2]Stanimir Rankovski is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, E-mail: s.rankovski@gmail.com.

inappropriate to be implemented on the same PCB due to an increase in overall dimensions.

Since the element of comparative versatility of the system is to be preserved, it is intended to be controlled by development boards with various microcontrollers with minimal form factor, minimum periphery and low cost.

The flexibility in implementing various microcontrollers and additional peripherals makes it possible to use the presented development system also in other disciplines of engineering education - digital circuits design, communication technologies, etc. [2], [3].

The presented results have been achieved using Nucleo L476RGT with a STM32L476RGT6 microcontroller and the integrated development environment Keil uVision5. Experiments with two other development boards - Arduino Uno and PIC32-PINGUINO OTG have been made, which integrated development environments are quite simpler.

In section II the components of the multifunction module are briefly described. In section III the suggested system is presented, and in section IV some sample algorithms and programs used in the testing of the system are given.

II. THE MULTIFUNCTIONAL DEMONSTRATION BOARD

The block diagram of the multifunctional demo-board is presented in Fig. 1 [1].
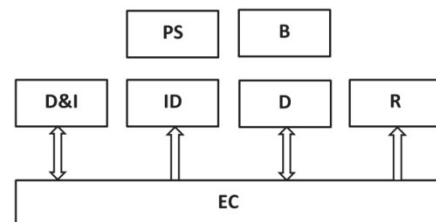
Fig. 1. Block diagram of the demo-board

PS   Power Supply
B   Breadboard
D&I   Display and Indication
ID   Input Devices
D   Drivers
R   Relays
EC   Expansion Connectors

It includes several commonly used I/O devices - five tactile push buttons, a rotary encoder, 10 LEDs, a 4-digit 7-segment display, and a 2x16 LCD display.

There is an 8-bit serial shift register SN74HC595N, which could be used to minimize the I/O pins of the microcontroller,

when implementing multiplexed LED display, a matrix LEDs, etc.

A relay block with two RAS-05-15 relays is provided, with a galvanic isolation of the load via optocouplers.

A motor driver block is also provided to control stepper and low power DC motors, relays, solenoids, etc. It consists of L293B IC - a quadruple high-current half-H driver, a stepper motor controller L297 and a high voltage, high current dual full-bridge driver L298.

The dual voltage power supply (+5V and +12V) is intended for the logic and the driver block.

The mini breadboard allows increasing the functions of the demo board by using a large number of additional modules.

All the I/O pins of the blocks are wired to expansion pin header strips in order to connect the desirable function with the pins of the microcontroller development board. This simplifies the wiring by avoiding a great number of configuring microswitches and jumpers, as it is in other development boards.

III. Nucleo Board and the Microprocessor Development System

Why ARM microcontroller?

As previously mentioned, the purpose of the development system is to use it for educational purposes in two disciplines. At Microprocessor Circuits 8-bit microcontrollers with a simpler CPU architecture are studied, as the students face the complex matter for the first time. Embedded Systems is being studied in the next semester, achieving two goals: expanding the knowledge and skills for designing and implementing embedded systems, and at the same time - the ability to study microcontrollers with more sophisticated CPU architecture such as ARM-based, which are widely spread too.

ARM Cortex is a wide set of popular 32/64-bit architectures and cores, divided into three main subfamilies:

- **Cortex-A** (**A**pplication) are processors for devices performing complex computing tasks and found in most of mobile devices, like phones and tablets.
- **Cortex-M** (e**M**bedded) is a range of energy efficient and easy to use processors designed for the low-cost embedded market.
- **Cortex-R** (**R**eal-Time) processors offer high-performance computing solutions for embedded systems where reliability, fault tolerance, maintainability and deterministic real-time response are essential.

Thus, Cortex -M based microcontroller is a very suitable choice to be studied.

Many manufactures offer a wide range of ARM microcontrollers: STMicroelectronics, Microchip, Maxim, Freescale, Holtek, Infineon, NXP, Silicon Labs, etc.

STMicroelectronics has a large portfolio of 32-bit Arm Cortex MCUs, namely STM32.

STM32 **advantages** [4]:

- Cortex-M based MCUs - a powerful alternative of 8-bit embedded microcontrollers, like PIC, MSP430, etc.
- Available completely free ARM based tool-chain;

- Pin compatibility - most of STM32 MCUs are designed to be pin compatible;
- 5V tolerant - they can interface other devices that do not provide 3.3V I/O, without using level shifters;
- Comparatively not expensive devices if the designer would like to migrate from 8/16-bit MCUs to a more powerful platform, while keeping a comparable target price;
- Integrated bootloader, allowing reprogramming the internal flash memory using some communication peripherals (USART, I²C, etc.).

Some **drawbacks**:

- Too many challenges to inexperienced users during the study process. If they are completely new to embedded development, they will suffer from the difficulties with the ST documentation;
- A lack of good documentation especially for its HAL;
- Lack of STMicroelectronics official development environment;
- Lack of MCUs for the IoT.

The Nucleo L476RGT6 development board

There are many development STM32-based boards on the market, including Arduino-compatible. For instance, the microcontroller on the STM32F103C8T6 board can even be programmed using Arduino IDE. But as the original STMicroelectronics boards like Discovery and Nucleo are extremely cheap, a Nucleo board has been used.

STM32 Nucleo boards are supplied with two sets of extension connectors, allowing easily applying hardware add-ons: Arduino compatible and ST morpho connectors.They also integrate an ST-Link debugger/programmer. They work with a wide range of development environments including IAR EWARM, Keil MDK-ARM, mbed and GCC/LLVM-based IDEs.

Fig. 2 shows Nucleo L476RGT board with STM32L476RGT6 microcontroller.
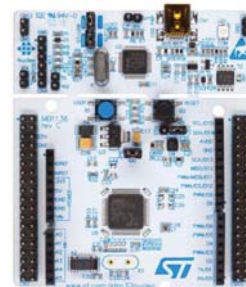
Fig. 2. Nucleo L476RGT6 development board

Except the above mentioned two sets of board expansion connectors, there are also 1 user LED, 1 user and 1 reset push-buttons, 32.768 kHz low speed external (LSE) crystal oscillator on the Nucleo L476RGT, flexible power-supply options: ST-LINK USB VBUS or external sources and on-board ST-LINK/V2-1 debugger/programmer. Also free software libraries and examples are available with the STM32Cube MCU Package [5].

The development system

The suggested development system is shown in Fig. 3. Except the control microcontroller board (in our case Nucleo L476RGT6) and in addition to the multifunctional demo-board, several often used peripherals have been added: a matrix keypad 4x3, a low power 5V DC motor, a stepper motor PM35L-048-HPD4 [6], a servo motor SG90 and a LED matrix 8x8.
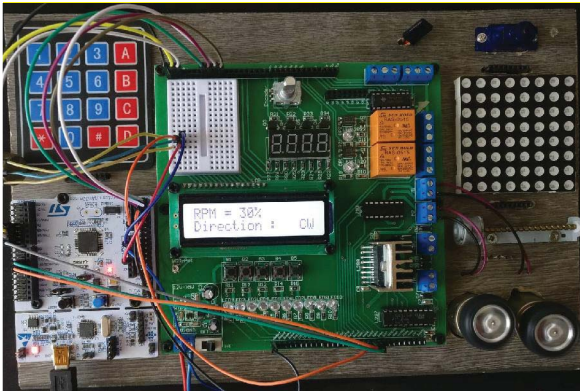
Fig. 3. The suggested development system with Nucleo L476RGT6

Nucleo is firmly mounted on the microprocessor system base, but it could be easily exchanged with other development boards which are compatible with the mounting holes: Arduino Uno, Olimex's PIC32-Pinguino-OTG with PIC32MX440F256H microcontroller [7], Netduino 3 with 168Mhz Cortex-M4 (STM32F4) of Wilderness Labs [8], etc. (Fig. 4). The system would operate with many other low size development boards, but they could be just put on or out of the base board.
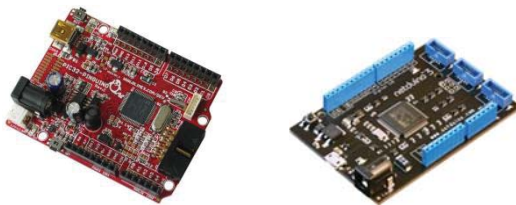
Fig. 4. Arduino Uno compatible development boards - PIC32-Pinguino-OTG and Netduino 3

IV. SOFTWARE DEVELOPMENT TOOLS AND SYSTEM VERIFICATION

The operation of the development system has been verified by simple programs for the input and output devices, motors and additional sensor blocks.

The STM32 toolchain is supported by three groups of software development tools:

- Tools to configure and generate code - STM32CubeMX;
- A comprehensive set of partner compile and debug IDEs: IAR EWARM (IAR Embedded Workbench), Keil MDK-ARM, GCC-based IDEs, ARM mbed online, etc.

- Monitor, program and Utilities tools: STM Studio, ST-Link utility, ST-MCU-Finder, etc.

STM32CubeMX

STM32CubeMX is a graphical tool for STM32 microcontrollers. It allows microcontroller and board selection from a list of STM boards, type of package, hardware peripherals, etc. and also an easy switching to another microcontroller. It generates after that the code necessary to configure the ST HAL (Hardware Abstraction Layer).

STM32CubeMX generates initialization C code and includes a utility tool for settings, like pin multiplexing, clock tree setting, peripheral configurations, etc.; generates IDE-ready projects for various integrated development environment toolchains; calculates the power consumption, etc.

Fig. 5 shows Nucleo L476RGT board selection. In Fig. 6 Pinout & Configuration window is shown. A sample program controls DC direction and the rmps and outputs data on the LCD. The necessary pins for the PWM, direction and LCD have been configured.
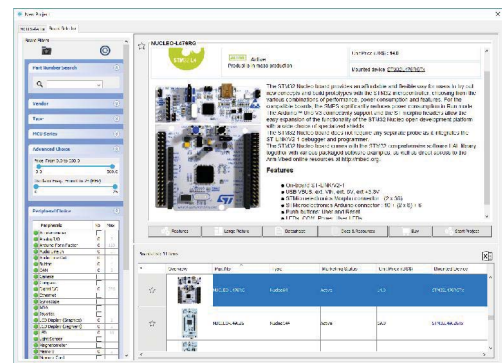
Fig. 5. Development board selection window

Fig. 6. STM32CubeMX Pinout & Configuration window

Fig. 7 shows timer and clock configuration for our example.

The Integrated Development Environment μVision

The ARM Keil products are the world leading software development tools for Cortex-M processor based devices [9]. So the Keil μVision5 IDE is used in the present work.

The μVision IDE combines project management, run-time environment, build facilities, source code editing, and
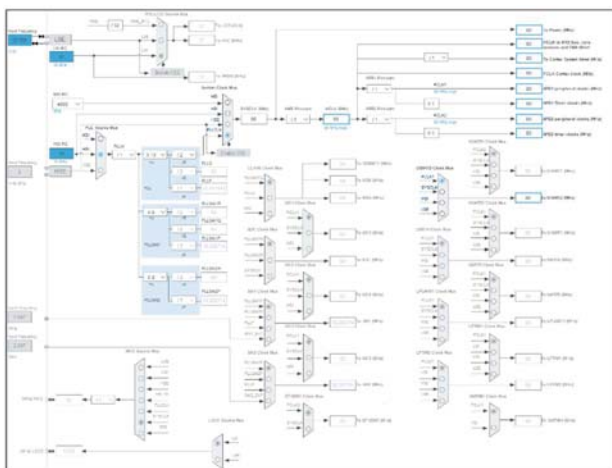
Fig. 7. STM32CubeMX Clock Configuration window

Fig. 9. Part of the sample program in the μVision 5 Editor window

program debugging utilities [10]. The debugger supports options, like breakpoints, watch windows, and execution control and full visibility to device peripherals. Fig. 8 shows the Project Manager and Run-Time Environment window.

using Nucleo L476RGT with 32-bit ARM-based STM32L476RGT6 microcontroller are presented.
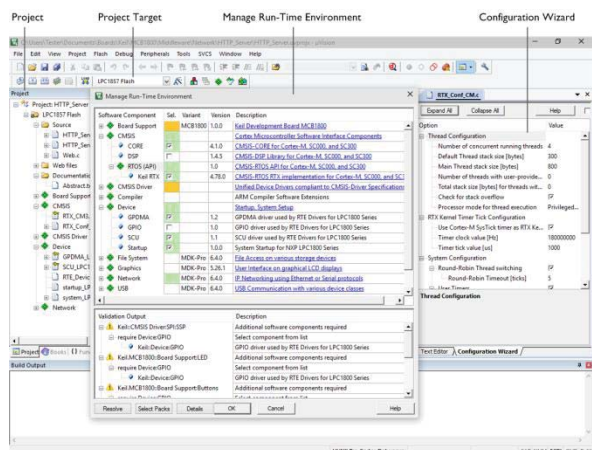
Fig. 8. μVision 5 Project Manager and Run-Time Environment window

Figs. 9 and 3 show respectively a part of the main function in a sample program for DC motor control and its operation. The whole project, including all the files, which could be seen in the picture, has been generated by the STM32CubeMX.

V. Conclusion

A microprocessor development system based on a multifunctional demonstration board is presented in the paper. It is to be used in Microprocessor circuits and Embedded systems courses at the Bachelor degree of Electronics, where various 8-, 16- and 32-bit microcontrollers are used, but could be useful in other disciplines too. Thus, the goal is the system to be comparatively "universal" and to be controlled by various microcontroller development boards. It has been tested with several development boards and the some results

REFERENCES

[1] V. Rankovska, S. Rankovski. "Multi-functional Demo Module for Microprocessor Development Systems for Educational Purposes". ET2018, Conference Proceedings, pp. 55-58, Sozopol, Bulgaria, 2018.
[2] Kandov, I., G. Goranov, A. Aleksandrov, J. Kanev. "Web Based System for Control and Measure Parameters of Sensor Network". ITHET 2014, Conference Proceedings, York, United Kingdom, 2014.
[3] G. Georgiev, I. Balabanova, P. Kogias, S. Sadinov and S. Kostadinova. Research Article Identification of Sine, Squire, Triangle and Sawtooth Waveforms with Uniform White and Inverse F Noises by Adaptive Neuro - Fuzzy Interface System. Journal of Engineering Science and Technology Review 11 (3), 2018, pp. 128 – 132.
[4] UM1718 User manual. STM32CubeMX for STM32 configuration and initialization C code generation. STMicroelectronics, 2019. (accessed: March 2019) https://www.st.com/content/ccc/resource/technical/document/user_manual/10/c5/1a/43/3a/70/43/7d/DM00104712.pdf/files/DM00104712.pdf/jcr:content/translations/en.DM00104712.pdf
[5] UM1724 User manual. STM32 Nucleo-64 boards. STMicroelectronics, 2017. (accessed: March 2019) https://www.st.com/content/ccc/resource/technical/document/user_manual/98/2e/fa/4b/e0/82/43/b7/DM00105823.pdf/files/DM00105823.pdf/jcr:content/translations/en.DM00105823.pdf
[6] https://www.eminebea.com/en/product/rotary/steppingmotor/pm/standard/pm35s-048.shtml
[7] https://www.olimex.com
[8] https://www.wildernesslabs.co/Netduino
[9] https://www.st.com/content/st_com/en/partner/partner-program/partnerpage/Arm_Keil.html#
[10] http://www2.keil.com/mdk5/uvision/

Enhancing Teaching of Wireless Communications by Combining Simulation, Measurements and Demonstration

Slavica Marinkovic[1], Amela Zekovic[2], Ivan Pavlovic[3], and Milutin Nesic[4]

Abstract – **This paper presents set of laboratory exercises that introduce students to antenna parameters, propagation effects and wireless link design. The students are introduced progressively to the topics which are of interest for the wireless communications from the point of view of applied studies. The exercises focus on linking theory and practice and are realized with limited resources, taking into account available equipment at the Electronics and Telecommunications Laboratory at VISER. The teaching is improved by combining different teaching methods such as simulation, measurements and practical demonstration. Students get practical experience with RF signals and measurements, antenna parameters and signal propagation effects.**

Keywords – **Teaching wireless communications, Practical teaching, Simulation, measurements and demonstration in wireless communications education.**

I. INTRODUCTION

Wireless communications are very important part of the telecommunication curriculum. Understanding fundamentals of this subject lays down foundation for understanding of numerous applications ranging from radio frequency identification and near filed communications to more sophisticated systems such as global positioning systems, wireless local area networks or cellular systems. Applied studies incorporate knowledge of all aspects of telecommunication systems as academic studies however with less involved mathematical foundation and with more practice oriented instruction. It is therefore of great importance to design laboratory exercises that support theory by illustrating, demonstrating, simulating or measuring and implementing real world examples. This is not always easy due to limited resources in terms of appropriate hardware, software and instruments. A number of papers deal with practical instruction in engineering education in the field of wireless communications. In order to emphasize the importance of the subject of wireless communication education and the need to pay special attention to practical instruction we here mention some of the work in this field. Although the purpose is the same (teaching and demonstrating concepts, facilitating learning and inspiring interest in the subject) our approach does not directly build on the work in the quoted references. We are oriented towards applied studies, introductory wireless lab classes, low cost solutions, use of available free software and special attention has been given to designing gradually more involving learning tasks. In [1] the authors present a Wireless Communication System Laboratory Course for graduate and senior level undergraduate students. This course covers broad range of subjects and is designed to maintain high students' motivation by including the use of popular wireless devices such as cell phones, notebook computers, Bluetooth devices. In [2] authors present teaching wireless communication and networking fundamentals using Wi-Fi projects. The authors have concluded that Wi-Fi projects are easy to use and set up for demonstrations, low cost, facilitate an interactive hands-on learning experience, useful and challenging in sense of motivating students to test their knowledge. When it comes to radio link design several papers propose various software tools [3], [4].

This paper presents four introductory laboratory exercises for the third year course in Mobile communications at the Electronics and Telecommunications (ELITE) study program at the School of Electrical and Computer Engineering of Applied Studies in Belgrade (VISER). The students are introduced step by step to subjects covering fundamental of RF signal measurements, antenna parameters and wireless channel propagation effects. The Mobile communication course has been introduced in 2007 in the curriculum. Over the years the laboratory sessions have evolved from pure software simulation covering the basic points of wireless link design to the contemporary laboratory that involves measurement equipment and practical realizations. Here we present this introduction to the subject that includes measurements of RF signal characteristics, examining antenna parameters, simulation and calculation of propagation losses and implementation of point to point Wi-Fi link.

The goal of the presented exercises is to approach topics in wireless communications from different perspectives and in different settings with progressively increasing complexity of the studied example. Pedagogically, this means that students are motivated to constantly review previously acquired knowledge and build on it which leads to better understanding and memorizing of theoretical knowledge. Also, since real world examples are demonstrated, the students are encouraged to integrate previously acquired knowledge in various communication courses.

[1]Slavica Marinkovic is with the School of Electrical and Computer Engineering of Applied Studies Belgrade, 283 Vojvode Stepe, 11050 Belgrade, Serbia (e-mail: slavica.marinkovic@viser.edu.rs)..

[2]Amela Zekovic is with the School of Electrical and Computer Engineering of Applied Studies Belgrade, 283 Vojvode Stepe, 11050 Belgrade, Serbia (e-mail: amelaz@viser.edu.rs).

[3]Ivan Pavlovic is with the School of Electrical and Computer Engineering of Applied Studies Belgrade, 283 Vojvode Stepe, 11050 Belgrade, Serbia (e-mail: ivanp@viser.edu.rs).

[4]Milutin Nesic is with the School of Electrical and Computer Engineering of Applied Studies Belgrade, 283 Vojvode Stepe, 11050 Belgrade, Serbia (e-mail: nesic@viser.edu.rs).

II. Introducing RF signals, noise and distortion measurement in the frequency domain

Although this topic has been covered in the second year telecommunication courses at ELITE study program, due to importance of the subject and need that students get working knowledge of measurements with spectrum analyzer instrument, this topic has been repeated in the third year Mobile communications course. In the telecommunications courses preceding this courses software, combined, and hardware realization of spectral analysis laboratory sessions have been introduced in order that students master this complex and important subject. Details of this have been presented in [5]. In Mobile communications course students revise and expand their knowledge. The exercise consists of connecting signal generator and spectrum analyzer with coaxial cable and performing measurements. In the first part of measurements students use SMA and in the second part BNC connectors respectively. Frequency of the RF carrier (f_{c_in}) is measured, as well as the signal (L_{c_in}) and noise (N) levels. Based on the measurements the carrier frequency error (FE) is calculated (relative to the frequency f_{c_out} set by the signal generator). The RF carrier frequency (f_{c_out}) is varied and the output signal level (L_{c_out}) is kept constant. The attenuation of the cables and connectors (ΔL) is calculated.

TABLE I
RESULTS FOR THE RF MEASUREMENTS EXPERIMENT

f_{c_out} [Hz]	100 M	300 M	3 G	10 G
L_{c_out} [dBm]	0	0	0	0
f_{c_in} [Hz]	100000052.1	300000243.2	3000002426.1	10000007980.5
L_{c_in} [dBm]	−0,77	−1,40	−5,88	−13,08
N [dBm]	−75	−75	−70	−70
ΔL [dB]	0,77	1,40	5,88	13,08
FE [ppm]	0,521	0,811	0,801	0,798

The results are reported in the form of tables and graphs. Table I shows results obtained for the RF carrier measurements and SMA connector.

TABLE II
RESULTS FOR THE HARMONIC DISTORTION MEASUREMENTS

L_{out} [dBm]	L_{in} [dBm]	ΔL_{2nd} [dB]	ΔL_{3rd} [dB]
0	−2,56	−52,24	−80,85
2	−0,56	−52,67	−77,61
4	1,45	−52,37	−73,02
6	3,49	−52,50	−80,24
8	5,50	−52,51	−75,85
10	7,50	−50,37	−72,44

Students observe cable and connector losses as a function of frequency. The differences between BNC and SMA connector losses are becoming distinct at higher frequencies.

Subsequently students perform signal harmonic distortion measurements. An example of the harmonic distortion measurement, that is, the level of the second and third harmonic components and the fundamental is shown in Figure 1. This figure shows image captured by spectrum analyzer for RF carrier frequency 300 MHz and signal level (L_{out}) of 14 dBm. Students further change output signal level and record level
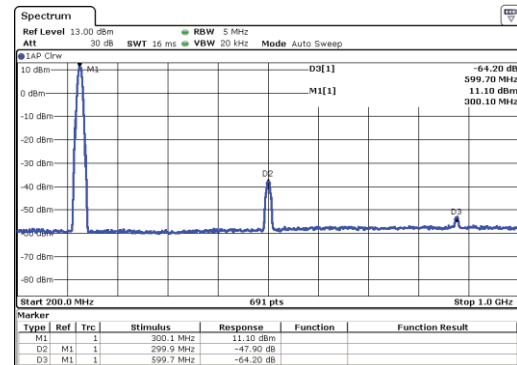
Fig. 1. Harmonic distortion measurements

differences between the harmonics and the fundamental. This is shown in Table II. The results were obtained with signal generator Agilent E8257D and spectrum analyzer Rohde Schwarz FSV.
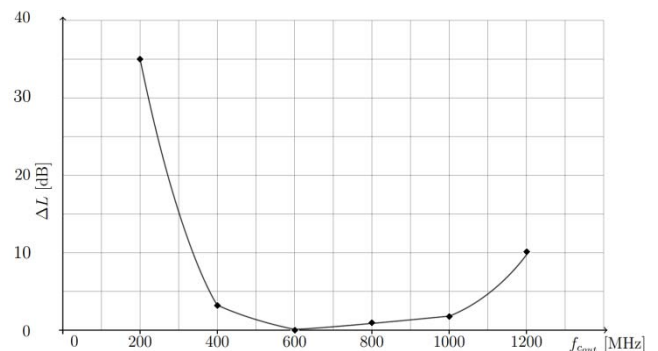
Fig. 2. Results for the antenna bandwidth estimation

III. Examining design and antenna characteristics

The goal of this exercise is that students get familiar with antennas and their characteristics which is achieved by measurements using signal generator, spectrum analyzer and printed log-periodic antennas. In the introductory part students are instructed about general antenna design and parameters as well as particularities of printed log periodic antennas. The first task consists of estimating antenna bandwidth. This is done by connecting one antenna to signal generator and using it as a transmitter and the second antenna to spectrum analyzer and using it as a receiver. The transmitting frequency is changed for constant signal level and the received signal level is measured.
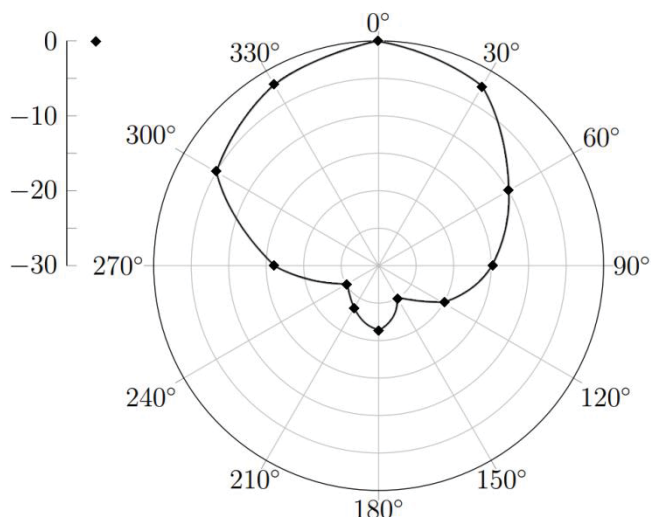
Fig. 3. Estimated antenna radiation pattern

The obtained results are shown in Figure 2 where normalized received signal level with respect to frequency is plotted.

The second part of the exercises cosniders estimating antenna radiation pattern by using the same setup. The receiving antenna is rotated and normalized antenna pattern is sketched. An example of obtained results is shown in Figure 3.

IV. CREATING AND INVESTIGATING CHARACTERISTICS OF COMMUNICATION SYSTEM IN RADIO MOBILE SOFTWARE

The goal of this exercise is to introduce students to the use of Radio Mobile software tool [6] for investigating characteristics of wireless communications system. The laboratory exercise includes investigating radio link equipment characteristics and electromagnetic waves propagation effects. The students observe the impact of the transmitter power, antenna gain and radiation pattern, terrain profile on the performance of the communication link including Fresnel zones. This exercise demands that students are familiar with data needed to perform a basic link performance such as coordinates of the transmitter and receiver, radiated power, operating frequency band, antenna radiation pattern as well as with theoretical foundations of the wireless link budget analysis. Some of these parameters were introduced in the previously described lab sessions.

The example in Figure 4 shows the terrain profile between two locations that were chosen for the point to point link that will be realized and examined in the next laboratory exercise. The figure shows line of sight as well as Fresnel zones and predicted values for the path loss and received power for the realistic parameters. The simulation takes into account terrain profile and detailed data on clutter is not taken into account.
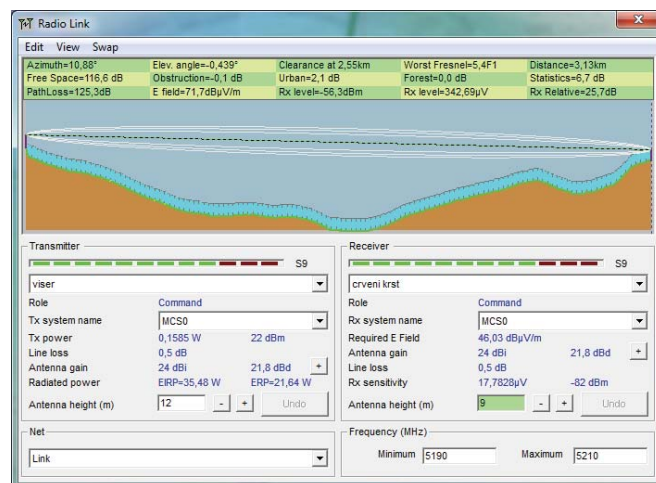
Fig. 4. Results for the radio link simulation in Radio Mobile software

The antenna parameters and transmitter power as well as receiver sensitivity were taken from the manufacturer's web site [7]. The coverage map when the transmitter is at VISER school building is shown in Figure 5. The students can observe that we use highly directional antennas and that for point to point links antennas have to be aligned. Also due to this property influences of other sources of interference are limited.
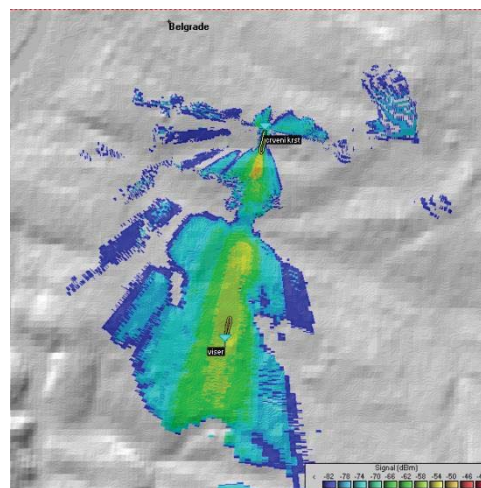
Fig. 5 Coverage map obtained by the simulation in the Radio Mobile software

V. VERIFYING WIRELESS SYSTEM DESIGN IN PRACTICE

In this lab session students get hands on experience with a practical Wi-Fi point to point link realized between VISER school building and the building that is 3,12 km away. Students first estimate path losses for particular antennas and their locations with Radio Mobile software and compare it to measurements obtained by the MikroTik LHG 5 802.11 wireless device with an integrated dual polarization grid antenna [7].
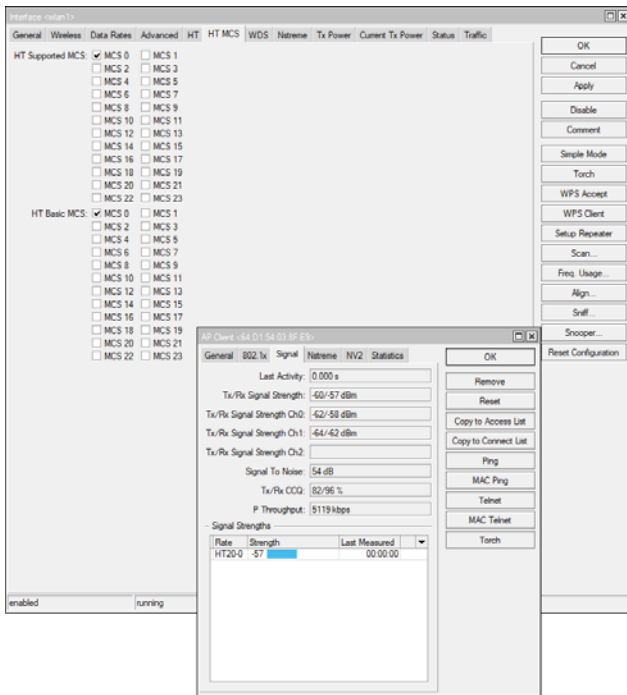
Fig. 6 Results for the received signal strength obtained from the wireless MikroTik LHG 5 802.11 device
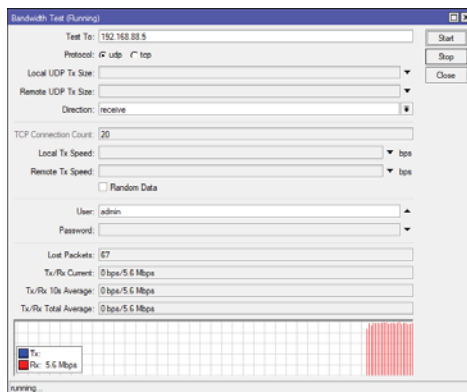
Fig. 7. Estimated link bitrate obtained from the wireless MikroTik LHG 5 802.11 device

Students are motivated to integrate their previously acquired knowledge and apply it to real world example. Figure 6 shows obtained received power results for the modulation and coding scheme that uses binary phase shift keying and rate 1/2 convolutional code. This is obtained from the MikroTik LHG wireless device measurements. Students can observe that there is difference in simulation results and results obtained by experiment.

The values predicted by simulation are better than the ones obtained directly from the wireless device due to fact that environment is dense urban environment and simulation does not take into account datasets with clutter height.

Figure 7 shows results for the estimated bitrate. The obtained values are slightly lower than that given in 802.11n specifications. This laboratory session also requires that students are familiar with theory and parameters used in Wi-Fi standard.

VI. CONCLUSION

In this paper we have presented laboratory sessions that introduce students to topics in wireless communications step by step and in an easy to grasp way. They start from simple measurements and observations and end up with practical realization. The communication system elements and parameters are examined individually and as part of the complex system. Some aspects are reinvestigated but in a different setting which makes repetition of knowledge more attractive to students as well as more productive. Also in this way the integration of knowledge from various courses is facilitated. Although topics are covered from the point of view of applied studies and represent introduction to the subject they are fundamental and lend themselves to expansion to higher level of knowledge. This could be used either for broadening course topics or as topics for student projects.

REFERENCES

[1] S. Guzelgoz and H. Arslan, "A Wireless Communications Systems Laboratory Course," in IEEE Transactions on Education, vol. 53, no. 4, pp. 532-541, Nov. 2010.

[2] Sarkar, Nurul & M. Craig, Trevor. (2006). Teaching Wireless Communication and Networking Fundamentals Using Wi-Fi Projects. Education, IEEE Transactions on. 49. 98 - 104.

[3] Juan Llacer, L, Rodriguez, J V, Molina Garcia Pardo, J M, Pascual García, J, Martínez Inglés, M. RADIOGIS: Educational software for learning the calculation of radio electric coverage in wireless communication systems. Comput Appl Eng Educ. 2019.

[4] M. J. Madero-Ayora, M.A. Sarmiento-Vega, J.J. Murillo-Fuentes, L. Salamanca-Mino, Active Learning of Radiocommunication Systems with the Help of Radio Planning Tools., Proceedings of the 8th WSEAS International Conference on Engineering Education (EDUCATION '11)

[5] S. Marinković, A. Zeković, M. Štimac, "Različiti pristupi u praktičnoj nastavi u oblasti spektralne analize", INFOTEH-JAHORINA Vol. 13, March 2014.

[6] "Radio mobile", January 2018, Available online at: http://www.ve2dbe.com/english1.html

[7] MikroTik LHG5 802.11 device manufacturer's data: https://mikrotik.com/product/RBLHG-5nD

Transport Policy and electric road vehicles in the European Union – Trends and Impacts

Verica Danchevska[1], Zoran Joshevski[2] and Dejan Danchevski[3]

Abstract – **The importance of traffic policy and the complexity of the traffic system are the reason for state interventions in almost all segments and modes of traffic. For these reasons, the following should be added: interconnection of traffic and the national economy, high capital investments, especially in the large traffic infrastructure, very rapid development of traffic technical and technology, increased consumption of some poor and non-renewable resources (eg oil , arable land), endangering the quality of the environment, traffic safety, etc.**

Bearing in mind the direct and indirect effects of traffic in general, which means the traffic policy - macro and micro, it can be established that traffic policy is a fun

The rapid introduction of electric vehicles in the transport sector will be a major challenge for the operation of the distribution network. If the charging of electric vehicles is not under control, a large overload will occur in the distribution network, especially in areas where we expect a high density of charging points, such as clusters of family houses, garages, car parks and the rural part of the network. So essential are solutions that will minimize the need for additional investment in the network, and solutions that will reduce the negative effects on the electrical grid caused by electric vehicles and stations (devices) for their charging. The paper discusses the EU has taken various actions to support electric mobility. For instance, EU-level measures have been encouraging the use of renewable electricity and smart charging; helping to develop and standardise charging infrastructure; and supporting research on batteries. Local, regional and national-level incentives (such as the introduction of lower taxes or the provision of free public parking for electric vehicles) are also promoting electric mobility. Countries that offer generous incentives and good charging infrastructure typically have a bigger market share for electric road vehicles.

Keywords – **European Union, Transport policy, electric vehicles, distribution grid, smart meters, smart grids**

I. INTRODUCTION

With advances in technology and changes in society, the way people and goods travel from point A to point B on the road is evolving drastically. In general, cars and other vehicles are becoming increasingly efficient, cleaner, automated and connected. More and more people are open to car- pooling, car-sharing and ridesharing. As

demand for mobility rises, the transport network grows and becomes increasingly congested. Along with these trends, electric vehicles (whether hybrid or all- electric) are also gaining in popularity.

Electric vehicles are not a new invention. They were among the first vehicles in the 1800s, and by 1900 accounted for around a third of all vehicles on the road. However, battery and electrical network weaknesses and the rise of the cheap oil era, together with the technological advancement of diesel and petrol vehicles, shifted interest away from electric vehicles for many decades. Although all vehicles can be electric, this briefing focuses only on electric road vehicles (EVs).

With the cheap oil era drawing to an end and climate change posing an increasingly serious threat, EVs have once again come into the spotlight. All levels of governance within the EU have recognised that EVs are needed to make transport smarter and more sustainable. However, the EU market for EVs is still in its infancy and is largely dependent on support policies. The EU still needs to address problems, such as high upfront cost and limited offer of electric models compared to the conventional ones, as well as limitations of the electrical network, charging infrastructure and batteries. In broader terms, the development of the EV market also depends on the level of ambition of the EU emission regulations; the incentives offered to users of EVs; fuel prices; general travel behavior and advances in research.

II. A GROWING MARKET

The EU market is still dominated by petrol and diesel vehicles, but the share of electric vehicles is growing fast. According to the European Automobile Manufacturers' Association (ACEA), the market share of electric cars in the EU was about 2 % in the third quarter of 2018, around 30 % higher than in 2017. While most of these cars are in use in a few northern and western Member States, their largest sales growth in recent years has been registered in southern and eastern ones. In most Member States, hybrid car sales exceed fully electric car sales.
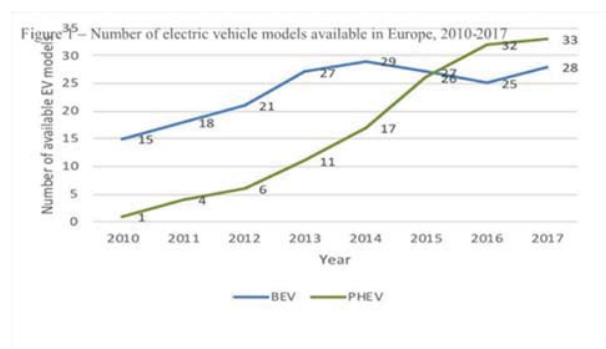

Figure 1. Number of electric vehicle models available in Europe, 2010-2017

[1]Verica Danchevska is with the Faculty of Technical Sciences at University St. Kliment Ohridski, Bitola Macedonia, E-mail: verica.dancevska@uklo.edu.mk .

[2]Zoran Joshevski is with the Faculty of Technical Sciences at University St. Kliment Ohridski, Bitola Macedonia, E-mail: zoran.josevski@uklo.edu.mk .

[3]Dejan Danchevski is with the Faculty of Technical Sciences at University St. Kliment Ohridski, Bitola Macedonia, E-mail: dejan.dancevski@uklo.edu.mk .

III. ENVIRONMENTAL AND HEALTH IMPACTS

A. Greenhouse gas emissions

Despite technological improvements, transport is still responsible for around one quarter of Europe's greenhouse gas (GHG) emissions, which are a substantial contributor to climate change. While in other sectors GHG emissions have gradually declined since 1990, those from transport only started to decrease in 2007 and still remain higher than in 1990 (see Figure 4). Within the transport sector, road transport is by far the biggest emitter, accounting for about 80 % of all EU GHG emissions from transport.

The EU is committed to reducing GHG emissions from transport and other sectors. Overall, its aim is to cut GHG emissions by at least 40 % below 1990 levels by 2030. For transport in particular, the aim is to cut GHG emissions by 60 % compared to 1990 levels by 2050.
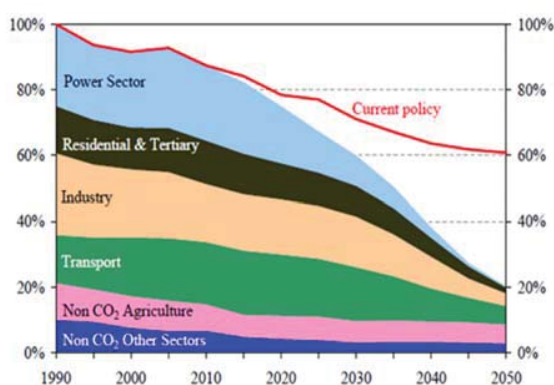
Figure 2 – EU emissions of greenhouse gases by 80% reduction by 2050

B. Air pollution

Transport is the main cause of air pollution in cities. Fossil fuels used in transport emit a range of substances, including nitrogen oxides (NOX) and particulate matter (PM), which inflict significant harm to air quality and human health; more than 400 000 citizens in the EU die prematurely each year and millions more have respiratory and cardiovascular diseases as a result of poor air quality.

Road transport is responsible for around 40 % of NOX emissions in the EU, most of which come from diesel-powered vehicles. The pollution from road transport is all the more harmful, as emissions released by vehicles occur close to the ground, often in areas where many people live and work.

Electric mobility could significantly reduce air pollution caused by transport, due to less fossil fuel use, especially if the electricity is produced from renewable sources. EVs emit no tailpipe CO_2 and have significantly lower NOX emissions than conventional vehicles. However, EVs still emit PM from road-, tyre- and brake wear. Besides, EVs are often heavier than equivalent conventional vehicles and vehicle weight tends to correlate with an increase in non-exhaust emissions of PM. In addition, electricity generation also produces pollution; however, since power stations are usually located away from cities, the pollution they cause has a lower impact on human health than pollution from conventional vehicles.

C. Noise

Transport is also a major source of noise pollution. High-levels of noise can lead to hearing loss, sleep disturbance, poor mental health and well-being, increased risk of heart disease and change in blood pressure, among other effects. According to 2018 data from the European Environment Agency, almost 88 million people living in urban areas and almost 41 million people living outside urban areas in the EU are exposed to road, rail and air traffic noise levels exceeding EU thresholds.

In general, EVs have lower noise levels than conventional vehicles, especially at lower speed, where the minimal noise they produce can even pose a risk for other road users if they do not hear the EV coming.3 However, on rural roads or motorways, where speeds are higher, the difference in noise levels is much smaller. The extent of noise reduction also depends on the proportion of EVs in the overall vehicle fleet. In brief, even though EVs do make some noise, if they are used mostly in urban areas and once their proportion in the overall vehicle fleet becomes significant, they can help to achieve noise levels that are less harmful for human health and well-being.

D. EU Support

Electric mobility is an area where the EU has shared competences with the Member States. The latter promote electric mobility through local, regional and national incentives such as lower taxes or free public parking for EVs. The EU complements these efforts by encouraging measures aimed at increasing resource efficiency and recycling, as well as measures to help break the oil dependency, optimise and improve the efficiency of the transport system, develop sustainable fuels, scale up the use of renewable electricity and remove obstacles to the electrification of transport. Often actions in support of electric mobility are part of wider measures taken with a view to developing a more sustainable transport system.

As part of the next long-term (2021-2027) EU budget, in June 2018 the Commission proposed to spend 60 % of the CEF €42.3 billion budget on projects that contribute to achieving climate objectives, for instance, through the development of charging infrastructure for EVs. The Parliament and the Council are now negotiating on the proposal. In its negotiating position (adopted in December 2018), the Parliament voiced its support for allocating 60 % of the CEF budget to projects contributing to climate action, while also calling for an increase of the overall CEF budget by almost €6 billion compared to the Commission's proposal.

E. Electrical network

Replacing millions of petrol and diesel vehicles with electric ones also means placing a bigger load on the electrical network that is already under pressure at certain times of the day in some areas. According to a 2018 JRC study, the EV is normally the biggest consumer of electricity in its owner's home. The same study calculates that if 15 % of the cars on EU roads were electric in 2030, this would place an extra demand on the electrical network of about 95 TWh per year. This would be about 3 % of total electricity consumption in the EU in 2030. However, by promoting charging at the most convenient times for the

electricity grid and at the lowest cost for consumers (smart charging), the EU can limit the costs needed for investing in the electricity grid and shorten the delays in the uptake of EVs. Some electricity grid reinforcements will still most likely be needed, especially for high-power recharging points. In the long run, EVs could also feed energy back to the electricity grid and even earn their owners money, if charging occurs during non-peak electricity times and selling back to the grid occurs during peak times.

To help to connect EV recharging points to the distribution network while increasing the share of renewable energy in the electricity grid, in 2016 the Commission proposed a directive setting new rules for the internal electricity market (on which the Council and the Parliament reached a provisional agreement in December 2018). In this proposal, the Commission called on Member States to create a framework that facilitates the connection of EV recharging points to the distribution network. It furthermore proposed that customers have access to electricity price comparison tools, smart charging, and dynamic electricity price contracts.

IV. Conclusion

It is hard to tell how the future of electric mobility will evolve, but it is clear that it holds a lot of potential to make mobility more sustainable and smarter. Electric road vehicles could reduce the EU's dependence on foreign oil, while lowering the pollution from transport. However, the extent to which EVs will effectively lower pollution will depend on their share in the overall vehicle fleet as well as on how environmentally friendly they remain during their whole life cycle.

Most likely, electric mobility will continue to blend with other trends such as digitalisation and connectedness, automation and shared mobility. Instead of owning an EV, people might choose to order an automated EV from a shared fleet with a click on their smartphone. This could also help to reduce congestion, especially in urban areas.

According to forecasts by the International Energy Agency, the market share of EVs in Europe could be around 23 % in 2030 when accounting for all road transport vehicles except two- and three- wheelers. Since a number of motor vehicle manufacturers have announced that they would stop launching new petrol and diesel models in the coming years and launch only fully electric or hybrid ones, it is likely that the offer of EVs will increase and their prices will drop. Range anxiety could diminish significantly with technological advancements, as the driving ranges of EVs get longer and charging times shorter. Motor vehicle manufacturers and governments are also continuing to develop their charging infrastructure for EVs and are helping customers to find the nearest recharging point (for instance via an app). However, whether these technological advancements will indeed convince customers to replace their diesel or petrol car with an electric one, will also depend on other factors, such as fuel prices.

The EU has an important role to play in supporting the transition to a more sustainable and smarter mobility. The more ambitious it is in its policies that drive vehicle technology

improvements and encourage the use of renewable energy and smart electricity networks, the more impressive the uptake of EVs is likely to be. The future of electric mobility will also be determined by incentives at the local, regional and national levels of governance. Namely, countries that offer generous EV incentives and good charging infrastructure, are typically observing a bigger increase in EVs than countries with low or no incentives. Whether they continue to see growth in the share of EVs will largely depend on how persevering they are in applying these incentives.

References

[1] Danchevska, V., "Transport Policy as a Function of Sustainable Development in the Economy of the Republic of Macedonia", Doctoral dissertation, 2005

[2] European Environmental Agency, Electric vehicles in Europe, November 2016.

[3] European Parliament, Charging infrastructure for electric road vehicles, June 2018.

[4] International Energy Agency, Global EV Outlook 2018, 2018.

[5] Joint Research Centre, Electric vehicles in Europe from 2010 to 2017: is full-scale commercialisation beginning?, 2018.

[6] A European Strategy for Low-Emission Mobility, Brussels, 20.7.2016. COM(2016) 501 final.

[7] A Roadmap for moving to a competitive low carbon economy in 2050, Brussels, 8.3.2011.

[8] COM(2011) 112 final

[9] http://ec.europa.eu/transport/themes/strategies/news/2016-07-20-decarbonisation_en.

[10] https://www.bloomberg.com/news/articles/2017-07-06/the-electric-car-revolution-is-accelerating.

[11] Facilitating e-mobility: EURELECTRIC views on charging infrastructure, March 2012.

[12] Žutobradić, S., Wagmann, L, Mihalek. E., „Radeka, I, Šagovac, G., „Istraživanje karakteristika opterećenja kućanstava na području grada Zagreba", Energija, 2001.

[13] „ITRES Tool for assessing the technical and economic impact of electric vehicles on distribution networks", Imperial College London, 26.01. 2014.

[14] European electricity industry views on charging Electric Vehicles, A EURELECTRIC concept paper, April 2011.

[15] SMART CHARGING: steering the change, driving the change, A EURELECTRIC paper, March 2015.

AUTHOR INDEX