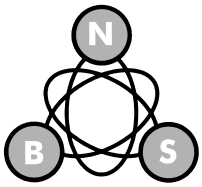
XLVI INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES



ICEST 2011

Proceedings of Papers

Volume 3

Serbia, Niš, June 29 - July 1, 2011

ICEST 2011 - XLVI INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES, Serbia, Niš, June 29 - July 1, 2011

Proceedings of Papers - Volume 3 of 3 volumes

Editor: Prof. Dr. Bratislav D. Milovanović

Technical Editor: Dr. Zoran Ž. Stanković

Technical Co-Editor: Dr. Biljana P. Stošić

Published by: Faculty of Electronic Engineering, University of Niš, Serbia
Printed by: UNIGRAF, Niš, Serbia

Number of copies printed: 80

Printing of this edition has been financially supported by Serbian Ministry of Science

ISBN: 978-86-6125-033-0

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

621.39(082)
537.8(082)
681.5(082)
004(082)
621.3(082)
37.02(082)

INTERNATIONAL Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST (46 ; 2011 ; Niš)
Proceedings of Papers. #Vol. #3 / XLVI International Scientific Conference on Information, Communication and Energy Systems and Technologies - ICEST 2011, Niš, June 29 - July 1, 2011 ; [organized by Faculty of Electronic Engineering, Niš [and] Faculty of Telecommunications, Sofia [and] Faculty of Technical Sciences, Bitola ; editor Bratislav D. Milovanović]. - Niš : Faculty of Electronic Engineering, 2011 (Niš : Unigraf).
- XVIII, 571-1028 str. : ilustr. ; 29 cm

Tiraž 280. - Bibliografija uz svaki rad. -
Registar.

ISBN 978-86-6125-033-0

a) Телекомуникације - Зборници b)
Микроталасна техника - Зборници c) Системи
аутоматског управљања - Зборници d)
Рачунарство - Зборници e) Електротехника -
Зборници f) Образовна технологија -
Зборници
COBISS.SR-ID 186468108

TABLE OF CONTENTS

VOLUME 1

ORAL SESSIONS

SIGNAL PROCESSING I

SIGNAL PROCESSING II

TELECOMMUNICATION NETWORKS AND SERVICES I

TELECOMMUNICATION NETWORKS AND SERVICES II

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS I

RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS II

METROLOGY AND REMOTE SENSING

TELECOMMUNICATION SYSTEMS AND TECHNOLOGIES

VOLUME 2

ELECTRONIC COMPONENTS, SYSTEMS AND TECHNOLOGIES I

ELECTRONIC COMPONENTS, SYSTEMS AND TECHNOLOGIES II

EDUCATION QUALITY

INTERNET TECHNOLOGIES

CONTROL SYSTEMS

COMPUTER SYSTEMS

POWER TRANSMISSION AND DISTRIBUTION SYSTEMS I

POWER TRANSMISSION AND DISTRIBUTION SYSTEMS II

STUDENT SESSIONS

STUDENT SESSION I

STUDENT SESSION II

VOLUME 3

POSTER SESSIONS

PO1 – TELECOMMUNICATION SYSTEMS AND TECHNOLOGIES

PO2 – RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS

PO3 – SIGNAL PROCESSING, METROLOGY AND REMOTE SENSING

PO4 – CONTROL SYSTEMS AND ROBOTICS

PO5 – RENEWABLE ENERGIES

PO6 – COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES

PO7 – ELECTRONIC COMPONENTS, SYSTEMS AND TECHNOLOGIES

PO8 – POWER TRANSMISSION, DISTRIBUTION SYSTEM AND ELECTRICAL MACHINES

PO9 – EDUCATION QUALITY

POSTER SESSIONS

Session PO1:

PO1 - TELECOMMUNICATION SYSTEMS AND TECHNOLOGIES

Effective P2P VoD Service Distribution over HFC Networks

Jordan Nenkov[1] and Lidia Jordanova[2]

Abstract – **The paper deals with a concept for realizing a Peer-to-Peer (P2P) VoD system over HFC networks in a way to ensure possibilities to increase the number of both the subscribers and the movies supported. Different methods to transfer video information from the VoD server to subscribers are considered. Three strategies to limit the downstream video traffic are suggested.**

Keywords – **P2P VoD, multicast video, MPEG2-TS, TVoD.**

I. INTRODUCTION

Due to recent advances in broadband Internet access technology the video-on-demand (VoD) service seems to gain an increasing popularity among media streaming services. VOD networks are developed to deliver video files to distant users with minimal delay and free interactivity [1].

Traditionally, the VoD service is based on a centralized architecture. However, this architecture cannot provide the quality of service needed to a large population of users due to its limited outbound channel capacity from the server to the clients. Recently, a peer-to-peer (P2P) architecture was proposed to meet the challenge of providing live and interactive video broadcast to a large number of clients over a wide area [2]. A P2P-based architecture is appropriate enough for the design of a scalable VoD services distribution architecture as the computing and bandwidth requirements are pushed toward the network clients side. Besides, it allows optimal use of the network resources by building multisource streams from neighbouring contributing clients to a requesting client. This in turns results in minimizing the VoD request rejection rates for a very large content library [3].

The aim of this work is to compare the methods of video information transmission from VoD server to subscribers and to choose the most appropriate one in terms of price, quality and complexity of implementation.

II. VOD SERVICE STRATEGIES

There are three main types of service strategies: broadcast, unicast and multicast.

[1]Jordan Nenkov is with the Faculty of Telecommunications, Kliment Ohridski 8, 1756 Sofia, Bulgaria, E-mail: Jordan_n2002@yahoo.com.

[2]Lidia Jordanova is with the Faculty of Telecommunications, Kliment Ohridski 8, 1756 Sofia, Bulgaria, E-mail: jordanova@tu-sofia.bg.

A. Unicast Communication Strategy

This is a point-to-point communication channel which means that unicast establishes communication between a single sender and a single receiver over the network.

With unicast, as shown in Fig. 1, an independent flow of information is sent to each one of the clients requesting the multimedia content.

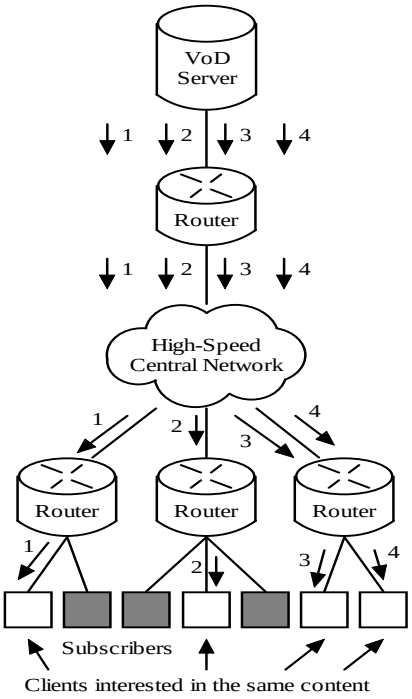


Fig. 1. Unicast communication strategy.

The main advantage of this technique is the control mechanism. Each communication channel is independent, thus the control is oriented to each client and there are no interactions between the channels. On the other hand, the bad scalability of the scheme is a great constraint of the unicast usage. Each channel consumes server and network resources in a linear fashion, which results in a rapid consumption of the resources predicted during the system's design. Unicast converts the server into a bottleneck for the VoD system.

B. Broadcast Communication Strategy

This service strategy is based on communication from a single sender to all the receivers over a network. In this scenario the video is broadcast over a dedicated channel within a predefined schedule. This approach can provide

service to an unlimited number of requests for popular video content with a constant consumption of bandwidth. However, a great amount of resources can be consumed pointlessly if the popularity of the video is low. In addition, clients must wait until the scheduled time to be served. In Fig. 2 a broadcast communication scheme is shown, where the information is distributed to every end host in the system.

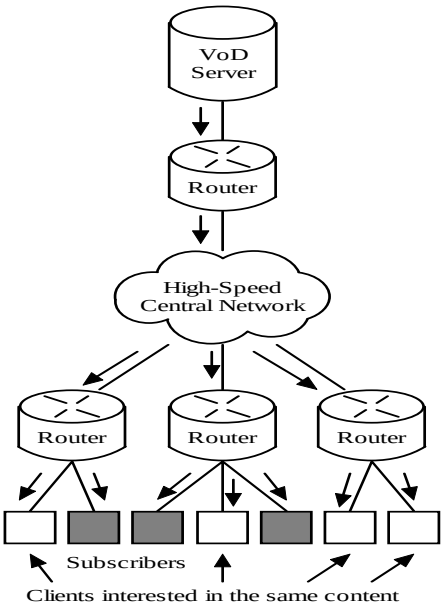


Fig. 2. Broadcast communication strategy

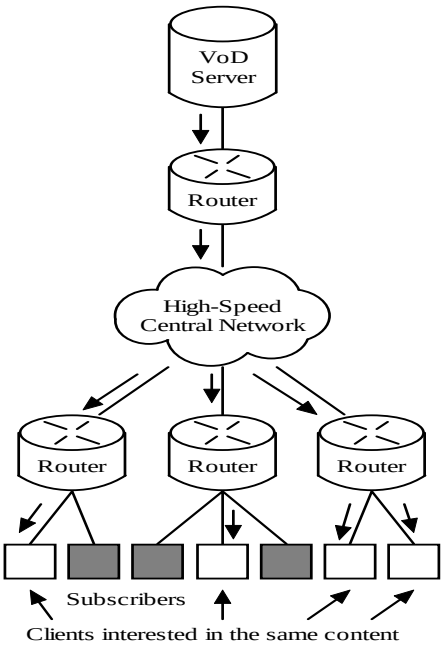


Fig. 3. Multicast communication strategy

C. Multicast Communication Strategy

The multicast technique consists of an information flow from a source to a group of receivers who have requested the same content. A number of requests for the same video are grouped and served by a single video stream. Therefore, the service scheme saves bandwidth by sharing video streams without wasting resources on non-requesting receivers. Multicast is shown in Fig. 3. It provides the best solution to scale large VoD systems.

III. The Collaborative Method for Downstream Video Traffic Decreasing

Let's analyze a collaborative method of reducing the downstream video traffic as shown in Fig. 4. After subscriber 1 has requested a video, he is not served immediately but after some time *t*, during which it is assumed that subscriber 2 will make a request for the same film. Once a subscriber has made a second request for the film and there is an awaiting subscriber (subscriber 1), the film (video stream) is fed to both subscribers simultaneously. In Fig. 4 the following symbols are used:

req. 1 and req. 2 are the requests for one and the same movie as made by subscribers 1 and 2 respectively;

reply 1 and reply 2 – request fulfilled for one and the same film, in reply to requests made by subscribers 1 and 2 respectively;

t1 – time when a request from subscriber 1 was issued;

t2 – time when a request from subscriber 2 was issued;

t3 – time to reply to both subscribers;

τ is the time between the requests for one and the same film issued by subscribers 1 and 2;

$\tau 1$ and $\tau 2$ – time to run between query and reply to subscribers 1 and 2 respectively.

The more the subscribers that have requested one and the same movie in the time interval τ, the more the ports on the VoD server that will be free for future requests. A disadvantage of this method is the delay in executing the movie requests. The value of τ must be chosen in a way to guarantee that the delay is below a given value acceptable for subscribers.

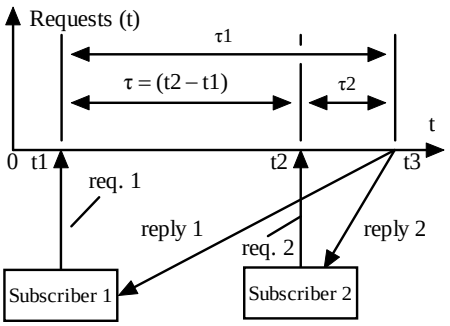


Fig. 4. Collaborative method for downstream video traffic decreasing

IV. Method of Downstream Video Traffic Decreasing with Reducing the Client Waiting Time

The method consists in the following. Every i^{th} video V_i is divided into 3 parts, the first W_1 minutes of each video V_i being referred to as prefix-1 $(pref\text{-}1)_i$ of V_i. If V_i is globally popular, it is replicated at all M Proxy Servers; otherwise, it is replicated across L Proxy Servers, in which the frequency of accessing the video V_i is high. The next part of W_2 minutes of video $V_{i,}$ referred to as prefix-2 $(pref\text{-}2)_i$ of $V_{i,}$ is downloaded from the Tracker and the rest of the video is referred to as suffix of the video and is stored at VoD Server, as shown in Fig.5.

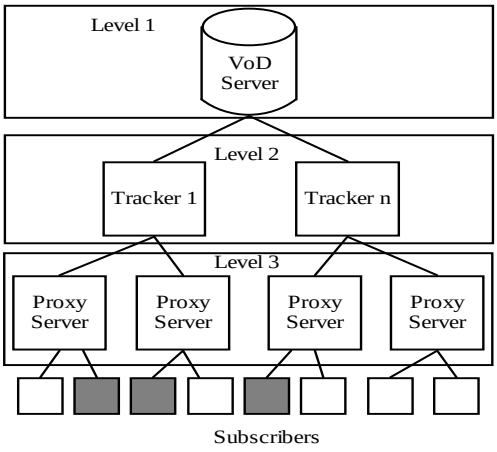


Fig. 5. Architecture realization of the method

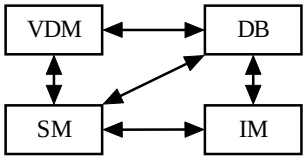


Fig. 6. A Tracker module

The proposed 3-layer architecture consists of a VoD Server, which is connected to a group of Trackers. Each tracker has various modules, as shown in the Fig. 6:

1. Interaction Module (IM) that interacts with the Proxy Server and the VoD Server.
2. Service Manager (SM) that handles the requests from the Proxy Servers.
3. Database (DB) that stores detailed information about availability and size of (pref-1) of videos at all the Proxy Servers.
4. Video Distributing Manager (VDM) which is responsible for decision-making about videos and sizes of (pref-1), (pref-2) of videos to be cached. It also handles the distribution and management of these videos to the Proxy Servers group, according to their global and local popularity.

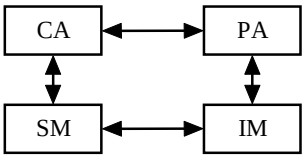


Fig. 7. A Proxy Server module

Each Tracker is in turn connected to a set of Proxy Servers. Each Proxy Server has various modules, as seen in Fig. 7:

1. Interaction Module (IM) that interacts with the user and the Tracker;
2. Service Manager (SM) that handles the requests from the user;
3. Popularity agent (PA) that observes and updates the popularity of videos at Proxy Servers and at the Tracker as well as;
4. Cache Allocator (CA) that allocates the Cache blocks using dynamic buffer allocation algorithm [4]. A large number of users are connected to each of these Proxy Servers as well. Each proxy server acts as a parent proxy server to its clients. The Proxy Server caches the (pref-1) of videos distributed by VDM and streams this cached portion of the videos to the clients.

Here it is assumed that:

1. The Tracker is of high computational power. It has various modules to coordinate and maintain a database that contains the information about availability of videos and size of (pref-1) and (pref-2) of video in each Proxy Server and Tracker respectively.
2. Proxy Servers and their clients are closely located at a relatively low communication cost [5]. The VoD server where all the videos are stored is placed far away from Proxy Servers which involves high-cost remote communication.

V. Method for Downstream Video Traffic Decreasing with Video Popularity Using

The distribution of the VoD-movie requests generally follows a Zipf / Pareto-like distribution [6], where the VoD movies are of two classes: popular and unpopular. The relative probability of a request for *i* (the most popular movie) is proportional to $1/i^{\alpha}$, with $0 < \alpha < 1$, and typically takes a value that is less than unity [7]. The assumption here is that all blocks of a popular movie belonging to the popular class are stored in the mesh network of the client, and if needed by a given client, they can be downloaded from its mesh network partners. For Zipf-like distributions, the cumulative probability that one of the *k* popular movies class is accessed (i.e. the probability of a movie request from the client mesh network) is given asymptotically by:

$$\psi(k) = \sum \delta / i^{\alpha} = \delta k^{1-\alpha} /(1-\alpha), \quad \delta = (1-\alpha)/V^{1-\alpha} \qquad (1)$$

where *V* is the total number of movies in the system [7]. Next, we estimate the probability (Pu) of a request for a movie that belongs to the unpopular class (i.e. does not exist in the

mesh network setup boxes) and therefore should be obtained from the central head-end video server. For a VoD system with V total movies and k popular ones at the mesh network the probability of a request for an unpopular movie stored in the head-end server is as follows:

$$P_u = 1-(k/V)^{1-\alpha}. \quad (2)$$

This method is applicable in networks with large numbers of subscribers. The more the subscribers and fewer the movie titles in the system, the better the results obtained with the above method.

VI. CONCLUSION

The paper deals with a comparative analysis of three types of communication strategies for transmission of video information from the VoD server to subscribers - unicast, broadcast and multicast. Three methods to reduce the downstream traffic from the VoD server to subscribers are considered. The following conclusions can be drawn:

1. The unicast communication strategy is appropriate when high security of information transmitted is required, but is inefficient in terms of price realization.
2. The broadcast communication strategy is a low-cost selling, but much of the resources are used unnecessarily when the popularity of films is low.
3. The multicast communication strategy saves bandwidth by sharing the video streams without loss of resources to unsolicited recipients. It provides the best solution for growing large VoD systems.
4. The collaborative method for downstream video traffic decreasing is easy to implement, but it results in movie requests execution latency which is a disadvantage of the method.
5. The method for downstream video traffic decreasing with reducing the client waiting time eliminates the shortcomings of the collective method but turns out to be complex and relatively expensive for implementation.
6. The method for downstream video traffic decreasing with video popularity using is applicable in networks with large numbers of subscribers. It gives much better results in systems with more subscribers and fewer movie titles.

REFERENCES

[1] F. Thouin and M. Coates, "Video-on-Demand Networks: Design Approaches and Future Challenges", IEEE Network Magazine, March/April 2007.

[2] Y. He, I. Lee, and L. Guan, “Distributed Throughput Maximization in P2P VoD Applications “ , IEEE Transaction on Multimedia , Vol. 11, No. 3, April 2009.

[3] A. Nafaa, S. Murphy, and L. Murphy, “Analysis of a Large-Scale VOD Architecture for Broadband Operators: A P2P-Based Solution” , IEEE Communications Magazine December 2008.

[4] H.S. Guruprasad, M. Dakshayini, “Dynamic Buffer Allocation for VoD System Based on Popularity” proceedings of NCIICT 2006, PSG College of Technology, Coibatore, 13- 17, www.psgtech. edu/NCIICT/files/NCIICT06.

[5] B. Wang, S. Sen, M. Adler and D. Towsley, “Optimal Proxy Cache Alloction for Efficient Streaming Media Distribution” IEEE Transaction on Multimedia, vol. 6, No. 2, April 2004.

[6] L. Jordanova, J. Nenkov, Distribution of Video-on-Demand Service over Cable Television Networks. Radioengineering, Vol. 2, No 09-021, 2009.

[7] L. Breslau, "Web Caching and Zipf-like Distributions: Evidence and Implication", Proc. IEEE INFOCOM 99, New York, March 1999, pp.126-34.

Forecasting FTTH as a New Broadband Technology

Valentina Radojicic[1], Goran Markovic[2] and Vladanka Acimovic-Raspopovic[3]

Abstract – **This paper proposes a model for FTTH forecasting as a new broadband technology. Our analysis is performed by using the Generalized Bass Model. We analyzed several possible scenarios with different percents of price reduction and marketing efforts as well as market potential. Based on the appropriate scenario it is possible to choose the best investment strategy.**

Keywords – **broadband traffic, forecasting, FTTH, transport network.**

I. INTRODUCTION

Huge investments have been made to roll out broadband networks in recent years. Long-term broadband demand forecasts have been and are crucial for investment decisions, rollouts and dimensioning of networks. Nowadays, the main broadband access technologies include DSL (Digital Subscribe Line) and Cable modem (Hybrid Fibre Coax). Other technologies like fiber and fixed wireless access are also entering the market. Especially in Japan, Hong Kong, China, and Korea, the growth of FTTH (Fiber-to-the-Home) has been significant in the last years [1]. There are several of technical, economic and business parameters that impact the right choice for each specific network situation. An operator runs the risk of picking an incorrect technology strategy if any of these key parameters are not identified and cost optimized [2]. Different technology options are available to operators today for their FTTH network deployment strategy decisions. Gigabit-Passive Optical Network (GPON), Ethernet Passive Optical Network (EPON) are called Point to Multi Point (P2M), Active Ethernet (AE) also known as Point-to-Point Ethernet (P2P) are the major competing technologies. Choice of active or passive architectures for deployment depends on the type of services to be delivered, cost of the infrastructure, current infrastructure and future plans for migrating to the new technologies.

With the aim of adequately planning required network resources, it is necessary to forecast traffic demands that are in direct correlation to the forecasted number of the customers. In this paper, we present a model for FTTH forecasting as a new broadband technology. The analysis made is based on diffusion theory, which takes into account advertising investment and effects of different prices. The Generalized Bass Model or GBM [3] has become especially popular, in both descriptive and normative applications. It is shown that in the GBM, the optimal evolution of advertising expenditures after launch is highly dependent on the initial level of advertising [4].

The paper is organized as followed. In the second section we present the various FTTH architectures available for deployment and worldwide forecasting of FTTH technology. The third section explains the GBM diffusion model that we used to forecast the future broadband demands. After that in the fourth section, we present the forecasted results for FTTH demand for Serbian market. Finally, we conclude the article giving the managerial implications.

II. FTTH ARCHITECTURES AND WORLDWIDE MARKET FORECAST

Network operators around the world are looking at transforming into Next-Generation-Network to remain competitive in a radically changing telecommunications environment. Broadband services continue its explosion, with the migration from copper broadband via xDSL to next generation FTTH deployments having begun in earnest. Optical fiber, as used in the core or metro network, can also be used in the access network as medium for digital transmission. Optical fiber can offer much higher bandwidths than are attainable with DSL or HFC [5]. The bandwidths that can be offered are largely depending on the fiber and architecture installed and the equipment used. In general fiber already runs up to a location close to the customer, and the different alternatives are indicated by FTTx with a specific character indicating where the fiber stops. Often used acronyms are FTTN (Fiber to the Node), FTTC (Fiber to the Curb) and FTTB (Fiber to the Building) with an ever advancing fiber running respectively up to the node, cabinet or building. The remainder of the access network is in these cases still bridged by DSL, HFC or wireless technologies. In the case the fiber runs all the way up to the customer's house, apartment or premises, this is called FTTH or FTTP.

FTTH Worldwide Market & Technology Forecast, 2006-2011 describes the key competing technologies, divided into two main categories – active fiber architectures, typically active Ethernet usually known as active optical networks (AONs), and passive fiber architectures, usually known as passive optical networks (PONs). The optical network termination (ONT) at the customer side performs the translation of the optical signal to the in-house wiring. Next to Ethernet it is not uncommon to perform a translation to a broad range of existing connectors at this point (coax, twisted pair, wireless). The inside optical wiring is connected to the outside plant by means of an optical connector plug often referred to as the optical network termination point (ONTP). The structure of an FTTH network is shown in Fig.1.

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

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

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

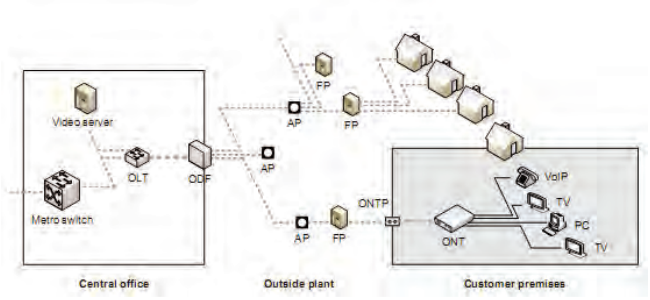


Fig. 1 Structure of FTTH network (OLT-optical line terminal; ONT-optical network termination; ODF- Optical Distribution Frame; FP-flexibility point; AP-aggregation points;)[7]

The optical signal is transported over the outside plant up to the central office. In the outside plant there are different aggregation points (AP) and often there is also a flexibility point (FP) comparable to a street cabinet in copper based networks, closest to the customer. At this point there are various options for the telecom operator. Passive optical networks (PONs) will aggregate the optical signal of different fibres into one fiber at such aggregation points using passive optical splitters and as such create a point to multipoint network with the optical fiber as shared medium. Active optical networks (AONs) will connect the customer with a dedicated fiber up to the OLT. Also the number of customers per PON, AP and FP are degrees of freedom for the operator installing the network. At the central office, all optical fibres connect to the ODF and from there to the optical line terminal which will aggregate all traffic, and translate between protocols where necessary. In case of a PON, access to the shared medium is divided between the different customers by means of some division multiplexing based protocol. At this point one or more additional wavelength(s) can be used for broadcasting content (for instance RF video) to all customers of a PON. Beyond the OLT, the traffic is sent into the metro network [7].

PONs became a popular solution among operators because they are seen as the least costly architecture for delivering FTTH in a mass market residential scenario. As one would expect, the initial population density developments were set in locations that maximize economies of scale, namely urban centres with high population density. The objectives of PON deployments are usually measured in Households Passed (HHP). In this context, a house is deemed as “passed”, when the distribution fiber reaches a Network Access Point (NAP) which, in urban settings, is usually placed inside buildings. The number of fibres that feed a given building is usually a function of the number of dwelling units in that building as well as of the Take Rate considered by each operator. The Take Rate is the predicted maximum percentage of users that will request the fiber service from the operator. Naturally this varies between operators and the type of areas served.

Active optical networks (AONs) will connect the customer with a dedicated fiber up to the OLT. Also the number of customers per PON, AP and FP are degrees of freedom for the operator installing the network. At the central office, all optical fibers connect to the ODF and from there to the optical line terminal which will aggregate all traffic, and translate between protocols where necessary. At this point one or more additional wavelength(s) can be used for broadcasting content (for instance RF video) to all customers of a PON. Beyond the OLT, the traffic is sent into the metro network [6].

Traditionally, the deployment of new telecommunications services/technology has taken years of effort and large amounts of investment. Thus, being able to predict the market acceptance before taking the business risk is critically important.

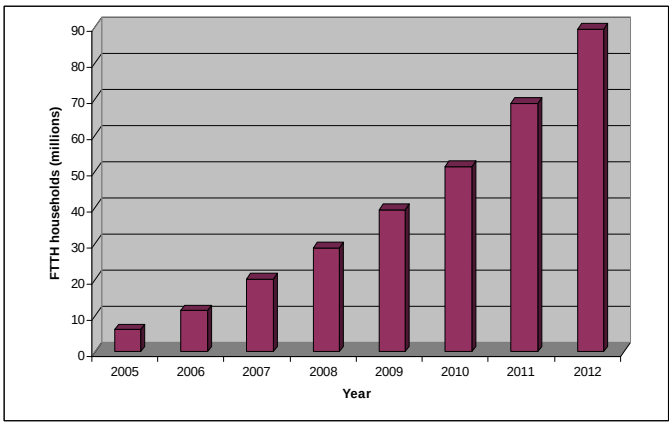


Fig. 2. Worldwide forecasting of FTTH technology [8]

Fig. 2. shows that the FTTH connected households will reach up to 90 million by the end of 2012.

III. Implementation of the GBM (Generalized Bass Model)

According to diffusion theory, a new service's sales growth at any time largely depends on the strength of word of mouth from its previous adopters. The most important model in this stream of research is the Bass model [9][10]. The Bass diffusion model has been widely used as forecasting procedure of new services/technologies and it was proposed to deal with the problem of initial adopters.

The mathematical structure of the Bass model is derived from a hazard function corresponding to the conditional probability that an adoption will occur at time *t* given that it has not occurred yet. If *f(t)* is the density function of time to adoption and F(t) is the cumulative fraction of adopters at *t*, the basic hazard function underlying the Bass model is given by Eq. 1:

$$\frac{f(t)}{1-F(t)} = p + q \cdot F(t) \qquad (1)$$

This model has three key parameters: the parameter of innovation (*p*), the parameter of imitation (*q*) and the market potential (*m*). Parameter *q* reflects the influence of those users who have already adopted the new technology, while *p* captures the influence that is independent from the number of adopters. The sale at time *t* is $S(t) = m[F(t) - F(t-1)]$, where *m* refers to the market potential for the new technology. These parameters can be estimated using cumulative sales data.

Generalized Bass Model, GBM incorporates control variables into diffusion models such as price and advertising

[3]. These are two marketing mix instruments that should ideally be used in extended diffusion models with marketing effects. However, researchers and managers may frequently have information on only one of the two (e.g. price). The GBM presents a surprisingly simplified structure:

$$\frac{f(t)}{1-F(t)} = Z(t)\cdot[p + q\cdot F(t)],\quad t \geq 0 \tag{2}$$

where:

$$Z(t) = 1+\alpha\cdot\frac{P(t)-P(t-1)}{P(t-1)}+\beta\cdot\frac{A(t)-A(t-1)}{A(t-1)} \tag{3}$$

follows:

$$Z(t) = 1+\alpha\cdot\Delta P+\beta\cdot\Delta A \tag{4}$$

where are:
F(t)=N(t)/m – the cumulative function of adoption for time *t*; *N(t)* – the cumulative sales; *m* – the market potential; *p* – the parameter of innovation (initial probability of adoption); *q* – the parameter of imitation (diffusion rate); α – the diffusion rate as a result of price decrease for 1%; *P(t)* – the current price; β – the diffusion rate as a result of increase advertising for 1%; *A(t)* – the current level of advertising expenditure; ΔP – the proportional change in price; ΔA – the proportional change in advertising efforts.

Assuming F(0)=0, the closed-form solution of differential Eq. (2) is:

$$F(t) = \frac{\left(1-e^{-Z(t)(p+q)t}\right)}{\left(1+\frac{q}{p}e^{-Z(t)(p+q)t}\right)} \tag{5}$$

Note, the speed of adoption at a particular point in time is affected not by the level of price or advertising at that time but by the proportional change in those marketing mix variables at that time. If the percentage changes in price and advertising remain the same from one period to the next, then function Z(t) reduces to a constant, yielding again the Bass model. The GBM allows to test the effect of marketing mix strategies on diffusion and to make scenario simulations based on intervention function modulation. Function Z(t) acts on the natural shape of diffusion, modifying its temporal structure and not the value of its internal parameters: as a consequence, the important effect of Z(t) is to anticipate or delay adoptions, but not to increase or decrease them. In other words, function Z(t) may represent all those strategies applied to control the timing of a diffusion process, but not its size [11].

Estimation of the parameters *p*, *q* and *m* is required to identify the diffusion curve. Bass model could be used to predict the timing and magnitude of the sales peak, and the shape of the diffusion curve. But, the most applications of the Bass model are used to make plans and decisions before the service/technology has been introduced to the market. Usually, no sales data exists with which to estimate *p* or *q*. Manager has not an intuitive estimate of *p* and *q*. In such a case, Bass parameters could be evaluated in two manners. One way is to use analogies with other similar services or diffusion process. The second way is analytical using comparative procedure with some other countries where a service/technology already exists.

If data sales does not exist, the market potential has to be estimated by taking into account different impact factors such as economy and social development of a particular area, presence of competitive broadband technologies, operators infrastructure investment strategies, etc.

IV. NUMERICAL RESULTS FOR SERBIAN MARKET

The users will be the focus of the operator and the estimation of their adoption behavior is probably the most important source of input for a planning activity and business model. Serbian broadband market is characterized by three long-standing market technologies: ADSL (Asymmetric Digital Subscriber Line), HFC (Hybrid Fiber-Coax), FWB (Fixed Wireless Broadband). However, it is expected that FTTH technology will be introduced this year. In this paper, we considered GBM as input model for our research.

We analyzed here several possible scenarios related with the cost for end users and marketing efforts. The cost of FTTH per home will be related with the chosen type of fiber architecture and cost of installation. For example, the cost of ONT in a PON will be higher (about 40%) than in case of an HRN (Home Run Network). We assumed this price difference because an HRN poses less stringent requirements considering the optical budget, bandwidth and protocol. In a case of an all buried network, the fiber cable to be installed for bridging the last meters is already available and connected in the pedestal. As such the trenching at installation time is minimized. In a case of an aerial customer connection, the last meters from the drop box up to the house have to be bridged at installation time. This difference leads to a less costly installation in case of an all buried network.

Here we proposed seven possible *Scenarios*. The Bass parameters, *p* and *q*, are estimated by comparative procedure based on worldwide forecasting of FTTH technology, using ordinary least squares (OLS) multiple regression by Eg. 6 - 8. In all proposed *Scenarios* the parameter of innovation are remaining unchanged.

In all considered *Scenarios* we assume that the substitution effects will happen between new (FTTH) and current technologies (HFC, ADSL, and FWB). It means, that overall market potential in Serbia, which is estimated to *m*=2.500.000 households, will be reduces in a case of FTTH to 500.000 (alternatively 800.000) households, because of low households economy and necessary investments in network infrastructure. In *Scenario 1* there is no change in price and marketing efforts. It corresponds to basic Bass model. All other Scenarios are compared with it. *Scenario 2*, *3* and *4* assume that the cost will be reduced for 20%, 40% and 60% respectively. The price reduction will make influence on the increasing of the parameter of imitation as we assumed in Table I. In addition, the price reduction could have an impact on market growth (*Scenario 4*). *Scenario 5*, *6* and *7* take into account marketing mix variables (price and advertising). Estimated parameters values for considered *Scenarios* are given in Table 1.

TABLE I
ESTIMATED PARAMETERS

Scenario	*p*	*q*	*m*	ΔP	*A*	ΔA	β
1	0.001	0.12	500.000	0	-0.37	0	/
2	0.001	0.25	500.000	-0.2	-0.37	0	/
3	0.001	0.29	500.000	-0.4	-0.37	0	/
4	0.001	0.29	800.000	-0.6	-0.37	0	/
5	0.001	0.25	500.000	-0.2	-0.37	0.4	0.35
6	0.001	0.29	500.000	-0.4	-0.37	0.6	0.35
7	0.001	0.29	800.000	-0.6	-0.37	0.8	0.35

The obtained results for proposed Scenarios are given by Fig. 3 and Fig. 4.

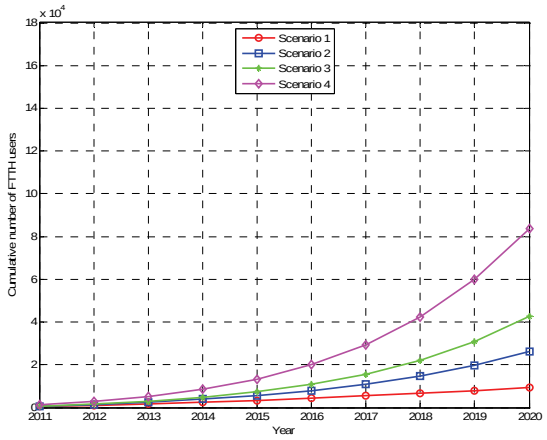


Fig. 3. Forecasted results for FTTH households for different percentage price reductions (20%, 40%, 60%)

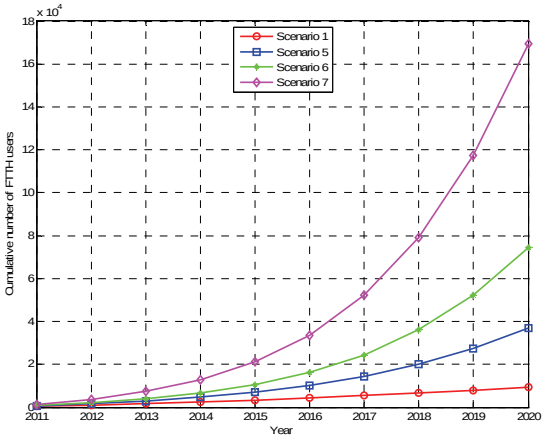


Fig. 4. Forecasted results for FTTH households for different percentage price reductions (20%, 40%, 60%) and advertising efforts (40%, 60%, 80%)

We can see that the cumulative number of FTTH users is heavily influenced by price reductions (Fig. 3) as well as advertising efforts (Fig. 4). For example, it is expected that the number of FTTH households in Serbia could reach up to 34.000 users in 2016, depending on the advertising investments and price reduction.

V. CONCLUSION

In this paper we suggest the GBM (Generalized Bass Model) for long-term forecasting of new broadband technology. The main advantage of this model is that it includes the marketing mix variables: price and advertising effects. It enables broadband operators to predict the number of users according to the price reductions as well as marketing investments. Also, the proposed model would enable operators to quickly make the right technology deployment decisions. We applied this model for FTTH technology deployment in the case of Serbian market. We analyzed several scenarios with different percents of price reduction and marketing efforts as well as market potential. In this way, managers can make the right technology investments and cost strategies that are important to capture new broadband market shares.

ACKNOWLEDGEMENT

This paper resulted from the researching project TR-32025 that is supported by the Serbian Ministry of Education and Science.

REFERENCES

[1] K. Stordahl, "Long-Term Broadband Evolution –Forecasts and Impact of New Technologies", Telektronikk, no. ¾, 2008.

[2] S. Kulkarni, M. El-Sayed, P. Gagen, B. Polonsky, "FTTH network economics: Key parameters impacting technology decisions",Infocommunications Journal, vol. LXV, no.2, 2010.

[3] F. M. Bass, T. V. Krishnan, D. C. Jain, "Why the Bass Model Fits Without Decision Variables", Marketing Science, vol.13, no. 3, 1994.

[4] V. T. Krishnan, D. C. Jain, Optimal Dynamic Advertising Policy for New Products, Management Science, vol. 52, no. 12, pp. 1957-1969, 2006.

[5] P. E. Green, "Fiber-to-the-Home: The Next Big Broadband Thing", IEEE Communications Magazine, vol. 42, no. 9, pp. 100-106, Sep. 2004.

[6] Cisco, "Hong Kong Broadband Driving World's Fastest Residential Broadband Service with Cisco Powered Network QoS-certified Metro Ethernet Solution" available at: http://newsroom.cisco.com/dlls/global/asiapac/news/2005/pr_1-26b.html

[7] K. Casier, "Techno-Economic Evaluation of a Next Generation Access Network Deployment in a Competitive Setting" , PhD, Faculty of Engineering of the Ghent University, 2009, available at: http://ibcn.intec.ugent.be/te/Members/PhD_KoenCasier.pdf

[8] FTTH Worldwide Technology Update &Market Forecast by heavy Reading, Feb. 2008. Avl. at: http://www.heavyreading.com/details.asp?sku_id=2087&skuitem_itemid=1071

[9] V. Mahajan et al. *New Product Diffusion Models: From Theory to Practice*, Boston: Kluwer Academic, 2000.

[10] F. M. Bass, "New Product growth model for consumer durables", Management Science 15, pp. 215-227, 1969.

[11] V. Mahajan, E. Muller, M. F. Bass, "New Product Diffusion Models in Marketing: A Review and Directions for Research", Journal of Marketing, vol. 54, pp. 1–26, 1990.

Network Selection Heuristics Evaluation in Vertical Handover Procedure

Bojan Bakmaz[1] and Miodrag Bakmaz[2]

Abstract – **Network selection heuristics are essential components of the heterogeneous wireless networks architecture deployment. Following the principles of heterogeneous networking, a mobile user may choose among multiple available connectivity alternatives based on the criteria related to networks performances, users preferences and services requirements. This paper seeks to provide a framework for network selection heuristics evaluation. Some perspective network selection heuristics, based on cost function, artificial neural networks, multi criteria decision making and fuzzy logic systems are systematically presented and analyzed in terms of efficiency and implementation complexity.**

Keywords – **Heterogeneous wireless networks, Mobility, network selection, QoS, Vertical handover.**

I. INTRODUCTION

Heterogeneous wireless networks inherits the vital complementary characteristics of both infrastructure and ad-hoc architectures, and thus has the potential of attaining the level of performance and efficiency required by the future ubiquitous wireless communications [1]. Following the principles of heterogeneous networking, users will be able to choose among multiple available connectivity alternatives based on the criteria related to networks performances, users preferences and services requirements. This process makes an important element in the complex vertical handover procedure. The need for a handover in the traditional wireless networks usually occurs when the terminal, due to the movement of the users, leaves the Point of Attachment (PoA) coverage area. In a heterogeneous environment, handover is more frequently used in order to improve communication and rider in order to maintain connections [2].

Major challenges in heterogeneous handover management are seamlessness and automation aspects in network switching. It is a strategic goal to define important advancements that happen and are predicted in technologies, networks, user terminals, services, and future business models that include all this issues while realizing and exploiting new wireless networks. On the other hand, because users could be always connected through the optimal Radio Access Network (RAN), it is necessary to develop an adequate mechanism for its selection.

ITU's concept of Optimally Connected, Anywhere, Anytime proposed in M.1645 [3] states that future wireless networks could be realized through the coalition of different RANs. According such a scenario, the heterogeneity of access networks, services and terminals should be fully exploited to enable higher utilization of radio resources. The main objective is to improve overall networks performances and perceived QoS.

Third Generation Partnership Project (3GPP) is defining an Access Network Discovery and Selection Function (ANDSF) [4] to assist mobile terminals in vertical handover between 3GPP and non-3GPP networks, covering both automated and manual selection as well as operator and user management.

IEEE 802.21 is developing standards to enable handover and interoperability between heterogeneous link layers [5]. This standard defines the tools required to exchange information, events and commands to facilitate handover initiation and preparation. IEEE 802.21 standard does not attempt to standardize the actual handover execution mechanism. Therefore, the Media Independent Handover framework is equally applicable to systems that employ mobile IP at the network layer as well as systems that use Session Initiation Protocol (SIP) at the application layer.

A great number of heuristics related to the handover initiation and optimal access network selection are proposed in the open literature. The suggested solutions are using different criteria and mathematical tools for solving the above mentioned problems. Unfortunately, currently proposed solutions do not meet all the requirements in terms of functionality and efficiency.

II. OPTIMAL NETWORK SELECTION FRAMEWORK

Network selection in heterogeneous environment is essentially a resource allocation problem and is typically addressed as user-centric, network-centric or a hybrid approach. In the network selection scenario users are always trying to seamlessly access high-quality wireless service at any speed, any location, and any time through selecting the optimal RAN. Therefore, ensuring a specific QoS is one of the main goals in the process of network selection.

There are four groups that are often analyzed as the criteria in the network selection, and those groups are related to the entities - the participants in handover decision:

- Network-oriented metrics (coverage, link quality, bandwidth, etc.),
- Service-oriented metrics (QoS level, security level, etc.),
- User-oriented metrics (user's preferences, perceived QoS, etc).
- Terminal-oriented metrics (velocity, energy consumption, etc.).

[1]Bojan Bakmaz is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: b.bakmaz@sf.bg.ac.rs.

[2]Miodrag Bakmaz is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: bakmaz@sf.bg.ac.rs.

Received Signal Strength (RSS) is the most widely used criterion because of its measure simplicity and close correlation to the link quality. There is a close relationship between RSS readings and the distance from the mobile terminal to its PoA. Traditional horizontal handover techniques are basically analyzing metrics through the variants of comparing RSS of the current PoA and candidate network PoA. In combination with threshold and hysteresis, RSS metrics represent a satisfying solution for a homogeneous network environment. In a heterogeneous environment RSS metric is not sufficient criterion for initiating a handover, but in combination with other metrics it could be applied as an ultimate condition.

Available bandwidth represents important indicator of RAN traffic performances and transparent parameter for users of the multimedia services. This is the measure for per user bandwidth allotted by the network operator which is dynamically changeable according the utilization of the network. The maximum theoretical bandwidth is closely related to the channel capacity. Transition to a network with better conditions and performances would usually provide improved perceived QoS.

The QoS level can be defined through the metric values of delay, jitter, package loss, etc. and it can be declared by the service provider on the basis of ITU recommendation Y.1541, which defines the upper bounds of QoS parameters for specific applications or classes of services. By declaring the QoS level in this way, we will avoid a complex examination of QoS parameters by users as well as the additional load of user's terminals and other network elements.

Security level, as well as the previous criteria, may be declared by the service provider, and it represents the security measure for the information transfer in the network. For most users, depending on the application, security plays a great part in making a decision on the adequacy of a network for transferring the desired content. When the information exchanged is confidential, a network with high encryption is preferred. The security level concept, sometimes called Level of Security (LoS), is similar to level of service in QoS management. LoS is a key piece of information within a security profile and is used to determine whether data are allowed to be transferred by a particular network or not.

Cost of service can significantly vary from provider to provider, but in different network environments. In some cases cost can be the deciding factor for optimal network selection, and it includes the traffic costs and the costs of roaming between heterogeneous networks. In some context cost of service is in tight relation with available bandwidth, QoS level, security level, but in next generation wireless environment, cost of service is fast time differentiable function dependable of many others parameters [6].

Mentioned metrics are affecting the moment of the handover initiation and optimal access network selection. The number of criteria, and dynamic variability of some parameters significantly increases the complexity of the handover heuristic, and because of that, the choosing of adequate criteria is of great importance. After the definition of the convenient parameters, the question often arises is how to transfer the metrics information from the network entities to the user's multimode terminals. Through the End to End Reconfigurability (E^2R) project, concepts and solutions for a Cognitive Pilot Channel (CPC) were developed [7]. It was concluded that CPC will be able to provide enough information for network selection, when users are preceding either initial connection or handover.

Performance analysis of the network selection heuristic can be performed through the determination of mean and maximum handover delays, number of handovers, number of handovers failed due to the incorrect decisions, handover failure probability, resource utilization, etc [8].

Handover delay refers to the duration between the initiation and completion of the handover process. It is related to the complexity of the applied heuristic. Reduction of the handover delay is especially important for delay-sensitive voice and multimedia sessions.

Reducing the number of handovers is usually preferred, as frequent handovers would cause wastage of network resources. A handover is considered as superfluous when a mobile terminal is coming back to the previous PoA is needed within certain time duration ("ping-pong" effect), and such handovers should be minimized.

A handover failure occurs when the handover is initiated, but the target network does not have sufficient resources to complete it, or when the mobile terminal moves out of the coverage area before the process is finalized. In the first case, the handover failure probability is related to the resource availability (e.g. channel availability) of the target network, while in the second case, it is related to the terminal mobility.

Resource utilization is defined as the ratio between the mean amount of utilized resources and the total amount of resources in a system. In the case of efficient channel utilization, the ratio between the mean number of channels that are being served and total number of channels in a system is taken into account.

For efficient network selection strategy the following important issues have to be fulfilled:

- Only considerable parameters must be analyzed,
- Equilibrium among user's preferences, service's requirements and network's performance must be achieved,
- Technique has to be reliable and transparent to the user,
- Heuristic has to minimize handover latency, blocking probability and number of superfluous handovers,
- Flexible and suitable implementation in real environment is necessary.

III. COST FUNCTION BASED HEURISTICS

Perspective network selection heuristic for the each active session that relies on a cost function is proposed in [9]. In this scenario, the mobile terminal maintains a list of current active sessions, arranged in priority order. Than, the cost function is evaluated for the highest priority service. The optimal target network is chosen by minimizing the per-session cost

$$\min C_s^n = \sum_s W_{s,j}^n Q_{s,j}^n , \quad E_{s,j}^n \neq 0 , \tag{1}$$

where $Q_{s,j}^{n}$ is the normalized QoS provided by network *n* for parameter *j* (e.g. bandwidth, delay) and service session *s* (video, voice). $W_{s,j}^{n}$ is corresponding weight coefficient which indicates the impact of the QoS parameter on the user or the network, and $E_{s,j}^{n}$ is the network elimination factor, indicating whether the minimum requirement of parameter *j* for service *s* can be met by network *n*. The algorithm of the proposed heuristic is shown in Fig. 1.

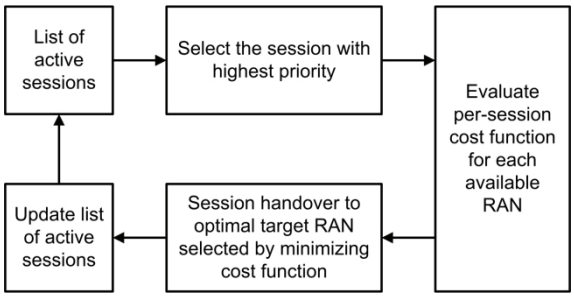


Fig. 1. Example of cost function based heuristic

Similar to the previous technique, the article [10] analyzes the application of cost function in the process of evaluating the qualitative performance of potential target networks. By using the normalization and weights distribution methods, cost function determines a network quality factor

$$Q_i = f(w_c \frac{1}{C_i}, w_s S_i, w_p \frac{1}{P_i}, w_d D_i, w_f F_i), \quad (2)$$

where C_i is the cost of service, S_i security, P_i power consumption, D_i network condition and F_i network performance, while the w_c, w_s, w_p, w_d and w_f are weights for each of the network parameters which are proportional to the significance of a parameter to the vertical handover decision. Due to the heterogeneous parameters, it is necessary to make a normalization of the function. High overall throughput and user's satisfaction can be regarded as major advantages of this heuristics.

The fundamental benefit of cost function usage and handover independent initiation for different services is reduced failure (blocking) probability. However, parameter normalization and weights coefficients determination techniques are not discussed.

IV. Neural Networks based Heuristics

Network selection heuristic based on the artificial neural network is proposed in [11]. Applied feedforward neural network topology, which consists of input, hidden and output layer, is shown in Fig. 2. The input layer is made of the *h* nodes representing different criteria for optimal network selection, while the hidden layer consists of the *n* nodes that represent the available access networks. Logistic sigmoid activation function $f(x) = 1/(1+e^{-x})$ is applied to determine the cost function. For the training process error backpropagation algorithm is used. Output layer is formed by a node that generates the identification of the optimal access network.

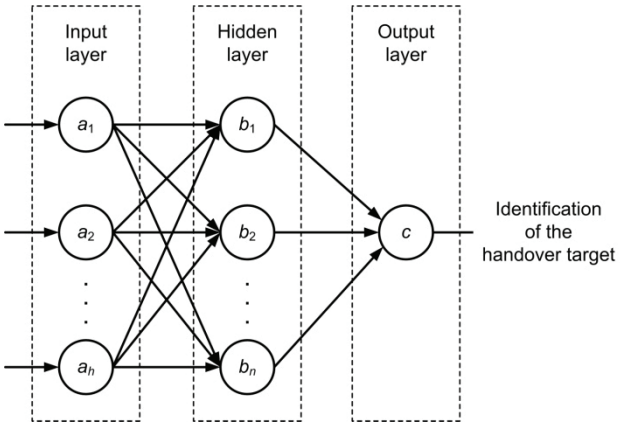


Fig. 2. Topology of the neural network as network selection heuristic

During the simulation, the authors in [11] adopted the same cost function as in [10]. The performed simulations have shown high accuracy and reliability of the model while selecting the optimal network. The lacks of the algorithm are reflected to the complexity of the system and to increased handover delays due to the training process.

V. Multi-Criteria Decision Making based Heuristics

With tight association to the nature of problem, a great number of the network selection heuristics available in the open literature are based on Multi-Criteria Decision Making (MCDM). MCDM tools relay on certain indices to estimate the performance of alternatives and finally rank them.

In [12] the authors develop a network selection mechanism for an integrated WLAN/cellular system. The design goal is to provide the user the best available QoS at any time. The suggested network selection mechanism relies on the combination of Analytic Hierarchy Process (AHP) and Grey Relational Analysis (GRA) of the multiple criteria analysis method. This method mathematically presents a complex solution and unnecessarily takes into account a large number of QoS parameters (delay, jitter, response time, bit error rate, etc.) only for 3G and WLAN networks. Processing a large number of parameters leads to the computational time increasing, while the terminal and infrastructure network elements are additionally loaded. Thus, this model is interesting from theoretical point of view, but not adequate for a direct implementation. These lacks are recognized in [13], but in general forms.

Network selection solution proposed in [14] represents interesting and promising solution while combining the heuristics of the fuzzy logic systems and MCDM (Fig. 3). In the process of handover initiation, proposed technique uses fuzzy logic analyzing the criteria such as: *RSS*, bandwidth (*B*), network coverage (*NC*) and terminal velocity (*V*). Based on 4 related functions and 81 predefined rules, a system determines whether handover is necessary or not. By application of AHP method and Saaty's scale on criteria such as cost of service,

preferred interface, battery status and QoS level, the optimal access network is determined.

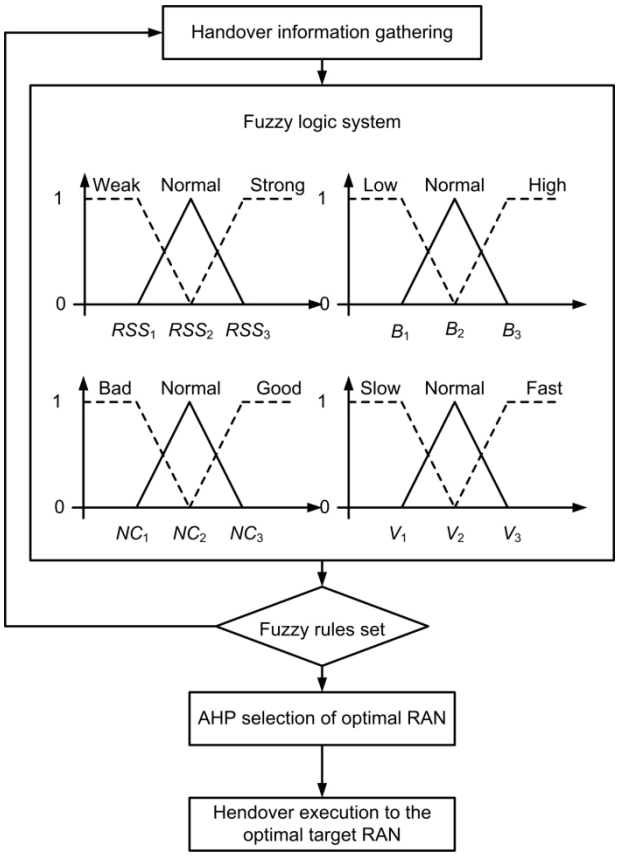


Fig. 3. Example of network selection heuristic based on fuzzy logic system and MCDM

On the other hand, by applying fuzzy logic in decision making process the number of unnecessary handover are reduced, as well as signaling traffic and handover delays. Inflexibility of the impact of user's preferences on the system is basic lack of applied AHP method, which could possibly be exceeded by using some other MCDM techniques, e.g. TOPSIS [15].

VI. Conclusion and Future Work

The scope of this research is to address the issue of network selection in vertical handover procedure. The main challenges involved in the network selection are pointed out and synopsis of approaches encountered in the open literature is presented. Currently proposed network selection heuristics for vertical handover require more significant challenges to overcome, before being successfully deployed in real environment. Selection of the most suitable heuristic is the crucial research direction in the field of heterogeneous wireless networks.

Concerning further research, all activities will be dedicated to possible quantitative evaluation of perspective network selection heuristics taking in to account various mobility models and traffic characteristics (e.g. traffic load, blocking probability, etc.).

Acknowledgement

This research is partially supported by the Ministry of Education and Science, Republic of Serbia, Project No. TR-32025.

References

[1] J. Akhtman, L. Hanzo, "Heterogeneous Networking: An Enabling Paradigm for Ubiquitous Wireless Communications", Proceedings of the IEEE, vol. 98, no. 2, pp.135-138, 2010.

[2] X. Gao, G. Wu, T. Miki, "End-to-End QoS Provisioning in Mobile Heterogeneous Networks", IEEE Wireless Communications, vol. 11, no. 3, pp. 24 – 34, 2004.

[3] ITU-R Rec. M.1645, "Framework and Overall Objectives of the Future Development of IMT-2000 and System beyond IMT-2000", 2003.

[4] 3GPP TS 24.312, "Access Network Discovery and Selection Function (ANDSF) Management Object (MO)", 2009.

[5] IEEE 802.21, "Media Independent Handover Services", www.ieee802.org/21.

[6] D. Niyato, E. Hossain, "Competitive Pricing in Heterogeneous Wireless Access Networks: Issues and Approaches", IEEE Network, vol. 22, no. 6, pp. 4 – 11, 2008.

[7] Y. Ji, et. al., "CPC-assisted Network Selection Strategy", 16th IST Mobile and Wireless Communications Summit, Budapest, Hungary, 2007.

[8] X. Yan, Y. A. Şekercioğlu, S. Narayanan, "A Survey of Vertical Handover Decision Algorithms in Fourth Generation Heterogeneous Wireless Networks", Computer Networks, vol. 54, no. 11, pp. 1848-1863, 2010.

[9] F. Zhu, J. McNair, "Multiservice Vertical Handoff Decision Algorithms", EURASIP Journal on Wireless Communications and Networking, vol. 2006, no. 2, pp. 1-13, 2006.

[10] N. Nasser, A. Hasswa, H. Hassanein, "Handoffs in Fourth Generation Heterogeneous Networks", IEEE Communications Magazine, vol. 44, no. 10, pp. 96-103, 2006.

[11] N. Nasser, S. Guizani, E. Al-Masri, "Middleware Vertical Handoff Manager: A Neural Network-based Solution", Proc. IEEE ICC´07, pp. 5671-5676, Glasgow, Scotland, 2007.

[12] Q. Song, A. Jamalipour, "Network Selection in an Integrated Wireless LAN and UMTS Environment Using Mathematical Modeling and Computing Techniques", IEEE Wireless Communications, vol. 12, no. 3, pp. 42-48, 2005.

[13] Q. Song, A. Jamalipour, "A Quality of Service Negotiation-Based Vertical Handoff Decision Scheme in Heterogeneous Wireless Systems", European Journal of Operational Research, vol. 191, no. 3, pp. 1059-1074, 2008.

[14] M. Kassar, B. Kervella, G. Pujolle, "An Intelligent Handover Management System for Future Generation Wireless Networks", EURASIP Journal on Wireless Communications and Networking, Article ID 791691, 12 pages, 2008.

[15] B. Bakmaz, Z. Bojkovic, M. Bakmaz, "Network Selection Algorithm for Heterogeneous Wireless Environment", Proc. IEEE PIMRC 2007, pp. 347-350, Athens, Greece, 2007.

Problems in Configuration of VPNs over MPLS Network

Veneta Aleksieva[1]

Abstract – **Recent years many internet service providers offer not only access to the Internet services, QoS, traffic engineering, but also Intranet VPNs, Extranet VPNs, VPNs with network management. Some problems exist during the creation, monitoring and usage of VPNs over MPLS. In this paper are presented these problems and are suggested some decisions, which are able to overcome them.**

Keywords – **MPLS, VPN, LSP, backbone networks**

I. INTRODUCTION

Recent years there is a very active research in the field of multiprotocol label switching (MPLS), and more and more networks are supporting MPLS [1]. One of the most notable applications of MPLS is traffic engineering (TE) [2], since label switching paths (LSPs) can be considered as virtual traffic trunks that carry flow aggregates generated by packet classification.

VPN solutions support remote access and private data communications over public networks as a cheaper alternative to leased lines. VPN clients communicate with VPN servers utilizing a number of specialized protocols as PPTP, L2TP etc. Building of VPNs in an enterprise network in the WAN transport uses Frame Relay, ATM or any other layer-2 transport technology, including MPLS [3]. There are two different methods to construct VPNs across IP backbone, i.e., CPE (Custom Premises Equipment) based and network based. Most of the current VPN implementations are based on CPE equipment.

IP/MPLS VPNs are compelling for many reasons. It defines IP VPNs's meaning that the VPN service accepts IP datagrams from customer sites and delivers them also as IP datagrams to other customer's sites. The connection between a customer's site and the core network, also referred to as an attachment circuit, may be a Layer 2 service such as ATM, SDH, Ethernet, but the VPN service handles only IP datagrams transmitted over this link[4,6]. One example is presented in Figure 1. (In this example: CE means customer edge, PE means provider edge, such as Ingress Label Switching router(LSR) or Egress LSR, which belong to the provider, P means providers' LSR.) For enterprises, they enable right-sourcing of WAN services and yield generous operational cost savings. For service providers, they offer a higher level of service to customers and lower costs for service deployment. When used with MPLS, the VPN feature allows several sites to interconnect transparently through a service provider's network. One service provider network can support several different IP VPNs. Each of these appears to its users as a private network, separate from all other networks. Within a VPN, each site can send IP packets to any other site in the same VPN.

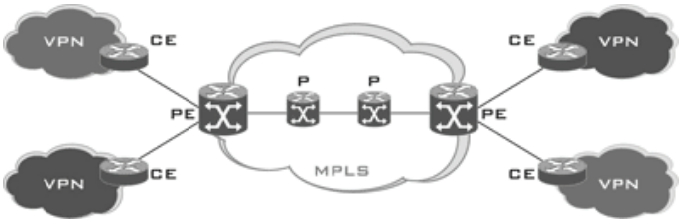


Fig. 1. Example of MPLS VPNs

In this paper are presented problems which related to MPLS VPNs and some suggestions for its overcoming. In particular, it is presented architecture of MPLS VPNs with QoS routing capability as well as some methods for operating QoS routing in MPLS VPNs.

II. ADVANTAGES AND DISADVANTAGES OF MPLS VPNS

When uses leased lines, each company will be responsible for the security of the information and network in a Point to Point connection. Here, in MPLS VPNs, security has the following characteristics:

- **Confidentiality, integrity, availability**— Security can usually be defined using these three properties. In the MPLS context, every VPN customer will have slightly different requirements for these parameters, but generally, customers will expect their data to be confidential, such that they are not accessible outside their VPN. They will expect the data not to change in transit, and they will expect the MPLS VPN service overall to be available to them.
- **Defense in depth**—It is good practice to add several layers of defense around everything that needs to be protected. This design principle is also important in MPLS networks.
- **Secure failure**—When the primary method fails, the backup method also needs to be secured appropriately. This is usually done through out-of-band access, mostly over the telephone network. It is important that this backup mode be as secure as the principal access mode.

MPLS VPNs are advantageous because they allow computers and devices to communicate with each other across large distances without using cables or wireless devices. MPLS VPNs cost less to maintain than other types of networks and can be created at any time by any computer in the world. Likewise, MPLS VPNs only have to look at the top label in a label stack in order to forward a data packet to another device. This allows MPLS VPNs to be much faster and more efficient than other types of networks.

[1] Veneta P. Aleksieva is with the Department of Computer Science and Engineering, Technical University of Varna, str."Studentska "1, 9010 Varna, Bulgaria, e-mail: ven7066@abv.bg

Although MPLS VPNs can be advantageous, they also have several disadvantages. The most notable disadvantage of an MPLS VPN is that it does not provide any security for the data packets that are sent out. This is because MPLS VPNs depend on each device within the network to forward the data packet to the next device. Therefore, once a data packet has been sent out, any device in the network could potentially intercept the data packet and view its contents. However, encryption protocols are available that they could be used in conjunction with an MPLS VPN.

Moreover, the benefits of the MPLS VPN Service for the customer are:

- Simple network implementation
- Easy to configure and manage
- QoS, CoS and better Traffic Engineering
- Easy network expansion at customer premises
- Easy introduction of new services as Multicasting, VoIP or hosting over the same link
- Security is the responsibility of MPLS Network
- Network is very reliable due to built in redundancy
- Flexible reconfiguration -instantaneous addition and deletion possible
- Less cost per link than leased lines
- Offer different level of service and protect specific part of traffic
- Traffic engineering gives maximal effectiveness of bandwidth usage
- Existing equipment gives possibility to use human resources with less qualification and less salary
- Faster then Layer 2 VPNs
- Cheaper than leased lines
- A single point of contact with access to a large number of licensed and certified carriers and local access providers
- A single point of contact for network performance and capacity management
- A network with enhanced flexibility and scalability which enables the customer to let its network grow with the growth of its business.

MPLS-VPNs are divided into:

- access channel,
- pick throughput ,
- quantity of sent packets,
- CoS,
- size of routing table,
- members of VPN,
- protocol between customer router and ingress label switching router (LSR).

In each of them each PE router maintains a number of separate forwarding tables. One of the forwarding tables is the "default forwarding table". The others are "VPN Routing and Forwarding tables", or "VRFs". Management of MPLS is based on database LIB (Label Information Base). Ingress LSR puts label to the packet when packet input in MPLS network, but Egress LSR deletes label from IP packet when packet leaves the MPLS cloud. Method for rerouting and making decisions with IP packets with/without labels is presented in Figure 2.

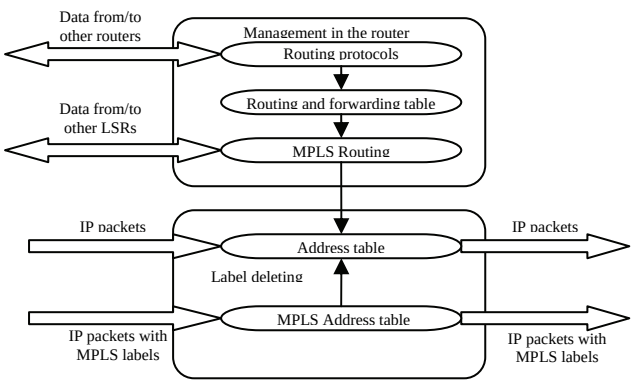


Fig. 2. Example of PE router's functions in MPLS VPNs

To be possible to monitor how work MPLS network, to predict some problems and manage them, there exist two main methods for managing LSRs:

- **with global routing table** (in Fig.3)- Loopback address of P and PE routers are inside in this table, but address of management workstation is inside in the VRF table. The connection between backbone MPLS network and this management workstation is with global static route in VRF table to the address of MPLS network and with global static route in global routing table to the address of management workstation.

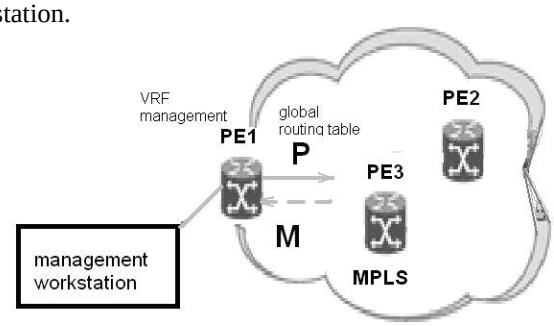


Fig. 3. Management of MPLS core with global routing table

- **with routing/forwarding table** (in Fig. 4) - This method is more simple. Here management network is directly connected to the interface, which is defined from global routing table, without association to VRF table.

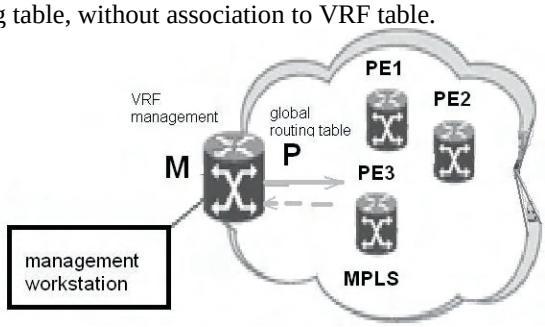


Fig. 4. Management of MPLS core with VRF table

III. Architecture of Hub and Spoke Model VPN

VPN decisions must differentiate different type of traffic quickly, to be possible ISP to group different customers and services. It is easy to make this with MPLS, because MPLS divides the traffic, protects it without encryption and

tunneling, offers scalable VPN service. All sites send traffic to the hub, which must know all sites for this VPN. If ISP must work with 100 sites, each of them with hub and spoke and 100 VPNs, logical topology will be created carefully and each device must carefully configure.

But if two VPNs have no sites in common, then they may have overlapping address spaces. Thus, a given address might be used in VPN V1 as the address of system S1, but in VPN V2 as the address of a completely different system S2. This is a common situation when each VPN uses a RFC 1918 [7] private address space. But MPLS overcomes this problem, because sends data based on labels, not based on IP addresses.

It is present the architecture as it is shown in Figure 5.

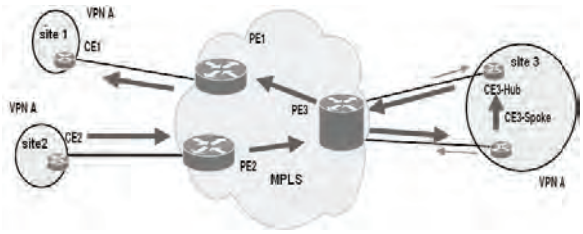


Fig. 5. Hub and spoke MPLS VPN

This structure has some advantages:

- ISP's work, when MPLS is used, is directly proportional to count of customer sites, which is included in VPN, while in Frame Relay it is directly proportional to the power 2 of count of customer sites.
- It supports optimal routing for customer traffic on ISP backbone, because here absent transit CEs.
- Customer doesn't manage own backbone, he only connect CE routers to the ISP.

On the other hand, this model has some disadvantages:

- Overload of P-routers with routing information –large resource of memory, CPU power, and bandwidth.
- In this architecture are existing customers with addresses schemes, which is difficult to co-ordinate with ISP's backbone topology and route aggregation absent.
- Because of the private addresses in customer's networks, unique addresses do not exist. In this case P-routers don't guarantee packet's delivery.
- In this architecture CE router hasn't the possibility to define where will send next packet. This gives chance to eavesdropping.

IV. Architecture of mesh model VPN

A MPLS VPN is built up by connecting MPLS sites through tunnels across IP backbone. Each MPLS site has a Bandwidth Broker (BB), which is to exchange route and signaling information and to manage and maintain VPN networks. Customer routers don't exchange each other information about routes. Data is sent from input customer router to ingress ISP router, then follows some LSRs to the egress LSR, then sent to the customer router in second site.

It is presented the architecture as it is shown in Figure 6.

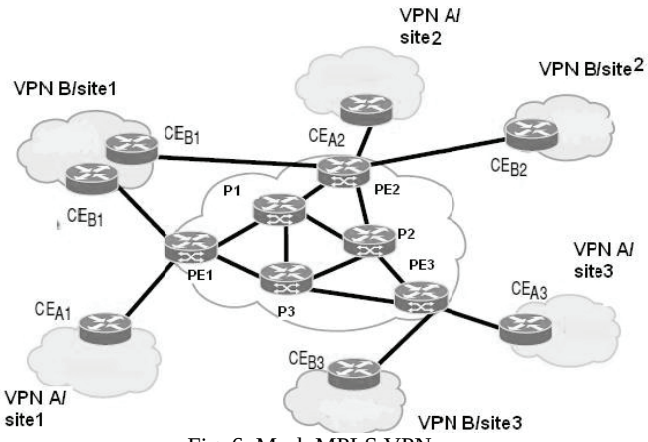


Fig. 6. Mesh MPLS VPN

This architecture has some disadvantages:

- To achieve optimal routing in customer's network, which uses ISP's network, it must be put a router in each end branch, which is connected to all other customers' routers in each branch. This means that this topology is full mesh.
- If used topology is different from full mesh topology, sometimes one customer's router sends packets to the central customer's router, which is placed in different customer's branch, using ISP network and this central customer's router makes decision about forwarding and send back packets to the destination router in the same source brunch. This means that this customer uses ISP's network pointless.
- If used topology is full mesh topology, the customer pays for virtual channels and ISP gives them resources.
- This topology is not well scalable, because here focus is on the links, not on the type of traffic.
- ISP must have tools to recognize different types of applications and based on this information to create security and QoS for customer's data.

On the other hand, this architecture gives some benefits:

- It is support large number VPN, without increasing of data quantity for routes, which is keeping in P-routers.
- It is not possible to send by chance traffic among VPNs, because each of them has own routing information.
- Routing table on PE is used only for directly connected to this router networks, and not for packets, which come from different PE. The route is calculate in the node, where is connected to the ISP backbone.
- If different sites use the same routes, they will consolidate in one routing table in PE, instead two different tables.

V. Configuration of MPLS for VPNs

Above are presented two different architectures for MPLS VPNs, but they use the same devices and this devices are configured in the same way, only the logical links among devices are different. When it must be configured MPLS for VPN, must be followed the next basic steps:

- Specify the interfaces used for communication between PE routers and between PE routers and provider routers:

LSR# edit protocols mpls interface interface-name

• For RSVP only, configure an MPLS LSP to the destination point on the PE router. The path name is defined on the source router only and it is unique between two routers.
LSR# edit label-switched-path path-name
• Specify the IP address of the LSP destination point, which is an address on the remote PE router.
LSR# set to ip-address
• Commit the configuration if you have finished configuring the device.
LSR# commit
• Configure OSPF with traffic engineering support on the PE routers.
LSR# edit protocols ospf traffic-engineering shortcuts
• Enable RSVP on interfaces that participate in the LSP. For PE routers, enable interfaces on the source and destination points. For P routers, enable interfaces that connect the LSP between the PE routers.
LSR# edit protocols rsvp interface interface-name
LSR# commit
• Configuring Routing Options for MPLS VPNs
• Configure the AS number.
LSR# set routing-options autonomous-system as-number
LSR# commit
• To configure a VPN routing instance on each PE router:
LSR# edit routing-instances routing-instance-name
LSR# set description “text”
• Specify the instance type, either l2vpn for Layer 2 VPNs or vrf for Layer 3 VPNs.
LSR# set instance-type instance-type
• Specify the interface of the remote PE router.
LSR# set interface interface-name
• Specify the route distinguisher using one of the following commands:
LSR# set route-distinguisheras-number:numberuser@host# set route-distinguisher ip-address:number
• Specify the policy for the Layer 2 VRF table.
LSR# set vrf-import import-policy-name vrf-export export-policy-name
• Specify the policy for the Layer 3 VRF table.
LSR# set vrf-target target:community-id
LSR# commit

VI. Conclusion

MPLS VPNs are used due to its distinguished benefits - fast forwarding, tunneling, etc. QoS routing is naturally used in MPLS VPNs for providing feasible routes with considerations on QoS constraints. QoS routing is beneficial for developing QoS guaranteed MPLS VPNs across IP networks. While other forms of VPN have desirable characteristics, only MPLS provides the network intelligence businesses demand with the reassurance of future capabilities. With its ability to reduce in-house IT resources, coupled with its inherent resilience, MPLS provides the most cost- effective and beneficial VPN solution. For very large organizations, MPLS VPNs offer additional virtualization options, along with advanced capabilities for rapid network failover (within 50 msec) and traffic management for optimizing the link usage. MPLS segmentation yields several benefits, including security through separation, isolation of unpredictable applications and traffic congestion, and prioritization of performance-sensitive applications.

In this paper, it is investigated both benefits and problems when introducing MPLS VPNs. Particularly, it is present architecture of MPLS VPNs with QoS routing capability and discuss some issues on running QoS routing in MPLS VPNs.

Acknowledgement

The work presented in this paper was supported within the project BG 051PO001- 3.3.04/13 of the HR Development OP of the European Social Fund 2007-2013

References

[1] Aleksieva V. P., Comparison Studies on Path Recovery Schemes in MPLS Network, ICEST’10, Macedonia, Bitola 2010, vol.2, p.497-500
[2] Awduche D. O., Malcolm J., Agogbua J., O`Dell M., McManus J. Requirements for traffic engineering over MPLS. IETF RFC 2702. – September 1999. – Available at: www.ietf.org/rfc/rfc2702.
[3] Behringer, M. H.; M. J. Morrow , MPLS VPN Security, Cisco Press, 2005, ISBN 1-58705-183-4,p.312
[4] Eric C. Rosen, Yakov Rekhter, BGP/MPLS IP VPNs, draft-ietf-l3vpn-rfc2547bis-03.txt, 2004
[5] B. Gleeson, et al: A Framework for IP Based Virtual Private Networks. IETF RFC2764, 2000
[6] E. Rosen, Y. Rekhter, BGP/MPLS IP Virtual Private Networks (VPNs), IETF RFC4364, 2006
[7] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, E. Lear, Address Allocation for Private Internets, IETF RFC1918, 2006

Concatenated "MMSE-Sequential Search" Algorithm for Multi User Detection in SDMA Uplink

Ilia Georgiev Iliev[1] and Mende Budzevski[2]

Abstract – **The concatenation of MMSE and sequential search is MUD method which combats the imperfect channel conditions and maintains low complexity at the receiver. In this work the combination of two methods is proposed, and it is regarded in Spatial Division Multiplexing Access scheme. The suppression of Multi Access Interference and MMSE error along with BER performance is studied in case of different number of users and receiving antennas.**

Keywords – **MUD, OFDM, MMSE.**

I. Introduction

Space Division Multiple Access (SDMA) based Orthogonal Division Multiplexing (OFDM) scheme has been presented as very attractive point of research recently. Considering every single antenna equipped user, communicating with the Base Station (BS) i.e. its multiple antennas, this uplink scheme can be regarded as Multiple Input Multiple Output (MIMO) model and inherits the benefits. Taking into consideration the OFDM approach, every single user transmits symbols within common frequency bandwidth as user differentiation is not maintained by allocating the users per different subcarriers. Once formed, OFDM symbols are transmitted over non-ideal channel that is approximated as flat slow Rayleigh fading channel in parallel with Additive White Gaussian Noise (AWGN).

A variety of multiuser detection (MUD) schemes were researched for user separation at the receiver, either linear or non-linear detectors. Minimum Mean Square Error (MMSE) MUD is a promising method to put up with the channel transfer function and restore the signal prior to the demodulation process. However, Multi Access Interference (MAI) caused by the imperfect channel condition can't be completely reduced by MMSE linear method and therefore, some non-linear methods are additionally invoked such as sequential search [1], genetic algorithms [2], parallel and successive interference cancellers [3] etc.

In this paper a potential sequential search (SS) is proposed to be concatenated with MMSE linear combiner in order to reduce the loss of information caused by MAI. This method was successfully proved to be functional in CDMA approach where MAI is the result of the non-ideal orthogonal spreading codes [4]. Moreover, modest complexity is maintained at the receiver in comparison with Maximum Likelihood (ML) non-linear detection which was found to give the best performance. Apart of the same cost function which is analyzed in both ML and SS, the latter tends to find the most accurate parallel combinations of users and compute the cost for every each of them. The minimum value of the cost drives up to the most probable combination that was sent by the users. Here in this paper, an iteration of this method is simulated and the performance of the system is estimated with or without the concatenation.

[1]Ilia Georgiev Iliev is with Dept. Of Radiocommunication and Videotechnologies in Faculty of Telecommunication in Tu-Sofia, N8, Kliment Ohridski bul. 1700, Sofia, Bulgaria. Email: igiliev@tu-sofia.bg.

[2]Mende Budzevski is PHD student with Dept. Of Radiocommunication and Videotechnologies in Faculty of Telecommunication in Tu-Sofia, N8, Kliment Ohridski bul. 1700, Sofia, Bulgaria. Email: mbudzevski@gmail.com

II. System Model

A. SDMA-MIMO-LFDMA channel model

Fig.1 shows the system model used in this paper where every user is equipped with single antenna, while the BS is equipped with multiple antennas. The number of users L and the number of receiving P antennas form SDMA-MIMO channel, essentially related with the PxL-dimensional matrix channel transfer function $\mathbf{H}_{PxL}$ in frequency domain.

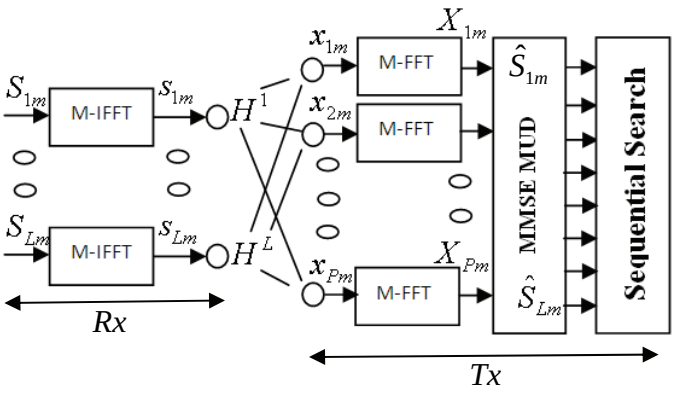


Fig.1 SDMA-MIMO Uplink Scheme

At every m-th subcarrier of the OFDM symbols received by the P-element receiver antenna array, the complex $Px1$-dimensional signal vector is formed as superposition of independently faded m-th subcarriers of the OFDM symbols associated with the L users and additionally distorted by AWGN. For every subcarrier the transfer function can be expressed as:

$$\mathbf{X}_m = \mathbf{H}\mathbf{S}_m + \mathbf{n} \quad (1),$$

where $\mathbf{X}_m$ is $Px1$ column vector that is constituted by symbols related to m-th subcarriers from every received OFDM

symbol from *P* antennas and $\mathbf{S}_m$ is *L*x1 column vector that is constituted by symbols related to *m*-th subcarriers from every transmit OFDM symbol per user. *Px*1-dimensional column vector **n** is the AWGN that exhibits a zero mean and a variance of σ_n^2. *PxL*-dimensional complex matrix **H** in frequency domain presents the channel transfer function. For example, the *l*-th column of the matrix:

$$\mathbf{H}^l = (\mathbf{H}_1^l ... \mathbf{H}_p^l), \quad l = 1,2...L \tag{2},$$

represents the complex transfer function that is associated with the transmission paths from the *l*-th user's antenna to each element of the *P*-element receiver antenna array.

B. Linear Detector – MMSE MUD

The OFDM symbols received at *P* antennas are transformed back to frequency domain by the *M*-dimensional Fast Fourier transform (*M*-FFT). Using the per-subcarrier approach, the column vector $\mathbf{X}_m$ is linearly combined with the aid of weight matrix **W**, resulting in:

$$\hat{\mathbf{S}}_m = \mathbf{W}^H \mathbf{X}_m \tag{3}$$

where **W** is *PxL*-dimensional matrix and the superscript *H* denotes the Hermitian transpose.

If we use (1) in case of avoiding IFFT and FFT operations, the last equation can be modified and invoked for the *l*-th user:

$$\begin{aligned} \hat{\mathbf{S}}_m &= \mathbf{w}^{lH} \mathbf{X}_m^l = \mathbf{w}^{lH} (\mathbf{H}\mathbf{S}_m + \mathbf{n}) \\ &= \mathbf{w}^{lH} \mathbf{H}^l \mathbf{S}_m^l + \mathbf{w}^{lH} \sum_{i=1, i \neq l}^{L} \mathbf{H}^i \mathbf{S}_m^i + \mathbf{w}^{lH} \mathbf{n} \end{aligned} \tag{4}$$

where $\mathbf{w}^{lH}$ is the *l*-th column of the *PxL* - matrix **W**.

Assuming the (4) and its components, the expression of undesired correlation matrix for the *l*-th user is:

$$\mathbf{R}_{a,I+N}^l = \mathbf{R}_{a,I}^l + \mathbf{R}_{a,N}^l = \sum_{i=1, i \neq l}^{L} \sigma_i^2 \mathbf{H}^i \mathbf{H}^{iH} + \sigma_n^2 \mathbf{I} \tag{5},$$

where σ_i^2 is the variance of the interfering users contribution, σ_n^2 is the variance of the AWGN and **I** is the *PxP* identity matrix.

The quality of the linear detector can be measured by the Signal to Noise and Interference Ratio (*SINR*). This parameter is defined by the variances of the desired signal, interfering signals and noise signal. Moreover, MAI that is supposed to be eliminated by the concatenation of the linear detector with additional non-linear detecting method is mutually related to $\mathbf{R}_{a,I}^l$ and it appears in the following expression for *SINR*:

$$SINR^l = \frac{\sigma_s^{(l)2}}{\sigma_I^{(l)2} + \sigma_N^{(l)2}} = \frac{\mathbf{w}^{lH} \mathbf{R}_{a,S}^l \mathbf{w}^l}{\mathbf{w}^{lH} \mathbf{R}_{a,N+I}^l \mathbf{w}^l} \tag{6}$$

Chanel estimation at the receiver is out of scope in this paper and therefore the channel matrix **H** is considered to be perfectly estimated. The linear combining using MMSE is method of finding the minimum mean square error of the cost function:

$$\Delta \mathbf{S}_m = \mathbf{S}_m - \mathbf{W}^H \mathbf{X}_m \tag{7}$$

The optimum weight matrix that minimizes the mean square of the cost function is constituted by the channel matrix and noise variance:

$$\mathbf{W}_{\text{MMSE}} = (\mathbf{H}\mathbf{H}^H + \sigma_n^2 \mathbf{I})^{-1} \mathbf{H} \tag{8}$$

Combining the receiving signal with (8) does not eliminate the MAI effect that basically depends proportionally on *L*. However, is very beneficial to deploy idealistic fading channel (flat and slow) because it eliminates the effect of Inter Sub-Carrier Interference (ICI). On the other hand, the perfect knowledge of the channel matrix at the receiver boosts up the performance of the linear detector and makes the process of MAI estimation straightforward.

C. Non-linear detector – Sequential search algorithm

In [4] this search algorithm was proposed in context of CDMA access and diversity at the receiver. Furthermore, the iterative form of the algorithm was proposed. In this paper, the same algorithm will be used and performed for QPSK modulated symbols. Additionally, the aim here is to prove that this concatenation of detectors is doable in context of SDMA-MIMO system model.

The method for mapping the modulated symbols with OFDM subcarriers was forced to maintain the fairness between the users [5] [6]. This means that *Q*, which is the number of modulated symbols per OFDM symbol, is equal per user. This leads to estimation of number of algorithm runs per OFDM symbol – *Q*. Hereafter, the algorithm specifications are described only for the *n*-th modulated symbol from every user, where $q = 1..Q$. If *Q* is equal to *M* (points of IFFT), the hardest case is encountered which is actually suitable for testing the efficiency of the non-linear algorithm in MAI environment.

The hard decision, made subsequent to MMSE MUD generates the input for the SS algorithm. For the *q*-th modulated symbols, the input combination can be expressed:

$$d_{Fq} = [d_{Fq1} ... d_{FqL}] \tag{9},$$

where d_{Fqi}, $l = 1..L$ is demodulated combination of *K* bits, where *K* is assigned to be '2' in term of the QPSK modulation.

Having in mind the Hamming distance between two combinations:

$$M_d = \{ d : H_d(d_{Fq}, d_q) = 1 \} \tag{10},$$

additional *LK* combinations are created based on the criteria.

Total of *LK*+1 combinations (including the initial one) are scope of the decision metric for SDMA-MIMO system, derived by the ML metric [3]:

$$\mathbf{\Delta}(\mathbf{S}) = \sum_{p=1}^{P} \mathbf{\Delta}_p(\mathbf{S}) \tag{11}$$

$$\mathbf{\Delta}_p(\mathbf{S}) = \left|\mathbf{X}_p - \mathbf{H}_p\mathbf{S}\right|^2 \tag{12}$$

The equation (11) solves the decision conflict so-called multiobjective optimization problem, since the optimization of the *P* metrics may result in more than one possible *L*-symbol solution. The equation (12) is the general expression of how the metric is calculated at every receiver. In (11) and (12), the parameter **S** is vector of *q*-th modulated symbols and has length of *L*. Once the needed symbols are extracted from the assigned subcarriers, the vector **S** can be formed and metric $\mathbf{\Delta}_p(\mathbf{S})$ can be calculated.

Solving (11) for every possible *LK*+1 combination will generate vector of *LK*+1 values, where the initial combination has index 1. From the theory of ML [3], the most likely combination that was sent is the one that leads to minimum value of the metric. Thus, the minimum is the factor in order to find the optimal combination from the pool. However, choosing the minimum value does not mean that the global minimum is attained. If the index with the minimum is not 1, than the optimal combination of the range becomes the initial input combination for the next iteration of the algorithm. As we stated above, in this paper the algorithm is broken at the first or third iteration and local minimum of the metric function is selected. Even with several iterations, this algorithm maintains the low computational complexity at the receiver in comparison to the extensive ML algorithm, based on full search.

III. Simulation Results

The algorithm is simulated in Matlab environment. The number of subcarriers assigned to particular user within OFDM symbol is equal per user and scheme without any user' differentiation is simulated. The modulation scheme is fixed to QPSK with Gray coding. The channel is modeled as AWGN with slow Rayleigh fading. Moreover, perfect channel estimation is assumed at the receiver.

These set options emphasize the MAI that depends only on the number of users. The other case which is more realistic for mobile channel is when fast fading (Doppler Effect) and frequency selectivity are taken into account. In that case MAI becomes more complex.

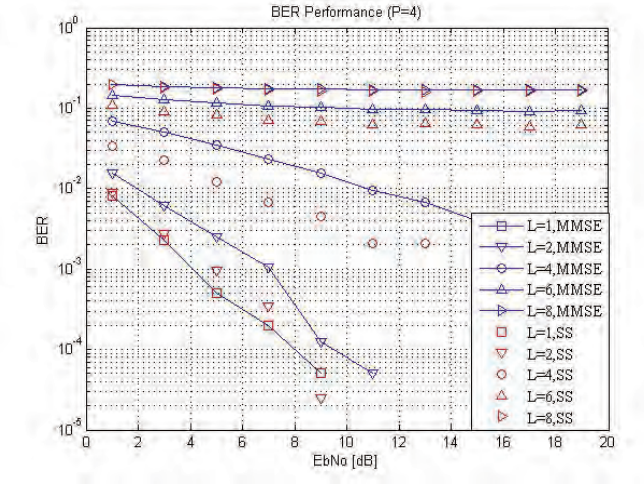


Fig.2 BER Performance – fixed P=4

Fig.2 shows the case of BER performance when the receiving antenna array has 4 elements. Both MMSE only and concatenation of MMSE and SS algorithms for MUD are simulated for different number of users. The parameter EbNo refers to the energy per bit to noise power spectral density ratio. When the case of single user is reviewed, we observe the lowest BER for the range of EbNo due to the lack of MAI effect. If L increases then the MAI will be obviously increased too due to the existing correlation between the users. This effect is inevitable and it is clearly shown on Fig.2 in case of L=2,4,6 and 8. All four multi-user scenarios has poorer performance than the single user scenario. The case L=8 has the worst performance and its curve for MMSE only MUD is on the top of the figure.

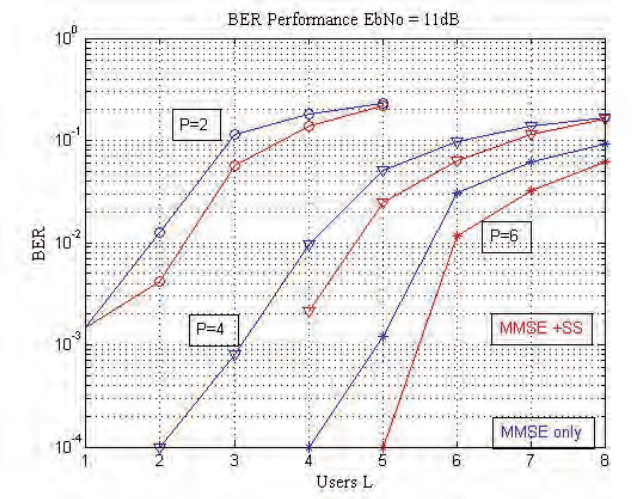


Fig.3 – BER performance, fixed EbNo=11dB

On the same figure, the effect of the concatenation of MMSE and SS MUD algorithms is presented as red symbols below every curve. The note here is that only three iterations have been invoked for every single simulated case. This fixed number of iterations can give us the real possibilities of this

non-linear method. Obviously, the number of iterations is chiefly and proportionaly related to the number of users and therefore, fixing the errors for cases with lower L is more significant rather than fixing for higher *L*. In this context, as can be seen on the figure, the curve for *L*=2 is closer to the single user case where MAI is not encountered.

Fig.3 is based on fixed EbNo to 11dB and the number of receiving antennas is a variable. Considering *P*, three cases are shown on the figure such as *P*=2,4 and 6. As stated previously, this scheme inherits the MIMO benefits and that is why the increasing of receiving antennas will gradually decrease the bit errors. Again, it can be concluded from these cases that MMSE SS MUD algorithm is more efficient when less users are communicating with the BS. Hence, less users, better quality. Summing up, the approaches to rich better quality when *L* is relatively high are either increasing of *P* or the number of iteration. The first approach is the matter of physical presence. The latter, will drastically increase the computational complexity of the receiver. The trade-off between two must be considered when receiver is designed.

A. Computational complexity of the algorithm

The computational complexity of each non-linear detector is defined as the average number of calculations of the metric function (11) and that is determined by the number of sums and multiplications made. In [4] SS algorithm has been compared to the Genetic and Optimum algorithm. In this case the number of computations depend on the number of combinations to be checked for minimum metric. They have been calculated $LK+1$ per iteration and per modulated symbol, where *K* was fixed to 2 because of the QPSK modulation. On the other hand, (11) is mutually dependant on (12) which means the number of receiving antennas play important role when the number of sums is considered in (12).

IV. CONCLUSION

The combination of linear detector MMSE and non-linear method SS in context of SDMA-MIMO system is practically doable and leads to some improvements on BER performance. This was shown in this paper either with only few iterations invoked of the SS algorithm. More iterations can be invoked and eliminate MAI in multi-user scenarios. The computational complexity of the proposed sub-optimal algorithm depends on the number of iterations set, but, however, it is lower in comparison to the ML optimal algorithm. Apart from the MAI elimination, the presented system model inherits the MIMO benefits and therefore, the more receiving antennas are implemented, the fewer bits are mistaken.

V. REFERENCES

[1] I. G. Iliev, M.Nedelchev, Performance Analysis of a suboptimal multi-user detection algorithm, ICEST 2007, Macedonia, pp.531 – 534.

[2] K.Yen and L.Hanzo, Hybrid genetic algorithm based multi-user detection schemes for synchronous CDMA system, in Proceedings of the IEEE Vehicular Technology Conference (VTC), (Tokio, Japan), May 15-18.

[3] L.Hanzo, M.Muster, B.J. Choi and T. Keller, OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting, Piscataway, NJ:IEEE Press, New York: Willey, 2003

[4] I. G. Iliev, M.Nedelchev, Antenna Diversity multi User Detection Algorithm for Synchronous CDMA System, ICEST 2009, Bulgaria, pp67-80

[5] H. G. Myung, D. Goodman, Single Carrier FDMA, a new interface for Long Term Evolution, Willey, 2009.

[6] Anja Sohl, Anya Klein, Comparison of Localized, Interleaved and Block Interleaved FDMA in terms of pilot multiplexing and channel estimation, EUSIPCO 2007, Poland

Approach to Formal Verification of Messaging Service Capability Server in Mobile Networks

Ivaylo Atanasov[1]

Abstract – **In this paper it is investigated how open access to messaging function in mobile networks may be deployed. The focus is on Open Service Access (OSA) interfaces for user interaction and Customized Application for Mobile network Enhanced Logic (CAMEL) applied to Short Message Service. Service Capability Server (SCS) makes translation between OSA interface methods and CAMEL Application Part (CAP) protocol. The formalism of labelled transition systems and the behavioural equivalence concept are used to verify the SCS functional behaviour.**

Keywords – **CAMEL, Open Service Access, Labelled Transition Systems, Bisimulation.**

I. Introduction

Open Service Architecture (OSA) allows third party access to communication functions in a network neutral way. Using OSA Application Programming Interfaces (API), application developers can create attractive applications without specific knowledge about underlying network technology and control protocols. Interoperability between OSA applications and specific network functions requires special type of application server called OSA Service Capability Server (SCS). The OSA SCS is responsible for translation of OSA interface method invocations into control protocol messages and vice versa.

The research focus is on OSAinterfaces for user interaction and Customized Application for Mobile network Enhanced Logic (CAMEL) applied to Short Message Service (SMS). The OSA User Interaction (UI) service provides API for call-related and call-unrelated user interactions [1]. The UI supports sending information or sending and collecting information. The mappings of OSA UI API onto CAMEL Application Part (CAP) protocol in the context ofSMS is defined in [2]. Some implementation aspects of CAMEL messaging service and OSA messaging service are discussed in [3,4] but no interworking issues are considered. In order to make interface to protocol translation, the OSA SCS needs to maintain two mutually synchronized state machines representing the application view on UI and protocol states. In the paper, we suggest a formal approach to verification of OSA SCS using the formalism of Labelled Transition Systems and the concept of bisimulation. The approach may be used for automatic generation of test cases during the OSA SCS functional verification [6].

The paper is organized as follows. In Section II, we discuss aspects of OSA deployment in a mobile network with CAMEL architecture.The formalism for Labelled Transition Systems is briefly introduced in Section III. A formal description of OSA SCS behavior is given in Section IV. Section V presentsformal descriptions of CAMEL state models for SMS events. Finally, the behavioral equivalence of state machines of OSA UI model and CAMEL SMS modelsis proved in Section VI.

[1]The author is with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, E-mails: iia@tu-sofia.bg.

II. Functional Architecture for Open Access to Messaging Functions

A functional architecture for deployment of OSA UI interfaces in CAMEL network is presented in Fig.1. Toward the network, the OSA SCS performs functions of CAMEL gsmSCF (Service Control Function) which provides CAMEL service logic. The network node - Mobile services Switching Center (MSC) or Serving GPRS Support Node (SGSN), provides functions of gsmSSF (Service Switching Function) which is responsible for switching between SMS processing and service logic in gsmSCF. The SMS-Center (SMSC) is a node where short messages are stored before delivering.

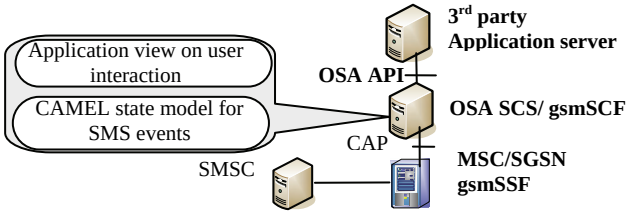


Fig. 1. OSA UI interface deployment in CAMEL network

In [1], it is defined a model that represents the application view of user interactions. In [5], two CAMEL models for SMS events are defined, one for mobile originating short messages and another for mobile terminating short messages. The behavior of OSA SCS regarding user interactions for short messaging needs to correspond to the specified models of both OSA application and CAMEL service logic. The formal specification of the models allows proving the behavioral equivalence, and hence the interoperability of OSA user interaction control and CAMEL service control.

III. Labelled Transition Systems and Behavioural Equivalence

To prove formally behavioral equivalence between state machines, the notion of *Labelled Transition Systems* is used [6].

Definition 1: A *Labelled Transition System* (LTS) is a quadruple $(S, Act, \rightarrow, s_0)$, where S is countable set of states,

*Act*is a countable set of elementary actions, $\rightarrow \subseteq S \times Act \times S$ is a set of transitions, and $s_0 \in S$ is the set of initial states.

We will use the following notations:

- $s \xrightarrow{a}$ s' stands for the transition (*s, a, s'*);
- $s \xrightarrow{a}$ means that $\exists$ s': s $\xrightarrow{a}$ s';
- $s \overset{\mu}{\Rightarrow} s_n$, where $\mu = a_1, a_2, ..., a_n : \exists s_1, s_2, \ldots, s_n$, such that $s \xrightarrow{a_1} s_1 ... \xrightarrow{a_n} s_n$;
- $s \overset{\mu}{\Rightarrow}$ means that $\exists$s', such as $s \overset{\mu}{\Rightarrow} s'$;
- $\overset{\hat{\mu}}{\Rightarrow}$ means $\Rightarrow$ if $\mu \equiv \tau$ or $\overset{\mu}{\Rightarrow}$otherwise,

where τ is one or more internal actions. More detailed notation description can be found in [6].

The concept of *bisimulation* [7] is used to prove that two LTSs expose equivalent behavior. The strong bisimulation possesses strong conditions for equivalence which are not always required. For example, there may be internal activities that are not observable. The weak bisimulation ignores the internal transitions.

Definition 2: [7] Two labelled transition systems $T = (S, Act, \rightarrow, s_0)$ and $T' = (S', Act, \rightarrow', s_0')$ are *weakly bisimilar* if there is a binary relation $U \subseteq S \times S'$ such that if $s_1 U\ t_1$: $s_1 \subseteq S$ and $t_1 \subseteq S'$ then $\forall a \in Act$:

- $s_1 \overset{a}{\Rightarrow} s_2$ implies $\exists\ t_2$: $t_1 \overset{\hat{a}}{\Rightarrow}' t_2$ and $s_2 U\ t_2$;
- $t_1 \overset{a}{\Rightarrow}' t_2$ implies $\exists\ s_2$: $s_1 \overset{\hat{a}}{\Rightarrow} s_2$ and $s_2 U\ t_2$.

IV. Formal Description of OSA User Interaction Model

The application view on UI object is defined in [1]. The behavior of the UI object is described by finite state machine. In Null state, the UI object does not exist. The UI object is created when the createUI()method is invoked or a network event is reported by reportEventNotification()method. In Active state, the UI object is available for requet messages which have to be sent to the network. Both sendInfoAndCollectReq()and sendInfoReq()methods have a parameter indicating whether it is a final request and the UI object has to be released after the information has been presented to the user. In Active state, when a fault is detected on the user interaction, an error is reported on all outstanding requests. A transition to Release Pendingstate is made when the application has indicated that after a certain message no further messages need to be sent to the end-user. There might be, however, still a number of messages that are not yet completed. After the last message is sent or when the last user interaction has been obtained, the UI object is destroyed. In Finished state, the user interaction has ended. The application can only release the UI object. A simplified state transition diagram for UI object is shown in Fig.2.

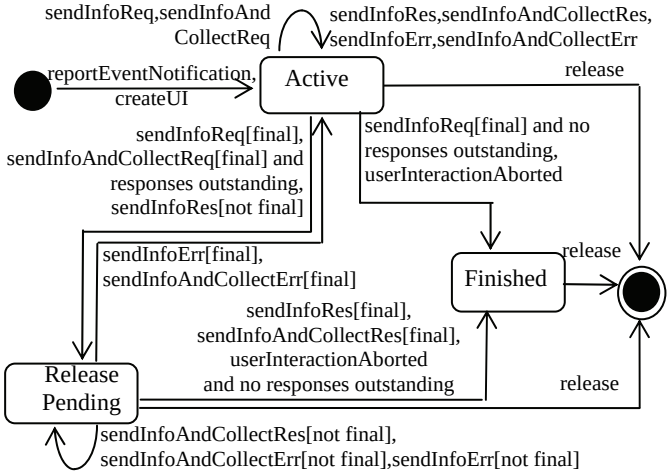


Fig. 2. OSA application view on the UI object

By $T_{AppUI} = (S_{AppUI}, Act_{AppUIH}, \rightarrow_{AppUI}, s_0)$ we denote a LTS representing the OSA application view on UI object where:

- S_{AppUI} = { Null, Active, ReleasePending, Finished };
- Act_{AppUI} = { createUI, reportEventNotification, sendInfoReq, sendInfoAndCollectReq, sendInfoRes, sendInfoErr, sendInfoAndCollectRes, sendInfoAndCollectRes, sendInfoAndCollectErr, sendInfoAndCollectErr, userInteractionAborted, release };
- $\rightarrow_{AppUI}$ = { Null createUI Active,
  Null reportEventNotification Active,
  Active sendInfoReq Active,
  Active sendInfoRes Active,
  Active sendInfoAndCollectReq Active,
  Active sendInfoAndCollectRes Active,
  Active sendInfoErr Active,
  Active sendInfoAndCollectErr Active,
  Active release Null,
  Active sendInfoReq ReleasePending,
  Active sendInfoRes ReleasePending ,
  ReleasePending sendInfoErr Active,
  ReleasePending sendInfoErr ReleasePending,
  ReleasePending sendInfoRes Finished,
  ReleasePending userInteractionAborted Finished,
  ReleasePending release Null,
  Finished release Null,
  Active sendInfoAndCollectReq ReleasePending,
  ReleasePending sendInfoAndCollectErr Active,
  ReleasePending sendInfoAndCollectRes ReleasePending,
  ReleasePending sendInfoAndCollectErr ReleasePending,
  ReleasePending sendInfoAndCollectRes Finished,
  Active sendInfoReq Finished,
  Active userInteractionAborted Finished };
- $s_{0'}$ = { Null }.

V. Formal Description of CAMEL State Models for SMS Events

CAMEL defines state models for SMS events which provide the possibility of triggering services as a result of messaging events [5]. Service logic may brake into sending a short message. CAMEL doesn't inspect the content of any message and it doesn't trigger services on that basis; the only events CAMEL triggers a service are the ones regardingsignaling conditions. CAMEL can recognize the origin and destination addresses of the message and can use this as criteria to start a service.

The Mobile Originating(MO) SMS state model is used to describe the actions in MSC and SGSN during Mobile Originating SMS and it is shown in Fig.3. The model is started when the gsmSSF sends to the gsmSCF and InitialDPSMS message.

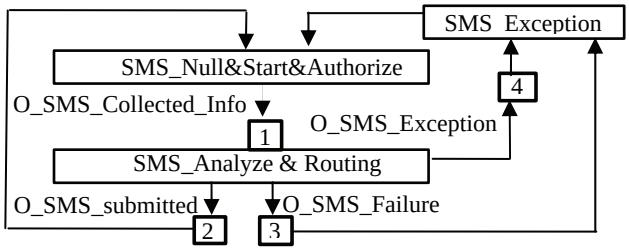


Fig. 3.CAMEL state model for MO SMS events

Entry events for SMSNull&Start&Authorize state are about previous MO SMS transfer to the SMSC completed or exception event. The detection point SMS_Collected_Info indicates that the subscription information is analysed and a MO short message is received. The CAMEL control flow between gsmSCF and gsmSSF corresponding to this detection point is shown in Fig.4.

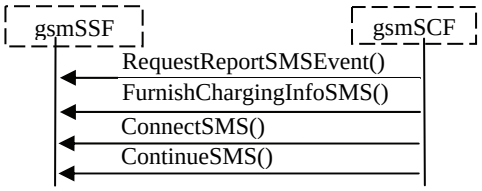


Fig. 4.CAP message flow on sending a short message

In SMSAnalyze&Routing state, information being analysed and/or translated to determine routing address of the SMSC and the short message is sent to the SMSC. The O_SMS_Submitteddetection point indicates that the short message is successfully submitted to the SMSC and it is reported by gsmSSF to gsmSCF sending anEventReportSMS message. The O_SMS_Failure detection point is armed when a failure has occurred in the SMS or command submission. The failure may have occurred internally in the MSC or SGSN or may have occurred externally, e.g. in the SMSC. Inthis case, the gsmSSF reports an error to the gsmSCF sending EventReportSMS message. An exception situation occurs when the gsmSSF reports DialogueAbort or DialogueError to the gsmSCF.

We decompose the SMSNull&Start&Authorize state into two states: $SMSNull_O$ and $Start\&Authorize_O$ to distinguish between different short messages in user interactions. Using the notations of LTS, we describe formally the CAMEL state model for MOSMS events by T_{OSMS} = (S_{OSMS}, Act_{OSMS}, $\rightarrow_{OSMS}$, s_0') where

- S_{OSMS}= {$SMSNull_O$, $Start\&Authorize_O$, SMSAnalyze&Routing };
- Act_{OSMS} = { $InitialDPSMS_O$, RequestReportSMSEvent, ConnectSMS, FurnishChargingInfoSMS, ContinueSMS, EventReportSMS, Release, EventReportSMS,DialogueAbort, DialogueError};
- $\rightarrow_{OSMS}$ = { $SMSNull_O$ $InitialDPSMS_O$ $Start\&Authorize_O$, $Start\&Authorize_O$RequestReportSMSEvent SMSAnalyze&Routing, SMSAnalyze&Routing FurnishChargingInfoSMSSMSAnalyze&Routing, SMSAnalyze&Routing ConnectSMSSMSAnalyze&Routing, SMSAnalyze&Routing ContinueSMSSMSAnalyze&Routing, SMSAnalyze&Routing EventReportSMSSMSNull$_O$, SMSAnalyze&Routing EventReportSMSStart&Authorize$_O$, SMSAnalyze&Routing EventReportSMSSMSNull$_O$, SMSAnalyze&Routing ReleaseSMSNull$_O$, SMSAnalyze&Routing DialogueAbortSMSNull$_O$, SMSAnalyze&Routing DialogueErrorSMSNull$_O$};
- s_0' = {$SMSNull_O$}.

The Mobile Terminating(MT) SMS state model is used to describe the actions in MSC and SGSN during Mobile Terminating SMS, and it is shown in Fig.5.

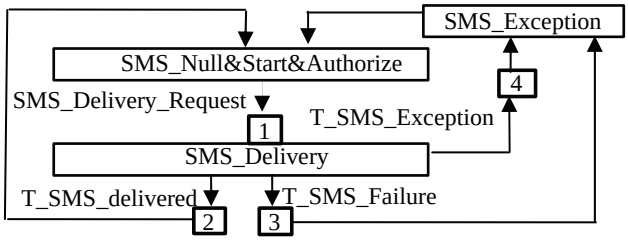


Fig. 5.CAMEL state model for MT SMS events

The $SMSNull\&Start\&Authorize_T$ state is entered when a short message is received in MSC from SMS-gateway MSC or previous MT SMS transfer completed or an exception event occurs. The SMS_Delivery_Request detection point indicates that the a mobile terminating SMS is received and it corresponds to the EventReportSMS message.The T_SMS_Delivereddetection point indicates that the short message has been successfully delivered which corresponds to theEventReportSMS message. The T_SMS_Failuredetection point indicates that the short message has failed which is reported byEventReportSMS message.

We decompose the SMSNull&Start&Authorize state into two states: $SMSNull_T$ and $Start\&Authorize_T$ to distinguish between different short messages in user interaction.Using the notations of LTS, we describe formally the CAMEL state model for MTSMS events by T_{TSMS} = (S_{TSMS}, Act_{TSMS}, $\rightarrow_{TSMS}$, s_0'') where

- S_{TSMS} = {$SMSNull_T$, $Null\&Start\&Authorize_T$, SMSDelivery};
- Act_{TSMS} = { $InitialDPSMS_T$, Release, EventReportSMS, EventReportSMS, EventReportSMS, DialogueAbort, DialogueError };
- $\rightarrow_{TSMS}$ = { $SMSNull_T$ $InitialDPSMS_T$ $Start\&Authorize_T$, $Start\&Authorize_T$EventReportSMS SMSDelivery, SMSDelivery EventReportSMSSMSNull$_T$, SMSDelivery EventReportSMSStart&Authorize$_T$, SMSDelivery EventReportSMSSMSNull$_T$, SMSDelivery ReleaseSMSNull$_T$, SMSDeliveryDialogueAbortSMSNull$_T$, SMSDeliveryDialogueErrorSMSNull$_T$};
- s_0'' = { $SMSNull_T$ }.

VI. Behavioural Equivalence between State Machines in OSA and CAMEL

To prove the interoperability between user interaction model in OSA and CAMEL state machines for SMS events we have to prove that the state machine representing the OSA user interactions and the CAMEL state machines for SMS events expose equivalent behavior. The behavioral equivalence is proved using the concept of weak bisimilarity.

Proposition 1: The labelled transition systems T_{AppUI}, T_{OSMS} and T_{TSMS} are weakly bisimilar.

Proof 1: To prove the bisimulation relation between labelled transition systems, it has to be proved that there is a bisimulation relation between their states. With U it is denoted a relation between the states of T_{AppUI}, $T_{OSMS}$and T_{TSMS} where U={(Null, $SMSNull_O$, $SMSNull_T$),(Active, $Start\&Authorize_O$, $Start\&Authorize_T$)}. Table 1 presents the bisimulation relation between the states of of T_{AppUI}, $T_{OSMS}$and T_{TSMS} which satisfies Definition 2.In [2], a mapping between the OSA User Interaction interface methods and CAP messages in the context of SMS is defined. We use this mapping to showaction's similarity. Based on the bisimulation relation between the states of T_{AppUI}, $T_{OSMS}$and T_{TSMS} it can be stated that the state machines expose equivalent behavior.

As an example,an application that uses OSA UI interfaces creates UI object and requests a message to be sent to the user, which starts the CAMEL state model for MO SMS events. Whent the SMS is submitted in the CAMEL network, the application is informed about the result of requested operation.

VII. CONCLUSION

In this paper an approach to formal description of OSA user interaction and CAMEL models for SMS events is suggested. The concept of bisimulation is used to prove the behavioral equivalence.

The approach is useful in testing the conformance of a black-box implementation of OSA SCS with respect to a specification, in the context of reactive systems.

ACKNOWLEDGEMENT

The research is funded by the Project DO-02-135/2008, funded Bulgarian Ministry of Education, Youth and Science.

REFERENCES

[1] 3GPP TS 29.198-5 "Open Service Access (OSA); Application Programming Interface (API); Part 5: User Interaction Service Capability Feature (SCF)", Release 9, v9.0.0, 2009.

[2] 3GPP TR 29.998-05-5 "Open Service Access; Application Programming Interface (API) Mapping for OSA: Part 5: User Interaction Service Mapping; Subpart 4: API to SMS Mapping", Release 9, v9.0.0,2009.

[3] T. Magedanz, M. Sher, "IT-based Open Service Delivery Platforms for Mobile Networks From CAMEL to the IP Multimedia System", in *Handbook of Mobile Middleware*, Auerbach Publishers, 2006.

[4] M. Wegdam, D. Plas, and M. Unmehopa, Validation of the Open Service Access API for UMTS Application Provisioning Proc. of PROMS 2001, LNCS 2213, pp. 210-221, 2001

[5] 3GPP TS 23.078, "Customized Applications for Mobile network Enhanced Logic (CAMEL) Phase 4; Stage 2", v10.0.0, 2010.

[6] T. J´eron, Symbolic "Model-based Test Selection",*Electronic Notes in Theoretical Computer Science*, 240, Elsevier, 2009, pp.167–184.

[7] X. Chena, R. Nicola, "Algebraic characterizations of trace and decorated trace equivalences over tree-like structures", *Theoretical Computer Science*, 2001, pp. 337–361.

.

Table 1.Bisimulation Relation between OSA User Interaction states and states of CAMEL models for SMS events

Transitions in T_{AppUI}	Transitions in T_{OSMS}	Transitions in T_{TSMS}
Null createUI Active Null reportEventNotification Active	$SMSNull_O$ InitialDPSMS$_O$ $Start\&Authorize_O$	$SMSNull_T$ InitialDPSMS$_T$ $Start\&Authorize_T$
Active sendInfoReq Active, Active sendInfoRes Active, Active sendInfoAndCollectReq Active, Active sendInfoAndCollectRes Active, Active sendInfoReq ReleasePending, Active sendInfoRes ReleasePending ReleasePending sendInfoErr ReleasePending, Active sendInfoAndCollectReq ReleasePending, ReleasePending sendInfoAndCollectRes ReleasePending, ReleasePending sendInfoAndCollectErr Active	$Start\&Authorize_O$ RequestReportSMSEvent SMSAnalyze&Routing, SMSAnalyze&Routing FurnishChargingInfoSMSSMSAnalyze&Routing, SMSAnalyze&Routing ConnectSMSSMSAnalyze&Routing, SMSAnalyze&Routing ContinueSMSSMSAnalyze&Routing, SMSAnalyze&Routing EventReportSMSSM$Start\&Authorize_O$	$Start\&Authorize_T$ EventReportSMS SMSDelivery, SMSDelivery EventReportSMSStart&Authorize$_T$
Active release Null, Active sendInfoReq Finished, ReleasePending sendInfoRes Finished, Finished release Null	$Start\&Authorize_O$ RequestReportSMSEvent SMSAnalyze&Routing, SMSAnalyze&Routing FurnishChargingInfoSMSSMSAnalyze&Routing, SMSAnalyze&Routing ConnectSMSSMSAnalyze&Routing, SMSAnalyze&Routing ContinueSMSSMSAnalyze&Routing, SMSAnalyze&Routing EventReportSMSSMS$Null_O$	$Start\&Authorize_T$ EventReportSMS SMSDelivery, SMSDelivery EventReportSMSSMS$Null_T$
Active sendInfoErr Active, Active sendInfoAndCollectErr Active, ReleasePending sendInfoAndCollectErr ReleasePending, ReleasePending sendInfoAndCollectRes Finished, ReleasePending userInteractionAborted Finished, ReleasePending release Null, Active release Null, ReleasePending sendInfoErr Active, ReleasePending userInteractionAborted Finished, ReleasePending release Null, Active userInteractionAborted Finished, Finished release Null	$Start\&Authorize_O$ RequestReportSMSEvent SMSAnalyze&Routing, SMSAnalyze&Routing FurnishChargingInfoSMSSMSAnalyze&Routing, SMSAnalyze&Routing ConnectSMSSMSAnalyze&Routing, SMSAnalyze&Routing ContinueSMSSMSAnalyze&Routing, SMSAnalyze&Routing EventReportSMSSMS$Null_O$, SMSAnalyze&Routing Release$SMSNull_O$, SMSAnalyze&Routing DialogueAbort$SMSNull_O$, SMSAnalyze&Routing DialogueError$SMSNull_O$	$Start\&Authorize_T$ EventReportSMS SMSDelivery, SMSDelivery EventReportSMSSMS$Null_T$, SMSDelivery Release$SMSNull_T$, SMSDelivery DialogueAbort$SMSNull_T$, SMSDelivery DialogueError$SMSNull_T$

Robust Header Compression for More Efficiency in Real-Time Transfer Date

Borislav Naydenov[1], Petar Petrov[2] and Alexander Milev[3]

Abstract – **The Robust Header Compression (ROHC) is a method to reduce the traffic. In the context of the problem, the objective is to analyze the principle of use and technical constraints of implementing ROHC and make a comparative assessment of appropriate use in wireless communication systems in terms of speed up of data transmission, occupied VoIP bandwidth and reliability of decompressed information.**

Keywords – **Robust Header Compression, VoIP traffic, Wireless LAN.**

I. Introduction

The evolution of telecommunication networks requires the use of approaches which allow to reduce the time, as in the implementation of interactive links and the transfer of data. One of the possible ways to optimize the time when transferring large files is through data compression. Data compression is widely studied and optimized. While the networks evolve to provide more bandwidth, applications, services and customers of these applications compete for this band. For network operators is essential to provide Quality of Service (QoS) in order to attract more customers and encourage them to use their network as much as possible.

Wireless networks are characterized by probability of bit error rates, mainly due to interference and greater delay. They are difficult to reach the requirements for wider bandwidth. It is therefore necessary available resources to be applied as efficiently as possible. The TCP header compression (HC) reduces overhead. The reduction in overhead for TCP traffic results in a corresponding reduction in delay; TCP header compression is especially beneficial when the TCP payload size is small, for example, for interactive traffic such as Telnet. In many services and applications such as VoIP, interactive gaming, messaging and other, payload in the IP packet is almost the same size, even smaller than the excess information, such as header. For connections of the type end to end, including multiple nodes, these protocol headers are important, but a links with type hop-to-hop, this service information is useless. It is possible to perform compression of header information, which will provide up to 90 percent size reduction of packet length. Thus reduces the load and save bandwidth of connection. IP header compression also provides significant advantages such as reduction of packet loss and improved interactive response time [1].

[1]Borislav Naydenov is with the Faculty of Electronic, Technical University - "Studentska" 1, Varna, Bulgaria, E-mail: borna@abv.bg.

[2]Petar Petrov is with the Faculty of Electronic, Technical University - "Studentska" 1, Varna, Bulgaria, E-mail: rls1@abv.bg

[3]Aleksandar Milev is with the Faculty of Mathematics and Informatics, University of Shumen, Bulgaria, E-mail: alex_milev@yahoo.com.

II. Actuality of the Problem

If necessary to transfer multiple small volumes of information with minimum delay in packet networks is possible to use a standardized approach to compress the header of packets. This question is especially actual in the transmission of voice information using Internet Protocol (Internet Protocol - IP). In slow serial connections such as wireless links in wireless local area networks (WLAN) and mobile communication networks, phone lines and other low speed lines, the use of packets with HC results in a significant reduction in the whole time to transmit information. There are other reasons that which demand to use the compression on slow and medium speed lines. The duration is reduced for interactive response time and packet loss rate over lossy links is getting smaller. This header compression scheme does not require that all packets in the same stream passes through the compressed link. However, for TCP streams the difference between subsequent headers can become more irregular and the compression rate can decrease. Neither is it required that corresponding TCP data and acknowledgment packets traverse the link in opposite directions.

For wireless links are proposed option for compression, called robust - Robust Header Compression [2]. ROHC involved mechanism for compression, which aims to achieve sustainability in relation on the packet loss and maximum efficiency in compression.

In the context of the problem, the objective is to analyze the principle of use and technical constraints of implementing ROHC and make a comparative assessment of appropriate use in wireless communication systems in terms of speed up of data transmission, occupied bandwidth and reliability of decompressed information.

III. Basic Approaches of Packet Header Compression

As a first implementation of the compression algorithm may point HC of packets of the Transport Protocol (Compressed Transmission Control Protocol - CTCP), standardized in RFC 1144 [3]. To make the optimization of service information, which are conveyed repeatedly with each packet in the slow line between two nodes, we need a compression protocol to be installed on network devices at both ends of the line. In this compression is achieved reductions of greater part of the header information of the Transmission Control Protocol (TCP), as is done shrinking it from 40 octets to 4 octets. Compressor, which implemented CTCP, detected retransmitted repeatedly and no amended on transport level information. After initial sending it stops retransmission. In the headers of next packets are put delta

encoding, which characterizes the successive changes in next header fields. This is known as a mechanism to adjust the context. For it is not necessary any additional signaling between the compressor and decompression.

The HC was improved when it was used to IP Compression (IP Header Compression - IPHC) with CTCP, RFC 2507 [4]. The mechanism to adjust the packets in CTCP is reinforced with a negative acknowledgment, called message CONTEXT_STATE, which accelerates the correction, when was detected an error. In RFC 2507 described how decreased IP and TCP headers per hop over point to point links.

IPHC does not compress the header of Real-time Transport Protocol - RTP. Compression of RTP header (Compressed RTP - CRTP) is an extension of the IPHC. CRTP compresses the 40 octets header packets including protocols IPv4/UDP/RTP to 2 octets. This is true if not activated error checking to header of the User Datagram Protocol – (UDP). Upon activation of this check the compression make the length of header to be 4 octets. This compression is not from one end to another end of connection, but from line to line [5]. Compression from line to line is characterized by good performances in which are small delays and low losses. IP / UDP / RTP compression is used with IPv4 and IPv6.

CRTP is characterized by loss of packets on slow lines, which in turn leads to the failure of decompression of next successive packets. With the implementation of IPHC is introduced a local adjustment mechanism called TWICE. But with this mechanism, CRTP does not reduce the number of lost packets. To improve the stability of the compression algorithm is proposed an improved version of CRTP - eCRTP [5], which is at the expense of reduced compression.

IV. Robust Header Compression

ROHC introduces a new mechanism for compression, which increases the resistance on the compression efficiency [2]. ROHC is expected to be the preferred compression mechanism over links where compression efficiency is important. However, ROHC was designed with the same link assumptions as CRTP, e.g., that the compression scheme should not have to tolerate misordering of compressed packets between the compressor and decompressor, which may occur when packets are carried. CRTP does not perform well on such links: packet loss results in context corruption and due to the long delay, many more packets are discarded before the context is repaired. To correct the behavior of CRTP over such links, a few extensions to the protocol are specified. It is based on Compressed Real-time Transport Protocol (CRTP), the IP / UDP / RTP header compression described in RFC 2508. The extensions aim to reduce context corruption by changing the way the compressor updates the context at the decompressor. With these extensions, CRTP performs well over links with packet loss, packet reordering and long delays [6]. ROHC was developed with wireless links as the main target, and introduced new compression mechanisms with the primary objective to achieve the combination of robustness against packet loss and maximal compression efficiency. If a packet that includes an update to some context state values is lost, the state at the decompressor is not updated. The shared state is now different at the compressor and decompressor. When the next packet arrives at the decompressor, the decompressor will fail to restore the compressed headers accurately since the context state at the decompressor is different than the state at the compressor. Decompressor fails not when a packet is lost, but when the next compressed packet arrives. If the next packet happens to include the same context update as in the lost packet, the context at the decompressor may be updated successfully and decompression may continue uninterrupted. If the lost packet included an update to a delta field such as the delta RTP timestamp, the next packet can't compensate for the loss since the update of a delta value is relative to the previous packet which was lost. But if the update is for an absolute value such as the full RTP timestamp or the RTP payload type, this update can be repeated in the next packet independently of the lost packet.

A "headers checksum" is inserted by the compressor and removed by the decompressor when the UDP checksum is not present so that validation of the decompressed headers is still possible. This allows the decompressor to verify that context sync has not been lost after a packet loss. Enhanced CRTP achieves robust operation by sending changes multiple times to keep the compressor and decompressor in sync. This method is characterized by a number "N" that represents the quality of the link between the hosts.

The Lightweight User Datagram Protocol (UDPLite), which is similar to the User Datagram Protocol (UDP) is propound [7]. It can serve to applications in error-prone network environments that prefer to have partially damaged payloads delivered rather than discarded. If this feature is not used, UDP-Lite is identical to UDP.

The Robust Header Compression (ROHC) protocol provides an efficient, flexible, and future-proof header compression concept. It is designed to operate efficiently and robustly over various link technologies with different characteristics. To improve and simplify the ROHC specifications, the new RFC explicitly defines the ROHC framework and the profile for uncompressed separately [8]. More specifically, the definition of the framework does not modify or update the definition of the framework specified by RFC 3095.

An updated version was defined for compression of RTP/UDP/IP, UDP/IP, IP and ESP/IP (Encapsulating Security Payload) headers [9]. Additional profiles for compression of IP headers, and UDP-Lite headers were later specified to complete the initial set of ROHC profiles for each of the above mentioned profiles, and the definitions depend on the ROHC framework as found in RFC 4995. Instead of compressing all RTP or all TCP packets that are going through network, it is possible to make RTP header compression to compress only those packets that belong to a class called "voice."

In [10] is proposed available method for the determination of Variable Sliding Window *(VSW)*, using a time interval, called Memory timeout interval during which the timeout values are registered. Indicating the actual value with *VSW (n)* and the previous value with *VSW (n-1)*. At current time, Variable Sliding Window dimension is expressed as:

$$VSW(n) = VSW_{\min}(n) + VSW_{BER}(n) \quad (1)$$

$VSW_{MIN}(n)$ is defined as follows:

$$VSW_{MIN}(n) = \left\lceil \beta . \frac{T.\left(\frac{MaxPkSize}{N}\right)}{MeanArrivalTime} \right\rceil \quad (2)$$

where MaxPkSize is the size, expressed in bit, of the biggest compressed or not compressed IP packet transmitted on Wireless channel; N bits are the maximum payload of a baseband packet; T ms is twice Wireless slot-time; MeanArrivalTime is the mean time gap between packets in the same application flow; β is a constant bigger than one.

V. OPNET IMPLEMENTATION

For investigation into VoIP traffic using ROHC we will use software OPNET [11], [12]. In this section we describe the basic steps in creating a model for the selected topology, as seen on Fig.1.

Select objects Application Config, Profile Config and QoS Attribute Config. Take two subnet models and ip32_cloud, which will simulate various traffic in our setting. With subnet elements we open two new areas, called sublevels of the topography. They are shown on Fig.2. Insert mobile computers that we chose – respectively 17 in the subnet London and 19 for subnet – Varna. In subnets located into Progect Editor of program, we insert 36 number of wireless ethernet_wkstn. To loading the links in a network we include two ethernet_server - respectively for HTTP and FTP traffic.

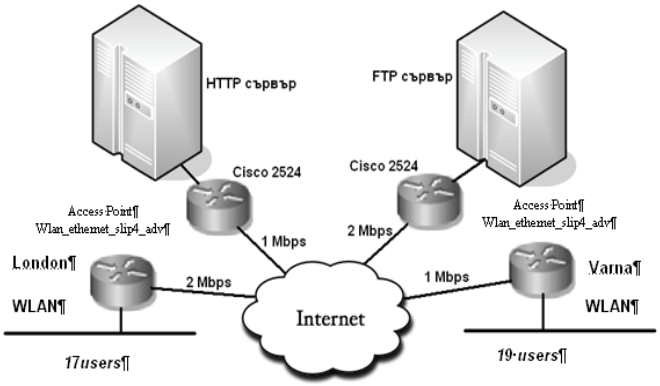


Fig. 1. Opnet project

The main components constitute our end-to-end system: the RTP/UDP/IP generator and sink, the RTP/UDP/IP header compressor/decompressor, and the transport network. The remaining sections are logically arranged in the organization seen in Fig. 1. This project can be regarded as two identical data paths, traveling in opposite directions.

Each data path consists of some traffic sources, each sending a unique compressed VoIP RTP/UDP/IP traffic stream to an aggregator. The aggregator combines the packet streams from the random number generators into a single point-to-point link, which is connected to the RTP/UDP/IP ROHC compressor/decompressor. The RTP/UDP/IP header compressor compresses the packet headers, which coming by line as necessary and sends each packet through the WLAN encapsulator. Sending the compressed packet, the WLAN encapsulator encapsulates the packet into an ethernet cell, and transmits the cell over a wireless channel to the terminal decapsulator through access point. The decapsulator removes the compressed packet from the ethernet cell and sends the packet to the RTP/UDP/IP header decompressor. The decompressor reconstructs the packets based upon its stored context state information.

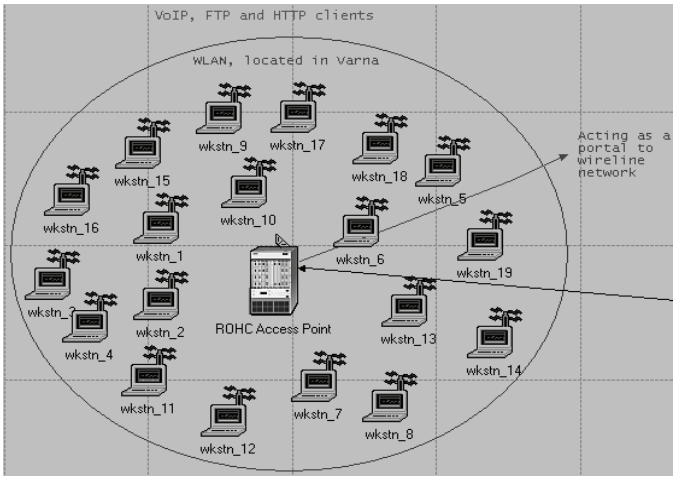


Fig. 2. WLAN subnetwork, located in Varna

The same data transfer process takes place in the opposite direction, using the counterparts of the same components used in the first direction of data transfer.

In the simulation we set up several streams of information such as voice traffic, FTP traffic and HTTP traffic. Then we create profiles, which can be set to already created applications to define the traffic generated for this profiles. Once the profiles have been created, we set up bi-directional compression of VoIP traffic from terminal to access point and contrariwise. We define the type of VoIP traffics generated by the application. For one of the stream we choose voice with parameters G.711, PCM Quality Speech - 64kbps, ToS - Interactive Voice, without RSVP parameters. For other flows we choose encoder scheme - G.723.1, G.728, G.729.

VI. SIMULATION RESULTS

Different main possible statistics are received, like packets end-to-end delay (sec) – Fig. 3 and traffic sent (bytes/sec), total delay of a packets from end to end (s), the speed with which information transmitted on the channel (bit/sec) and throughput of VoIP traffic through WLAN without HC (1) and with HC (2) – Fig.4. The analysis of current researches, in order to visualize the main results, some of them are presented in graphic form. The other parameter which will take into consideration is utilization of bandwidth (%) of VoIP traffic.

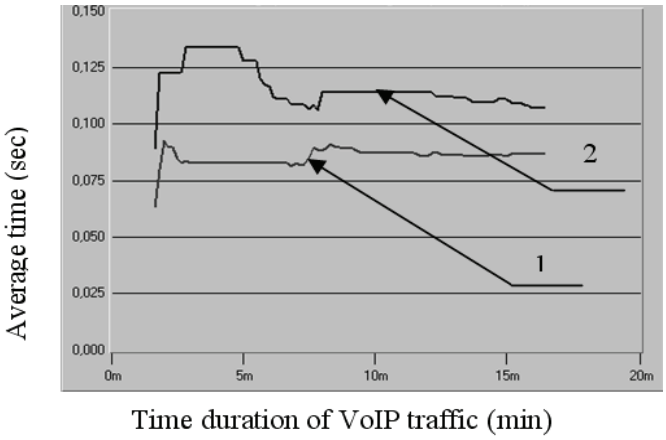


Fig. 3. Average packet end to end delay time (s) of VoIP traffic with ROHC (1) and without ROHC (2)

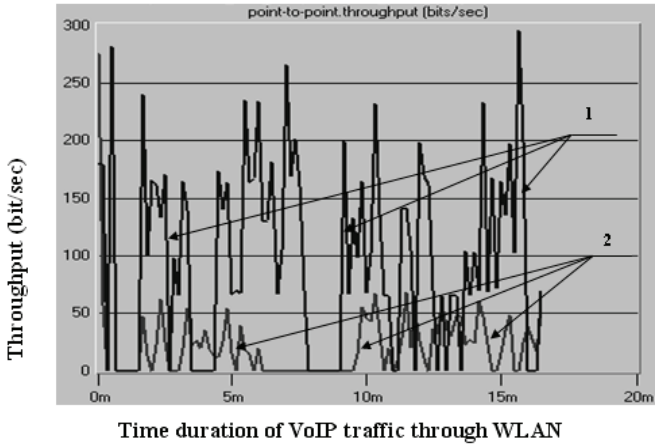


Fig. 4. Throughput of VoIP traffic through WLAN without HC (1) and with HC (2)

Table I shows the results of speed up of data transmission with HC, occupied bandwidth of VoIP traffic using priority (QoS), packet loss, reliability and Variable Sliding Window VSW_{MIN} *(n)* when is used ROHC and without ROHC.

TABLE I
ROHC PARAMETERS

ROHC Parameters	With ROHC	Without ROHC
Speed up of FTP date transmission	17 s	22 s
Occupied VoIP bandwidth	63,2%	74,5 %
Packet loss	4,1%	3,8 %
Reliability	10^{-2}	$10^{-2} – 10^{-3}$
$VSW_{MIN}(n)$	41	47

The voice codecs generating a VoIP packet every 20 ms, while the UDP checksums were enabled. In this scenario with ROHC scheme, we are interested in two result aspects: packet loss and occupied bandwidth. The first column represents the header compression scheme's performance regarding bandwidth savings, while the second column evaluates the parameters without use of ROHC.

VII. CONCLUSIONS

In narrowband networks application of HC is reflected in improvements in response time caused by the smaller size of the packets. The results prove that the smaller size of the packets also reduces the possibility of errors. For VoIP transmission quality is increased while using less bandwidth. HC improves quality of transmission and speeds up the network. It is a result of saving in bandwidth, reduces packet loss, improves interactive response time and reduces the cost of infrastructure due to including more users per channel.

HC is hop-to-hop process and is not applicable to end to end connection. For each node in the IP network, it is necessary to uncompress the packet to be able to perform operations such as routing, ensuring quality of service (QoS), etc. ROHC is best applied to specific links in the network, characterized by relatively narrowband, high bit error rate and more delay.

REFERENCES

[1] EFFNET AB, An introduction to IP header compression, White paper, February, 2004.
[2] C. Bormann, C. Burmeister etc., RObust Header Compression (ROHC), Framework and four profiles: RTP, UDP, ESP, and uncompressed, RFC 3095, July, 2001.
[3] V. Jacobson, "Compressing TCP/IP headers for low-speed serial links", RFC 1144, February 1990.
[4] M. Degermark, B. Nordgren, and S. Pink, "IP Header Compression", RFC 2507, February, 1999.
[5] S. Casner, V. Jacobson, Compressing IP/UDP/RTP Headers for Low-Speed Serial Links, RFC 2508, February, 1999.
[6] T. Koren, S. Casner, J. Geevarghese, B. Thompson, and P. Ruddy, "Enhanced Compressed RTP (CRTP) for Links with High Delay, Packet Loss and Reordering", RFC 3545, July 2003.
[7] L-A. Larzon, S. Pink etc., The Lightweight User Datagram Protocol (UDP-Lite), RFC 3828 July 2004.
[8] L-E. Jonsson, G. Pelletier, K. Sandlund, The RObust Header Compression (ROHC) Framework, RFC: 4997, July 2007
[9] G. Pelletier,K. Sandlund, RObust Header Compression Version 2 (ROHCv2):Profiles for RTP, UDP, IP, ESP and UDP-Lite, RFC 5225, April 2008
[10] L. Marzegalli, M. Masa, M. Vitiello, Adaptive RTP/UDP/IP Header Compression for VoIP over Bluetooth, ICC – 2009.
[11] URL:http://www.opnet.com.
[12] N. Rendevski, A. Tentov, E. Vlahu-Gjorgievska, B.Ristevski, FPGA Synthesis of Opensparc Processor's Cores, Network Implementations and Performances, 29th International Conference on Organizational Science Development, 2010, Portoroz, Slovenia.

Semi-Automatic Block System with Fiber Optic Chanel Data Transmission

Nikolay Nikolov [1], Dimitar Goranov [2] and Emiliya Dimitrova [3]

Abstract – **A device developed by DISSY Company is reviewed in the paper. It allows SAB regimes implementation by securing trains operation between two neighboring stations using an optic transmission medium.**

Keywords – **Telemechanic system, Semi-automatic blocking Stepanov, linean line circuit, local circuit, optic connection channel.**

I. INTRODUCTION

The semi-automatic blocking (SAB) known under the name Stepanov is a telemechanic system providing safe trains operation between two neighboring stations with a broad application in the railway system of our country. When SAB is installed, a train is allowed to enter an open line only at a permissible indicator of the starting signal given by the operation manager on duty, provided that there is not a train in the open line and a blocking permission is received from the reception station. It is closed automatically by the train leaving the station. After the opening of any exit signals in the departure station, all the rest (signals that could permit entering into the open line from one side, as well as from the other side) are closed, while the receiving station sends a blocking signal for train arrival to the departure station. The blocking signal confirming train arrival is given by the operation manager on duty, but only after an objective check of the actual arrival. In case of emergency the operational managers on duty use stamp buttons and take entirely or partially care of safety conditions keeping.

Information between two neighboring stations is exchanged through a polar compact direct current linear circuit with combined linear receiver in both stations. The linear circuit is two-conductive – cable (mainly in electrified sections) or an air circuit. One-conductive line is used in some scheme decisions and "earth" is used for a second conduit. The most common case is simulation of two-conductive telephone line between stations through differential transformers for the operation of the direct conductor and "earth" is the reverse conductor. A new regime for operation of the linear circuit shall be created for transmission of a subsequent command.

[1]Nikolay Nikolov is with „Metropolitan" EAD, 121 "Knyaz Boris I" Str., 1000 Sofia, Bulgaria, E-mail: nikolovna@gmail.com

[2]Dimitar Goranov is with DISSY LTD, 87-89 Pernik Str., 1309 Sofia, Bulgaria, E-mail: dgoranov@abv.bg

[3]Emilya Dimitrova is with Todor Kableshkov University of Transport, 158 Geo Milev Str., Sofia, Bulgaria,

E-mail: edimitrova@bitex.bg

The more frequent burglaries of copper cables in open lines make impossible SAB application. In this case the train operation is managed by telephones, but lots of disadvantages arise from this method.

The availability of optic fiber cable (NRIC property) between two neighboring stations and the fact that this cable is not a subject of burglaries allow SAB regimes implementation by securing trains operation between two neighboring stations using an optic transmission medium.

In accordance with NRIC Terms of Reference, DISSY Ltd. Company developed and introduced a device under the name DISIM-SAB [1], allowing classical SAB - Stepanov regimes implementation. The connection between its parts in both neighboring stations is fulfilled by a fiber optic cable. The device does not change the functional sequence of work with SAB.

DISIM-SAB got permission for commissioning by the Ministry of Transport, Railway Administration Executive Agency № BG-01-CCS-S and since 19.12.2007 it has been put in regular operation along Alexandar Dimitrov – Zemen section. At the end of 2008, due to introduction of this device, SAB was restored also along Radomir - Alexandar Dimitrov, Zemen – Razhdavitza and Razhdavitza – Kopilovtzi sections.

II. CHANGES IN CLASSICAL SAB – STEPANOV SCHEMES

Change in the existing linear circuit scheme has been done for interaction between SAB semi- parts in both neighboring stations by a fiber optic cable, as follows:

- The necessity of transformer drops out;
- The linear relay scheme is changed in such a way that it could be managed by code combinations in "Received Consent" – RC regime and „Track Arrival" – TA;
- Possibility for setting the relay into motion is foreseen in the scheme of TDR relay, when a code combination "Received Departure" – RD is received from the neighboring station;
- GAR relay scheme is changed in such a way that it is set into motion when „Track Arrival" – TA command is received, after a verification of the conditions necessary for regime fulfillment.

The scheme of local circuit relays connection is unchanged. Repeaters of some relays are supplemented.

The structural scheme of modified SAB-Stepanov is shown on Fig. 1, where:

1-given consent, 11-received consent,
2-given departure, 22-received departure,
3-track arrival.

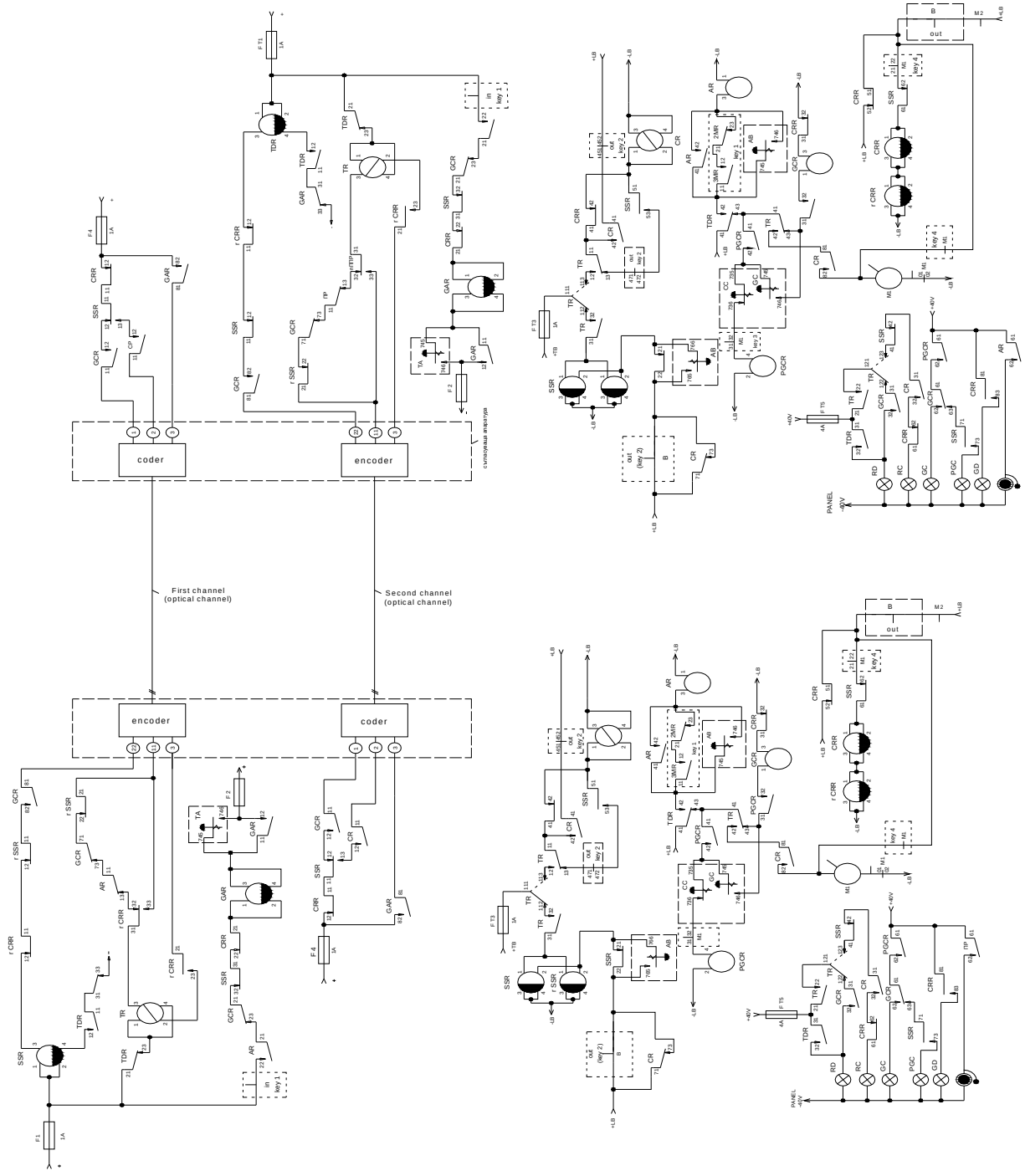


Fig. 1. Structural scheme of modified SAB-Stepanov

III. DISIM-SAB WORKING PRINCIPLE AND STRUCTURAL SCHEME

The signals arriving for connection into the optic channel are encoded, using 8-digit Heming code with Heming distance 4 between the different code combinations, i.e. a simultaneous transition 0 → 1 or 1 → 0 is necessary for the transformation of one code combination into another. The allowed code combinations responding to the condition for Heming distance 4 are shown in Table I.

TABLE I
ALLOWED CODE COMBINATIONS

№	Possible combinations	Function
1.	0000 0000	Standard status of decoding relays
2.	0001 0111	**GC (RC)**
3.	0010 1011	**GD (RD)**
4.	0011 1100	**TA**
5.	0100 1011	not applied
6.	0101 1010	not applied
7.	0111 0001	not applied
8.	1000 1110	not applied
9.	1010 0101	not applied
10.	1011 0010	not applied
11.	1100 0011	not applied
12.	1101 0100	not applied
13.	1110 1000	not applied

Code combinations (2) ÷ (4) from Table I are used for "given consent" (respectively "received consent") and "given departure" (respectively "received departure") and "track arrival" regimes implementation.

DISIM-SAB consists of:

- Coder;
- Encoder;
- Operating processor controller.

The block scheme of the device is shown on Fig. 2.

Contacts' reading from SAB relay apparatuses is done through a coder (diode matrix), that codes the entering analog signals into 8 digit code combinations (words).

The code combinations enter into the controller, which verifies their correctness, i.e. verifies the validity of the respective combination. If the result from the verification is positive, the respective key word is transmitted to the optic equipping of the channel for connection. If the code combination is invalid, it is not transmitted to the channel for connection.

Encoding of the accepted code combination is fulfilled with 8 receiving relays.

The encoder transforms coded signals for connection entering the channel into analog ones in a way acceptable for the interlocking relay apparatuses.

Encoding has to be completed (front and back contacts of all the 8 relays are included).

In case of a standard status, the encoder's relays perform the first code combination (1) in Table I.

When a coded signal arrives at the channel for connection, the processor controller sends the received signal to the encoder, if the following conditions are fulfilled:

- Triple subsequent full coincidence of the received code words information bits;
- Availability of continuous communication. The power supply of the encoder's relays will be stopped if the communication is interrupted for more than 100ms. The encoder's relays perform the first code combination (1) in Table I and doing so the transmission of an analog signal to SAB is interrupted.

Only a correctly set combination of the relay encoder is transferred to SAB set. The combination correctness is checked by a reverse reading of the contacts of the encoder's relays. The command is kept only if there is coincidence of the reverse reading and availability of continuous communication.

DISIM-SAB creates Events Protocol, i.e. it makes an archive in chronological order of the events arisen in DISIM-SAB. If it is necessary the Events Protocol could be downloaded using a portable computer and subsequently printed. When the computer memory is filled up, the record starts from the beginning, as the oldest records are firstly erased.

The available events for DISIM-SAB are shown in Table II.

A light-emitting diode indicating panel is foreseen for timely diagnostics of the different DISIM-SAB states. The panel is visible through the transparent cabin door.

DISIM-SAB works on two single mode optic fibers. Depending on the kind of the receiver-transmitter (Table III), the device works reliably in open lines with length of up to 35 km. Its working temperature interval is from - 5°C to +45°C.

The device allows a direct telephone connection of "local battery" type for implementation of "interstation connection". It is supplied by a station battery 24V (-10% +20%) with insulated terminals and works reliably in the presence of power supply pulsations up to 40%. When the supply voltage goes down to the lower limit, DISIM-SAB starts running in protective state and when it enters into the working zone, it proceeds to working state.

The device does not produce harmonious components and does not disrupt the normal work of the telecommunication and radio equipment.

DISIM-SAB operates under conditions responding to II climate group requirements in compliance with Bulgarian standart 17165-90. It keeps its working state and appearance under the following environmental parameters impact:

- Temperature of the ambient air: from 5 to 40°C;
- Air relative humidity at temperature of 28°C: from 5% to 85%;
- Atmospheric pressure: from 70 to 106 κPa;
- Air dustiness: not more than 0,75 mg/m3;

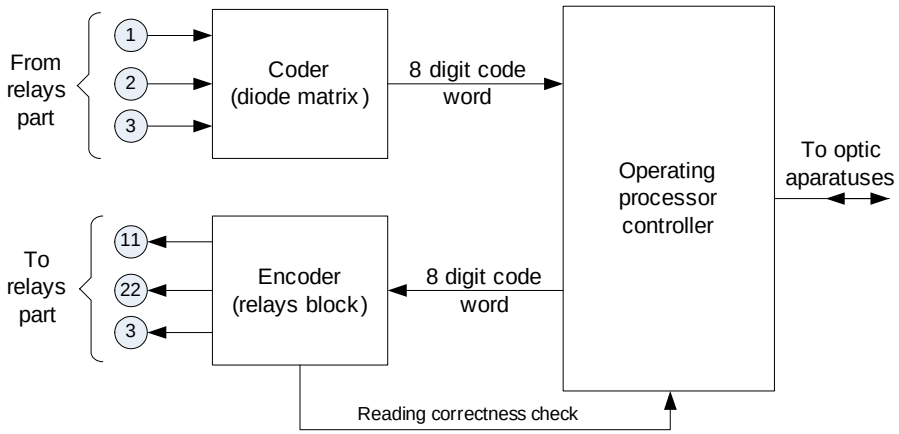


Fig. 2. Block scheme of DISIM-SAB

- Floor vibrations: with an amplitude not more than 0,1 mm, at frequency not more than from 10 to 25 Hz;
- Electrical component of disturbances electromagnetic field: not more than 0,3V/m;
- Absence of aggressive admixtures in the surrounding atmosphere.

TABLE II
EVENTS PROTOCOL

№	Event
1.	Power supply plugged in DISIM-SAB
2.	Fulfilled communication
3.	GC/ДС at the coder entrance
4.	GD/ДЗ at the coder entrance
5.	TA/ПП at the coder entrance
6.	RC/ПС at the encoder exit
7.	RD/ПЗ at the encoder exit
8.	TA/ПП at the encoder exit
9.	Business information
10.	Reset due to a system mistake
11.	Reset due to transfer mistake
12.	Reset due to acceptance mistake
13.	Protocol mistake – head of the telegram
14.	Protocol mistake – control amount of the telegram
15.	Invalid command of the encoder exit
16.	Invalid command of the coder entrance
17.	Reverse reading mistake
18.	U battery normal
19.	U battery low
20.	U battery high

TABLE III
TECHNICAL SPECIFICATIONS OF USED RECEIVER-TRANSMITTERS

Standard Compliance	IEEE 802.3z
Data Rate Gigabit	1.25Gbd
Data Rate Fiber Channel	1.0625Gbd
Media	SMF
Optical Receiver Sensitivity	-24dBm
Optical Transmission Power	0 ~5.0dBm
Power Budget	23.0dB
Center Wave Length	1550nm
Transmission Range	9/125um (up to 80km)
Connector	LC type

IV. CONCLUSION

The unusual operational conditions produce unusual technical decisions. This applies fully for the elaboration of semi-automatic block system with fiber optic channel data transmission. Along with modern technologies and elements usage, old and already rejected systems are returned and put into real operation. DISIM-SAB design, elaboration and introduction increased the product range of DISSY Company and strengthened the company's positions with innovative view and perspective developments.

To our great satisfaction, SAB with fiber optic channel data transmission has been working flawlessly from its installation for experimental operation until today and has positive references from NRIC employees working along Radomir – Kustendil section.

REFERENCES

[1] DISSY Ltd., Passport of a device for work on Semi-automatic blocking (SAB) through a fiber optic cable, Sofia, 2007.
[2] Paltekov, I. K. Automatic regulation of trains operation, Sofia, Todor Kableshkov University of Transport, 1985.
[3] Nedelchev, N. N., H. A. Hristov, I. H. Nenov, Autoblocking and autoregulation, Sofia, V.I.Lenin Technical University, 1980.

Estimation of Optical Link Length for High-Speed Aplications

Nataliya Varbanova[1], Krasen Angelov[2], Stanimir Sadinov[3]

Abstract – **In a fiber-optic link for high data rates, the system can be limited either by the losses (attenuation-limited transmission) or, assuming that the link is not limited by the source or detector speed, by the dispersion of the fiber (dispersion-limited transmission).**

Therefore a major task when designing optical links is the estimation of the optical link length in terms of various limiting factors.

Keywords – **optical link length, optical attenuation, optical dispersion.**

I. Introduction

Data transmission speed via optical link actually is not infinity. In a fiber-optic system at long distances or high data rates, the system can be limited either by the losses (attenuation-limited transmission) or, assuming that the link is not limited by the source or detector speed, by the dispersion of the fiber (dispersion-limited transmission) [1,4,9,11].

Estimation of maximum optical link length mainly depends on the following aspects:

- Source selection;
- Power budget;
- Dynamic range;
- Timing analysis (e.g. performance of the equipment);
- Attenuation-limited transmission length;
- Dispersion-limited transmission distance.

In optical links, dispersion is the phenomenon in witch the parameters of the medium are dependent on frequency of signals spreading through the medium or alternatively when the group velocity depends on the frequency [1,11].

Dispersion is sometimes called chromatic dispersion to emphasize its wavelength-dependent nature, or group-velocity dispersion to emphasize the role of the group velocity.

There are generally two sources of dispersion: material dispersion and waveguide dispersion [1,4,9,11]. Material dispersion comes from a frequency-dependent response of a material to waves. Waveguide dispersion occurs when the speed of a wave in a waveguide (such as an optical fiber) depends on its frequency for geometric reasons, independent of any frequency dependence of the materials from which it is constructed. In general, both types of dispersion may be present, although they are not strictly additive. Their combination leads to signal degradation in optical fibers for telecommunications, because the varying delay in arrival time between different components of a signal "smears out" the signal in time.

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

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

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

II. Components and System Requirements

Fig. 1 shows the three primary components in a fiber-optic link: an optical transmitter, a fiber-optic cable, and an optical receiver [1-3,5,10,12]. In the transmitter, the input signal modulates the light output from a semiconductor laser diode, which is then focused into a fiber-optic cable. This fiber carries the modulated optical signal to the receiver, which then reconverts the optical signal back to the original electrical RF signal.

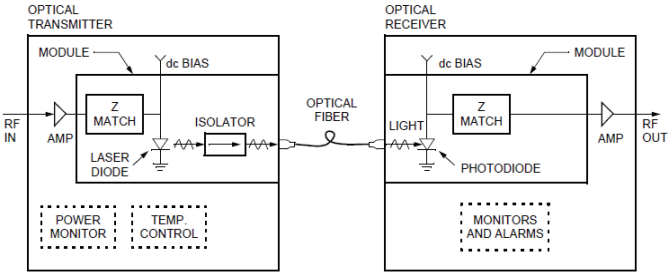


Fig. 1. Basic block diagram of the main components of the optical communication link

A. Source selection

The starting point for a link design is choosing the operating wavelength, the type of source (i.e., laser or LED), and the fiber type (single-mode or multi-mode fiber) [2,10,12]. In a link design, it is usually known the data rate required to meet the objectives. From this data rate and an estimate of the distance, it is chosen the wavelength, the type of source, and the fiber type.

Source data rate-distance performance limits are summarized in Table 1.

The choice of fiber type involves the decision to use either multi-mode or single-mode fiber [1,4]. And if the fiber is multi-mode, whether to use graded-index or step-index profiles. This choice is dependent on the allowable dispersion and the difficulty in coupling the optical power into the fiber.

TABLE I
SOURCE PERFORMANCE LIMITS

Source type	Short wavelength	Long wavelength
LED	< 150 Mbps.km	< 1,5 Gbps.km
Laser	< 2,5 Gbps.km	< 25 Gbps.km

If a LED is chosen, then the obvious choice of fiber is a multi-mode fiber because the coupling losses into a single-mode fiber are too severe. For a laser source, either a multi-mode or single-mode fiber can be used. The choice depends on the required data rate, as losses in both types of fiber can be made quite low [7].

B. Power budget

With a tentative choice of source, it is known the power P_T, available to be coupled into the fiber. If the receiver power P_R is necessary to achieve the required performance, then the ratio P_T/P_R is the amount of acceptable loss that can be incurred and still meet the specifications [7-10,12]. This can be expressed by

$$L_{losses,[dB]} + l_M = 10\log\left(\frac{P_T}{P_R}\right), \quad (1)$$

where l_M is the system margin.

The losses L_{losses} can be allocated in any desired approach by the system designer. Generally, the probable losses will be as follows:

- The source-to-fiber coupling losses l_T, *dB*;
- The connector insertion losses l_C and the splice insertion losses l_S;
- The fiber-to-receiver losses l_R (these losses is usually negligible);
- Additional losses l_A for device aging effects and future splicing requirements;
- Fiber attenuation losses αL (where α is the attenuation rate per kilometer of fiber and L is the fiber length).

Therefore Eq. (1) can then be expressed as:

$$10\log\left(\frac{P_T}{P_R}\right) = L_{losses,[dB]} + l_M = \alpha L + l_T + nl_S + l_R + l_A + l_M . \quad (2)$$

After solving Eq. (2) for the system margin l_M , the result will be:

$$l_M = P_{T,[dBm]} - P_{R,[dBm]} - \alpha L - l_T - nl_S - l_R - l_A . \quad (3)$$

A positive system margin ensures proper operation of the circuit; a negative value indicates that insufficient power will reach the detector to achieve the required *BER* [6].

C. Dynamic range

Using a "best case/worst case" approach it can be examined whether the link has sufficient dynamic range [4,5,10,12].

In Eq. (2) l_M can be written as follows:

$$l_M = l_{TR} - l_{SYSTEM} , \quad (4)$$

where l_{TR} is the ratio of the transmitter power to the required receiver power, expressed in dB, a l_{SYSTEM} is the total sum of the all system losses, given by:

$$l_{SYSTEM} = \alpha L + l_T + nl_S + l_R + l_A . \quad (5)$$

The dynamic range of the system is found by calculating the maximum and the minimum system margins. The two computations are summarized by

$$l_{M,\max} = l_{TR,\max} - l_{SYSTEM,\min} , \quad (6)$$

$$l_{M,\min} = l_{TR,\min} - l_{SYSTEM,\max} .$$

The system dynamic range $DR_{[dB]}$ is given by the difference in the values from Eq. (6):

$$DR_{[dB]} = l_{M,\max} - l_{M,\min} . \quad (7)$$

The optical receiver must have an equivalent dynamic range in order for the system to work properly [6]. The basic concern is to keep the power at the receiver above the minimum detectable power of the detector $P_{R,\ min}$ and below the maximum-rated power of the detector $P_{R,\ max}$. From Eqs. (2) and (3) the received power is deduced to be

$$P_{R[dBm]} = P_{T[dBm]} - l_{SYSTEM} . \quad (8)$$

Therefore the two boundary conditions will be respectively [9]:

- Maximum power output combined with minimum fiber attenuation;
- Minimum power output combined with maximum fiber attenuation.

D. Timing analysis

Rise time is a parameter of fundamental importance in high speed transmissions, since it is a measure of the ability of a circuit or system to respond to fast input signals. Rise time refers to the time required for a signal to change from a specified low value to a specified high value.

Rise time of a fiber-optic system Δt_{sys} is given by [1,10,12]:

$$\Delta t_{sys} = \left[\sum_{i=1}^{N} \Delta t_i^2\right]^{\frac{1}{2}} = \left[\left(\Delta t_S\right)^2 + \left(\Delta t_R\right)^2 + \left(\Delta t_{mat}\right)^2 + \left(\Delta t_{\mathrm{mod}\,al}\right)^2\right]^{1/2} = , \quad (9)$$

where Δt_i is the rise time of each component in the system. The four components of the system that can contribute to the system rise time are as follows:

- Rise time of the transmitting source Δt_S;
- Rise time of the receiver Δt_R;
- Material-dispersion delay time of the fiber link Δt_{mat};
- Modal-dispersion delay time of the fiber link Δt_{modal}.

Modal-dispersion delay time of the fiber link Δt_{mat} is given by [1,10,12]:

$$\Delta t_{mat} = -\frac{L}{c}.\frac{\Delta\lambda}{\lambda}.\left(\frac{\lambda^2 d^2 n}{d\lambda^2}\right). \quad (10)$$

For a step-index fiber with length L, the modal-dispersion delay is given by

$$\Delta t_{\text{modal}} = \frac{L(n_1 - n_2)}{c}, \quad (11)$$

and for a parabolic-index fiber the delay is estimated as:

$$\Delta t_{\text{modal}} = \frac{L}{c}.\frac{NA^2(0)}{8n_1^2}, \quad (12)$$

where NA is a numerical aperture.

If the system rise time is calculated, than using this value, it can be calculated the data rate that the system can support as [1,10,12]:

– for NRZ coding: $B_R \le \frac{0{,}7}{\Delta t_{sys}}$; (13)

– for RZ coding: $B_R \le \frac{0{,}35}{\Delta t_{sys}}$. (14)

The modal dispersion delay time depends linearly proportional to the distance. The modal dispersion contribution is small for short distances. To reduce the material dispersion, inspection of Eq. (10) reveals that $\Delta\lambda$ should be reduced. There are two methods by which this can be achieved:

1) To use an LED with a longer wavelength (while keeping $\Delta\lambda$ constant);
2) To use a laser source with its reduced value of $\Delta\lambda$.

E. Estimation of the maximal optical link length

To estimate the maximum optical link length in presence of dispersion and attenuation limitations in the fiber, as a function of the transmission data-rate speed in the fiber are derived the following relations [10,12]:

• Material Dispersion-Limited Transmission Length:

$$L_{\max} = \frac{0{,}35c\lambda}{B_R\Delta\lambda\left(\frac{\lambda^2 d^2 n}{d\lambda^2}\right)}. \quad (15)$$

• Modal Dispersion-Limited Transmission Length:

$$L_{\max} = \frac{2{,}8.c.n_1^2}{[NA(0)]^2 B_R} = \frac{1{,}4.c.n_1^2}{n_1^2\Delta B_R} = \frac{1{,}4.c}{\Delta B_R}. \quad (16)$$

• Attenuation-Limited Transmission Length:

$$L_{\max\,(att)} = \frac{P_{T[dBm]} - P_{R[dBm]}}{\alpha_{fiber}} = \frac{P_{T[dBm]} - (-65{,}0 + 20\log_{10}(B_R))}{\alpha_{fiber}}. \quad (17)$$

III. RESULTS

As an example, the estimation of a maximal length of high-speed optical link for short distances can be considered, as shown on Fig. 2. The link parameters are as follows [13-15]:

– Optical source type: LED;
– Optical source power: P_T = 2 *mW*;
– Optical wavelength: 850 *nm*;
– Multi-mode fiber with graded-index (g = 2);
– Data rate: < 500 *Mbps*;
– *BER* = 1.10^{-9};
– Optical receiver with pin-photodiode with sensitivity – 40 *dBm*.

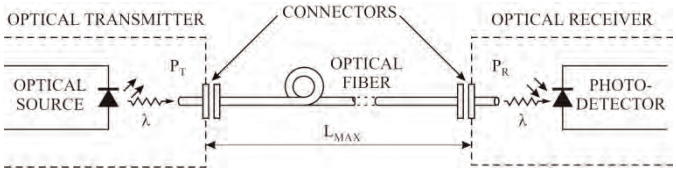


Fig. 2. Estimation of optical link length for high-speed aplications

Graphic appearance of the resulting dependencies of the maximum distance of transmission as a function of transmission speed in various limiting factors was carried out based on Eqs. (15) to (17) and is shown on Fig. 3.

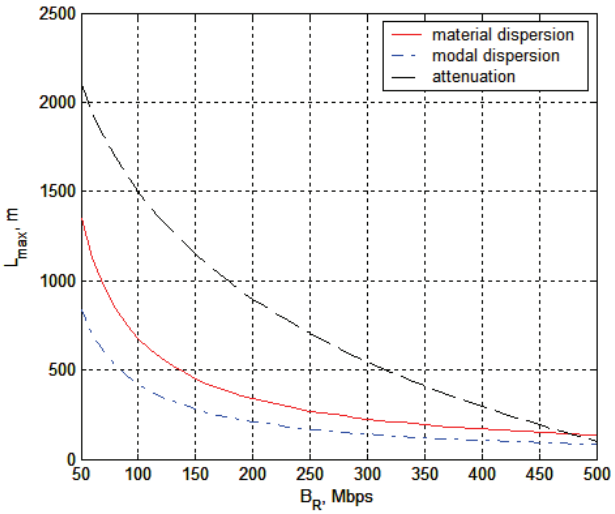


Fig. 3. Comparative effect of limiting factors on the length of the optical link

IV. CONCLUSION

Optical link design must meet both the link power budget and the system rise time analysis. In the link power budget analysis one first determines the power margin between the optical transmitter output and the minimum receiver sensitivity needed to establish a specified *BER*. Once the link power budget has been established, the designer makes a system rise time analysis to ensure that the dispersion limit of the link has not been exceeded.

In designing an optical-fiber system, one should take into account that the individual component parameters can vary considerably.

From the obtained results, the following important conclusions can be made:

- At lower data rates the main restriction on the length of the line comes from the dispersion; the modal dispersion has slightly more influence than the material dispersion;
- At higher data rates the maximal link length fell sharply as the fiber attenuation effect has a noticeable influence.

From the obtained results it is clear that taking into account the dispersion indicates a significant influence in determining the exact maximal optical link length.

It should be noticed that if the length of the line is greater than the fiber optic cable length, the optical splices must be taken into account in total attenuation calculations.

ACKNOWLEDGEMENT

This paper has been sponsored under the auspices of the "Increasing Efficiency and Quality of Service in Broadband Hybrid Cable Television Networks" project – a part of the University Center for Research and Technology (UTzNIT) at the Technical University of Gabrovo, contract E904/2009.

REFERENCES

[1] Е. Фердинандов, Ц. Мицев, В. Кърдев. Световодни комуникационни системи. Част 1 Преносна среда, Сиела, София, 2002.
[2] Е. Фердинандов, Ц. Мицев, В. Кърдев. Световодни комуникационни системи. Част 2 Предавателна част, Сиела, София, 2002.
[3] Е. Фердинандов, Ц. Мицев, В. Кърдев. Световодни комуникационни системи. Част 3 Приемна част, Сиела, София, 2003.
[4] Ст. Рабов, Л. Христов, Оптични комуникации, Нови знания, София, 2002.
[5] C. Gee, T. Chen, N. Bar-Chaim, I. Ury, K. Lau, "Designing High Dynamic Range Fiber-optic Links: A Comparison Between Directly-Modulated Fabry-Perot and Distributed-Feedback Laser Diodes," Microwave Journal, May 1993.
[6] J. Barry, E. Lee, "Performance of Coherent Optical Receivers", Proceedings of IEEE, Vol. 78, No. 8, August 1990.
[7] K. Angelov, K. Koitchev, N. Varbanova, Estimating Losses from Transient and Intersymbol Distortions in Hybrid Fiber-Coaxial Television Network, ICEST 2009, Proc. of Papers Vol. 1, pp.113-116, Sofia, Bulgaria, 2009.
[8] K. Angelov, K. Koitchev, S. Sadinov, An Investigation of Noise Influences in Optical Transmitters and Receivers in Cable TV Networks, ICEST 2006, Proceedings of Papers, pp.102-105, Sofia, Bulgaria, 2006.
[9] K. Koitchev K., K. Angelov, S. Sadinov, Determining Bit Error Rate in Digital Optical Transmission Network Using the Q-Factor, ICEST 2010, Proc. of Papers, Vol. 1, pp.53-56, ISBN: 978-9989-786-57-0, Ohrid, Macedonia, 2010.
[10] M. Fujita, S. Ramesh, R. Tatro, Fiber Optic Communication Link Design, EEE 165, CSU Sacramento, December 1, 2003.
[11] R. Freeman, Fiber-Optic Systems for Telecommunications, John Wiley & Sons, New York, 2002.
[12] RF and Microwave Fiber-Optic design Guide, Application note AP01-006OPTO, Agere Systems, April 2001.
[13] http://www.emcore.com/fiber_optics - EMCORE Corporation
[14] http://www.ortel.net - ORTEL Communications Ltd.
[15] http://www.tuolima.com/optical-cable-series.html - Tuolima.

Attractive Ways Forward to Maximise Capabilities of the FD-BPM Technique

Dušan Ž. Djurdjević[1]

Abstract – **Attractive ways forward to maximise capabilities of FD-BPM technique in photonics and nano-photonics design are presented and discussed. Novel improved FD schemes and formulas, the Alternating-Direction Implicit scheme, the explicit DuFort-Frankel scheme, the complex Jacobi iterative method, the efficient three-dimensional wide-angle beam propagation methods, the use of preconditioners allowing the solution of 3D problems of interest in reasonable computer runtimes – are just some of the few novel approaches highlighted.**

Keywords – **Beam propagation method, Finite-difference, Photonics, Optoelectronics, Numerical simulations.**

I. Introduction

Beam propagation methods (BPM) stand as a standard and computationally efficient design tools used in integrated photonics and optoelectronics in the last decades. BPM is an approach for numerical solving of the paraxial approximation of an exact vector Helmholtz's equation (also known as the Fresnel's equation). Although BPM can be formulated in time domain as well, frequency domain BPM techniques are still dominant in photonics analysis allowing suitable results with low run-time and memory computer costs.

The finite-difference beam propagation method (FD-BPM) is certainly the most popular BPM algorithm [1-3]. Since its first formulation in 1990 [4], the FD-BPM has undergone significant improvements, particularly during the last decade [5-7].

The main feature of the original FD-BPM, the paraxial approximation, at the same time presents the crucial limitation of the algorithm. A remedy was found in the early 1990s when the wide-angle (WA) BPM algorithm using Padé series expansion of the square root operator were introduced [8]. So far, several WA-BPM algorithms have been suggested allowing significant improvements of the computational efficiency of the standard paraxial BPM technique.

The FD-BPM is usually implemented in a rectangular co-ordinate system and accordingly the accuracy of the method is affected by inevitable so-called staircase approximation. Namely, if the structure under analysis contains oblique or curved interfaces or when the structure is changing in the direction of the propagation, the dielectric boundaries are modelled with error causing serious problems and certain restrictions of the method. A remedy was found by using improved FD formulas, or by using co-ordinate systems, which exactly describe the geometry of the photonic device studied. Several, usually the non-orthogonal, co-ordinate systems were recently proposed as well as novel very efficient forms of improved FD formulas.

The basic drawback of the three-dimensional (3D) implicit WA-BPM is its huge memory consumption and consequently lengthy computer runtimes. The way out has been sought during the last decade in the development of unconditionally stable Altering-Direction Implicit (ADI) schemes. ADI algorithms provide non-iterative solution and require less memory and computational time. One further attractive possibility is the use of the fast and unconditionally stable explicit BPM algorithms like the DuFort-Frankel (DFF) variant. The Spectral Collocation Method (SCM) is recently proposed to minimize memory storage and to offer highly accurate results [9]. The Fourier cosine BPM algorithm, based on simple and time efficient matrix calculations [10], deserves a particular attention.

Iterative FD-BPM schemes designed to achieve higher accuracy in numerical simulations (fine FD meshes for modelling complex geometries, the use of higher order Padé approximation in the WA-BPM) tend to be time very consuming and often instable. A novel complex Jacobi iterative algorithm [11] and construction of suitable preconditioners can substantially improve the convergence and minimize computer runtime involved in simulations.

Usage of the FD-BPM is not limited only for applications and devices in conventional integrated photonics; there are examples in the very recent literature about its possible application in the design of modern photonic devices, such as photonic crystals fibers, plasmonics, integrated optical memories, and other components in nano-photonics and bio-photonics.

This paper reviews the most recent advances in the field of FD-BPM implementation. Highlighted improvements are still attractive and active areas of research. Extensive bibliography follows the review presented in the paper.

Presented brief summary does not cover all advanced approaches proposed in the literature during the last decade. However, many of referred approaches have already significantly impacted the FD-BPM CAD manufacturing in photonics and optoelectronics.

II. Wide-Angle FD-BPM Propagation

The standard paraxial FD-BPM method limits the field simulations to paraxial beams along or close to the z axis. The WA-BPM schemes use Padé approximation of neglected second-order derivative with respect to the z in the wave equation, as it is assumed in the original BPM formulation. With the WA-BPM algorithms the field can be propagated through tilted and curved waveguide structures and circuits

[1]Dušan Ž. Djurdjević is with the Faculty of Technical Sciences, Knjaza Miloša 7, 38220 Kosovska Mitrovica, Serbia, E-mail: dusan.djurdjevic@pr.ac.rs

without loss of accuracy. The WA-BPM approach offers much more realistic results of the lightwave propagation. Unfortunately, the serious drawback of this approach is the increase of the matrix bandwidth, especially when higher order Padé operator is used, causing problems with available computational resources.

In order to improve the efficiency of the WA-BPM, the algorithm is combined with various multistep methods [12,13] and ADI methods [14-16]. The implementation of various improvements of the WA-BPM, including different meshes, non-standard co-ordinate systems [17] and multistep methods is perhaps the most attractive area in FD-BPM research.

III. Structure-Related FD-BPM propagation

The co-ordinate transformation approaches reformulate the BPM in non-orthogonal, so-called structure related (SR) co-ordinate systems [18]. Successful approach to eliminate non-physical scattering due to the staircasing effect in FD discretization of oblique dielectric interfaces in rectangular co-ordinate system is the use of the co-ordinate transformation methods, such as SR FD-BPM [18]. SR co-ordinate systems, such as tapered, oblique, bi-oblique co-ordinate systems, naturally follow the local geometry of the structure under analysis. The BPM Helmholtz's equation can be rewritten and numerically solved in any orthogonal or non-orthogonal transverse co-ordinate system.

SR FD-BPM algorithm allows simulations with noticeably reduced numerical noise and shortened simulation time. The non-orthogonal co-ordinate FD-BPM has been applied to the analysis of structures with oblique, bi-oblique, tapered, and tapered-oblique cross-sections in the transverse plane. Recently, the oblique FD-BPM has been proposed based on the fast explicit DFF algorithm [19].

IV. Improved FD formulas

Implementations of the FD-BPM for structures in a rectangular co-ordinate system are characterized by low-order truncation errors, e.g. standard difference equations in two dimensions in homogeneous regions are second-order accurate, $n = 2$, or $O(h^2)$, where h is the FD mesh size. Near the step-index dielectric interfaces, accuracy usually drops to $n - 1$, and near dielectric corner points difference equations are $(n - 2)$th-order accurate, resulting with $(n - 1)$th-order of accuracy of the modal index and modal electromagnetic field.

The starting point in improving FD discretization procedure was Stern's work [20] where the concept of a semi-vectorial mode has been introduced, resulting in $O(h^0)$ truncation error. Vassallo [21] proposed an improved three-point FD formulation for the semi-vectorial case providing $O(h)$ accuracy. Yamauchi et al. [22] improved Vassallo's approach to give $O(h^2)$ accuracy regardless of interface position. Chiang et al. [23] generalized Vassallo's and Yamauchi's approach to full-vectorial case to give $O(h^2)$ accuracy for oblique, even curved step-index boundaries. Hadley [24,25] derived highly accurate FD formulas, assuming TE polarization, with truncation error in the uniform region $O(h^4)$ to $O(h^6)$ depending on the type of grid employed, and up to the $O(h^5)$ near dielectric interfaces under certain grid-interface conditions. In [24,25] Hadley utilized 2D solutions of the Helmholtz's equation in cylindrical co-ordinates. This approach resulted in the tremendous increase in accuracy, however with increase in algebraic and numerical efforts in formulas derivation and implementation. Three distinct cases of uniform regions, dielectric interfaces and dielectric corners are handled separately and these derivations are finally incorporated into a TE mode waveguide modelling tool. Although Hadley's FD formulas are rather tedious to derive and implement, they have been incorporated in the improved accuracy eigenmode solvers for benchmark purposes in some waveguide simulations.

A novel technique for obtaining the truncation error with $O(h^{2N})$ accuracy (where N is number of sampled FD points) is recently proposed [26] for the 2D case. The extension to 3D cases is expected, promising benefits with tremendously less computation time and memory.

V. Improved FD discretisation

FD-BPM modelling of optical and photonic crystal fibers with non-cylindrical cross-sections and complex geometries, including nonlinearity effects involved, have been successfully solved by using FD dicretisation with triangular-mesh [27]. Nonlinear contributions to the index of refraction, due to high-power regimes of optical systems, are successfully treated within the algorithm. Curved dielectric boundaries of any shape can be accurately approximated with irregular deformable triangular FD-grid, although the derivation of accomplished FD formulas is more sophisticated and therefore more complicated.

The main disadvantage of the standard FD discretisation approach is a need to define a line-structured grid of points. Non-standard FD approaches, such as Generalized Finite-Difference Method (GFDM) [28], relax the grid requirements. By using the radial or polynomial basis functions and a moving least-squares scheme, the FD interpolation formulas can be constructed on localized sets of points to enable dealing with complex geometries.

Promising approach has been reported in [29], where the generalised two-dimensional full-vectorial FD approach has been used for the electromagnetic field discretisation near dielectric interfaces within rectangular grid featuring in $O(h^4)$, and higher, truncation error.

The irregular generalized FD schemes are still to come in the use in the whole scope of the FD-BPM applications.

VI. Improved iterative FD-BPM algorithms

The implicit FD-BPM methods (like CN – Crank-Nicolson method) are known as unconditionally stable FD-BPM algorithms. For fine FD meshes those algorithms require often time lengthy usually iterative procedures for matrix inversions. This is particularly problem when the propagation matrix is not sparse enough as in the original FD-BPM

applications. The WA-BPM and the reflective FD-BPM algorithms are the typical examples. In those cases every step-forward in cutting the computer runtimes is welcomed. A new complex Jacobi iterative (CJI) method, proposed by Hadley [8], is one of the most promising and competitive approach. The CJI method has been successfully applied to the 3D WA-BPM simulations [12,13,30] enabling a development of higher order 3D Padé approximant–based algorithms within modest runtimes and memory requirements.

Another way forward to improve the iterative FD-BPM algorithms is to speed-up the convergence rate by constructing and applying a suitable preconditioner. Usually, this is the typical linear algebra problem, where the CJI method [11] can serves as a preferable approach for obtaining precondition parameters for optimum algorithm convergence. An efficient idea is to use a preconditioner based on paraxial approximation [31].

VII. NON-ITERATIVE FD-BPM ALGORITHMS

The standard FD-BPM technique employs the CN scheme which is unconditionally stable, however, as an implicit procedure, has a drawback because it uses iterative matrix solver in every propagation step and thus requires huge computational resources. ADI-FD-BPM makes use of the highly efficient non-iterative Thomas algorithm (the direct tridiagonal matrix solver) by splitting the FD operator in two one-dimensional terms – i.e. two FD equations.

ADI schemes have been successfully applied to the WA-BPM algorithm to enhance the efficiency of the WA schemes [14,15]. Further, the Hoekstra's scheme has been utilized with the ADI and WA algorithm [32]. Non-iterative Local One-Dimensional (LOD) schemes were recently introduced for 3D FD-BPM [33].

The attractive alternative to implicit (iterative or non-iterative) schemes is sought within the use of the three-level explicit DuFort-Frankel (DFF) algorithm [19,34]. The DFF algorithm is the rare example of the explicit BPM procedure being unconditionally stable. The DFF approach does not need matrix solver; the associate computer code can be parallelized easily and very efficiently onto distributed memory parallel computers. Besides of these highly competitive benefits, the DFF scheme has some serious inherent disadvantages (or weakness). First of all, this is the appearance of the spurious mode solutions (the FDD algorithm is empirically constructed). To avoid and suppress spurious (fake) modes, the propagation step has to be reduced, or obtained solution has to be filtered. Furthermore, the FDD-BPM approach is still limited to paraxial and semi-vectorial cases. Therefore, the trade between unprecedented simplicity and efficiency of the DFF-BPM method and serious drawbacks on the other hand will certainly continue to deserve attention of researches in the future.

VIII. MODERN IMPLEMENTATION OF THE FD-BPM

Photonic devices are being constantly improved and developed in the last decade. These newly-designed structures have placed high demands on the numerical modellers. Although the finite element method (FEM) and finite-difference time domain method (FD-TD) are traditionally used today to solve the propagation and modal properties of these newly-designed photonic devices, the FD-BPM is highly applicable in this direction as well.

The employment of the FD-BPM in the design of novel optical fibers with complex geometries and modal properties operating in nonlinear and high power regimes [27,35,36] is already highlighted in Section 5. Numerical simulations of the wave propagation in plasmonics (metallic waveguides supporting surface plasmons having the enormous bandwidth of a light pulse) have been recently accomplished by the use of the FD-BPM [37]. Photonic crystal fibers [38] and other newly-designed modern photonic structures [39], which appear as the result of the advances in the modern nano-fabrication and characterization techniques, can be successfully modelled with the FD-BPM approaches.

IX. CONCLUSION

The FD-BPM remains one of the most widely used techniques for numerical field simulations in integrated photonics and optoelectronics. Some of very promising recently proposed FD-BPM procedures have been addressed and discussed. The overall conclusion is that the improvement of the capabilities of the FD-BPM is still a very attractive area of research in numerical photonics.

ACKNOWLEDGEMENT

This paper was based on research conducted within the Project (Project code: TR32052) funded by the Ministry of Science and Technological Development of the Republic of Serbia (MSTDRS).

REFERENCES

[1] C.L. Xu and W.P. Huang, "Finite-difference beam propagation method for guide-wave optics", PIER 11, pp. 1-49, 1995.

[2] R. Scarmozzino, A. Gopinath, R. Pregla and S. Helfert, "Numerical techniques for modeling guided-wave photonic devices", IEEE Journal of Selected Topics In Quantum Electronics, vol. 6, no. 1, pp. 150-162, 2000.

[3] D.Ž. Djurdjević, *Metoda prostiranja snopa i njena primena u projektovanju fotoničkih i optoelektronskih struktura*, Beograd, Vojnotehnički Institut, 2010.

[4] Y. Chung and N. Dagli, "An assessment of finite difference beam propagation method", IEEE Journal of Quantum Electronics, vol. 26, pp. 1335-1339, 1990.

[5] T.M. Benson, E. Bekker, A. Vukovic and P. Sewell, “The continuing role of beam propagation methods in photonics design”, ICTON-2007, Conference Proceedings, pp.240-243, 2007.
[6] T.M. Benson, P. Sewell, A. Vukovic and D.Z. Djurdjevic, “Advances in the finite difference beam propagation method”, In: Proc. IEEE on Transparent Optical Networks, vol. 2, pp. 36-41, 2001.
[7] T.M. Benson, D.Z. Djurdjevic, A. Vukovic and P. Sewell, “Towards numerical vector Helmholtz solutions in integrated photonics”, In: Proc. IEEE on Transparent Optical Networks, vol. 2, pp. 1-4, 2003.
[8] G.R. Hadley, “Wide-angle beam propagation using Padé approximant operators”, Optics Letters, vol. 17, pp. 1426-1428, 1992.
[9] C.C. Huang and C.C. Huang, “A novel wide-angle beam propagation method based on the spectral collocation scheme for computing titled waveguides”, IEEE Photonics Technology Letters, vol. 17, no 9, pp. 1872-1874, 2005.
[10] C.S. Hsiao, L. Wang and Y.J. Chiang, “An algorithm for beam propagation method in matrix form”, IEEE Journal of Quantum Electronics, vol. 46, no. 3, pp. 332-339, 2010.
[11] G.R. Hadley, “A complex Jacobi iterative method for the indefinite Helmholtz equation”, Journal of Computational Physics, vol. 203, pp. 358-370, 2005.
[12] K.Q. Le and P. Bienstman, “Fast three-dimensional generalized rectangular wide-angle beam propagation method using complex Jacobi iteration”, Journal of OSA B, vol. 26, no. 7, pp. 1469-1472, 2009.
[13] K.Q. Le and P. Bienstman, “Three-dimensional higher-order Pade approximant-based wide-angle beam propagation method using complex Jacobi iteration”, Electronics Letters, vol. 26, no. 3, doi:10.1049/el.2010.3204, 2010.
[14] J. Shibayama, T. Takahashi, J. Yamauchi and H. Nakano, “A three-dimensional horizontally wide-angle noniterative beam-propagation method based on the Alternating-Direction Implicit scheme”, IEEE Photonics Technology Letters, vol. 18, no 5, pp. 661-663, 2006.
[15] J. Shibayama, T. Takahashi, J. Yamauchi and H. Nakano, “A three-dimensional multistep horizontally wide-angle beam-propagation method based on the generalized Douglas scheme”, IEEE Photonics Technology Letters, vol. 18, no 23, pp. 2535-2537, 2006.
[16] E.V. Bekker, P. Sewell, T.M. Benson and A. Vukovic, “Wide-angle Alternating-Direction Implicit finite-difference beam propagation method”, IEEE Journal of Lightwave Technology, vol. 27, no. 14, pp. 2595-2604, 2009.
[17] S. Sujecki, “Generalized rectangular finite difference beam propagation method”, OSA Applied Optics, vol. 47, no. 23, pp. 4280-4286.
[18] D.Z. Djurdjevic, T.M. Benson, P. Sewell and A. Vukovic, “Fast and accurate numerical analysis of 3D curved waveguide couplers”, IEEE Journal of Lightwave Technology, vol. 22, no. 10, pp. 2333-2340, 2004.
[19] K. Chan, P. Sewell, A. Vukovic and T. Benson, “ Oblique Du-Fort Frankel beam propagation method”, Advances in Optoelectronics, Hundawi Publishing Corporation, doi:10.1155/2011/196707, 2011.
[20] M.S. Stern, “Semivectorial polarized finite difference method for optical waveguides with arbitrary index profiles”, IEE Proc. J, vol. 135, pp. 56–63, 1988.
[21] C. Vassallo, “Improvement of finite difference methods for step-index optical waveguides”, IEE Proc. J, vol. 139, pp. 137–142, 1992.
[22] J. Yamauchi et al., “Improved finite-difference beam propagation method based on the generalized Douglas scheme and its application to semivectorial analysis”, IEEE Journal of Lightwave Technology, vol. 14, pp. 2401–2406, 1996.
[23] Y.C. Chiang, Y.P. Chiou, and H.C. Chang, “Improved full-vectorial finite-difference mode solver for optical waveguides with step-index profiles”, IEEE Journal of Lightwave Technology, vol. 20, no. 8, pp. 1609-1618, 2002.
[24] G.R. Hadley, “High-accuracy finite-difference equations for dielectric waveguide analysis I: Uniform regions and dielectric interfaces”, IEEE Journal of Lightwave Technology, vol. 20, no.7, pp. 1210-1218, 2002.
[25] G.R. Hadley, “High-accuracy finite-difference equations for dielectric waveguide analysis II: Dielectric corners”, IEEE Journal of Lightwave Technology, vol. 20, no.7, pp. 1219-1231, 2002.
[26] Y.P. Chiou and C.H. Du, “Arbitrary-order interface conditions for slab structures and their applications in waveguide analysis”, OSA Optics Express, vol. 18, no. 5, pp. 4088-4102, 2010.
[27] G.R. Hadley, “Numerical simulation of waveguides of arbitrary cross-section”, International Journal of Electron. Commun., vol. 58, pp. 1-7, 2004.
[28] V.M.A. Leitao, “Generalized finite differences using fundamental solutions”, International Journal for Numerical Methods in Engineering, vol. 81, pp. 564-583, 2009.
[29] D.Z. Djurdjevic, “New finite-difference formulas for dielectric interfaces”, Facta Universitatis (Niš), Ser.: Elec. Energ, vol. 23, no. 1, pp. 17-35, 2010.
[30] K.Q. Le, R.G. Rubio, P. Bienstman and G.R. Hadley, “The complex Jacobi iterative method for three-dimensional wide-angle beam propagation”, OSA Optics Express, vol. 16, no. 21, pp. 17021-17030, 2008.
[31] N.N. Feng, C. Xu, W.P. Huang and D.G. Fang, “A new preconditioner based on paraxial approximation for stable and efficient reflective beam propagation method”, IEEE Journal of Lightwave Technology, vol. 21, no. 9, pp. 1996-2001, 2003.
[32] C. Ma and E.V. Keuren, “A simple three dimensional wide-angle beam propagation method”, OSA Optics Express, vol. 14, no. 11, pp. 4668-4674, 2006.
[33] H. Cheng, W.P. Zhang, Z.B. Li, W.Y. Zhou and J.G. Tian, “Nonparaxial split-step method with local one-dimensional scheme for three-dimensional wide-angle beam propagation”, IEEE Journal of Lightwave Technology, vol. 27, no. 14, pp. 2717-2723, 2009.
[34] P. Sewell, T.M. Benson and A. Vukovic, “A stable DuFort-Frankel beam-propagation method for lossy structures and those with perfectly matched layers”, IEEE Journal of Lightwave Technology, vol. 23, no. 1, pp. 374-381, 2005.
[35] G.R. Hadley, “Numerical simulation of light transmission through optical fibers: Linear and nonlinear”, OSA-2008, Conference Proceedings, pp. FThE1, 2008.
[36] G. R. Hadley, “Slanted-wall beam propagation”, IEEE Journal of Lightwave Technology, vol. 25, no. 9, pp. 2367-2375, 2007.
[37] B. Hu, P. Sewell, J.G. Wykes, A. Vukovic and T.M. Benson, “Fourth-order accurate sub-sampling for finite-difference analysis of surface plasmon metallic waveguides”, Microwave and Optical Technology Letters, vol. 50, no. 4, pp. 995-1000, 2008.
[38] Y.L. Li, Q.Z. Xue and C.H. Du, “Two-dimensional metallic photonic crystal with point defect analysis using modified finite-difference frequency-domain method”, IEEE Journal of Lightwave Technology, vol. 28, no. 2, pp. 216-222, 2010.
[39] Y.C. Chiang, Y.P. Chiou and H.C. Chang, “Finite-difference frequency-domain analysis of 2-D photonic crystals with curved dielectric interfaces”, IEEE Journal of Lightwave Technology, vol. 26, no. 8, pp. 971-976, 2008.

Investigation of Speech Coding Algorithms

Andon Nenov[1], Georgi Iliev[2], Maria Nenova[3]

Abstract – **In this paper, an investigation of different types of speech coding algorithms is presented. The goal of speech coding systems is to transmit speech with good quality. Different types and specifics of algorithms and their applications are shown. Features of Adaptive Differential Pulse Code Modulation, Code Excited Linear Predictive and MELP coders are presented too.**

The complexity of the algorithm is also a very important point for research. Since speech compression is used in real-time systems, digital signal processors are the best choice for running the algorithm. Implementations on DSP-based systems are not only robust and flexible but also very powerful.

Keywords – **Speech coding algorithms, Coders.**

I. Introduction

The speech coding is generally used for representation of the human voice as a digital signal. The main advantage of use of coding the speech signals is the ability to compress the signal, in a sense to reduce the bit rate of the digital speech signals [1].

The goal of all speech coding systems is to transmit speech with the highest possible quality using the least possible channel capacity. It is important to investigate the bit-rate and the relation with algorithmic complexity for the speech coders. Speech compression is used in real time systems and thus they are implemented on DSP-based systems. The algorithm complexity must be low, due to the power consumption requirements.

During the years development of coders operating at 4,8 kbits/s and below for narrow band and secure transmission in communications and voice applications in Internet arises [2]. Unfortunately the characteristics of those type low data-rate coders are poor. [3][4]

Other type of algorithms is MELP [2]. The MELP vocoder described in [5] is an enhancement of CELP with a number of additional features. These features include a mixed excitation signal (i.e., a mixture of noise and pulse excitation used as input to the synthesis filter), an adaptive spectral enhancement filter, Fourier magnitude modelling of the pulse excitation, and a pulse dispersion filter. The MELP vocoder encodes 22.5 ms of speech to a 54-bit frame.

[1]Andon Nenov is with the Ministry of Inferior, E-mail: and_mail@mail.bg

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

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

II. Coder Implementation

A. Problem Statement

In mobile communication systems one of the main problem is the limited bandwith. Therefore speech coders providing the quality of speech signals at low bit rate are needed. The main objective of the paper is to compare some of the most commonly used algorithms in wireless communication systems: CELP, VSELP, MELP and ADPCM.

Most of the already developed in this manner coders have been adopted in cellular phone standards. The focus is on those coders because of the fact that in UMTS and CDMA systems they are implemented.

On Fig. 1 is depicted one of the possibilities of classification of speech coders:

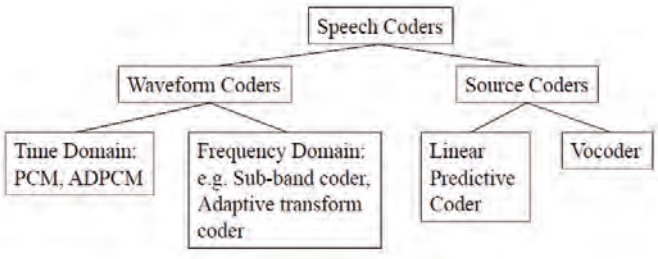


Fig. 1 Varieties of speech coders

A comparison of those types of coders, called by some researcher vocoders is going to be made in the next sections. The structure of coder and decoder and its components will be investigated.

B. ADPCM coders

The coders group called waveform coders can be investigated in the time domain and in the frequency domain. Specific their feature is that they are trying to achieve the time waveform of the speech signals. One of the biggest their advantages is that they are robust for the most of the speech characteristics and are very useful for applications in noisy environment.

The waveform codecs algorithm of work is based on the ability without knowledge of the prior information about the coded signal, to produce a reconstructed signal whose waveform is close to the form of the original signal. Their biggest advantage is that the low complexity. When the data rate is lower than 16 kbits/s the reconstructed speech quality that can be obtained degrades rapidly.

One of the main classes in the time domain group of algorithms is Adaptive Differential Pulse Code Modulation (ADPCM). ADPCM codecs are waveform codecs which

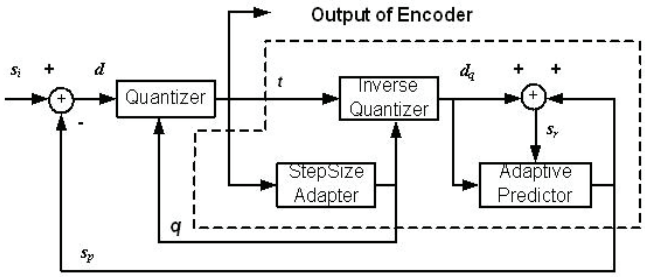


Fig. 2: ADPCM Coder Block diagram

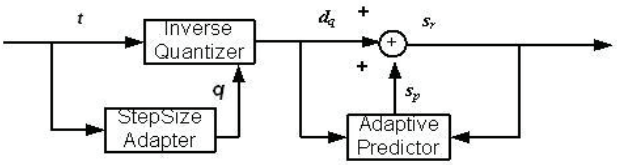


Fig. 3: ADPCM Encoder Block diagram

instead of quantizing the speech signal directly, like PCM codecs, quantize the difference between the speech signal and the prediction that has been made of the speech signal. If the prediction is accurate then the difference between the real and predicted speech samples will have a lower variance than the real speech samples, and will be accurately quantized with fewer bits than would be needed to quantize the original speech samples. At the decoder the quantized difference signal is added to the predicted signal to give the reconstructed speech signal. The performance of the codec is aided by using adaptive prediction and quantization, so that the predictor and difference quantizer adapt to the changing characteristics of the speech being coded.

The CCITT standardized a 32 kbits/s ADPCM, in G.721, which gave reconstructed the speech signal as in the 64 kbits/s PCM codecs. Later in recommendations G.726 and G.727 codecs operating at 40,32,24 and 16 kbits/s were standardized.

The ADPCM algorithm for compression of the signal is shown can be implemented with the blocks depicted on Fig.2. The process is on the base of iterations. On the next iteration, the predicted sample s_p and the quantizer step size index are saved in a structure.

The quantization step size and the predicted sample s_p are initially set to zero. The input s_i to the speech encoder is supposed to be a 16-bit 2's complement speech sample, while the value returned by the speech encoder is an 8-bit number which contains the 4-bit sign magnitude ADPCM code.

On the Fig. 3 the signal of the difference d is produced by subtracting the predicted sample s_p from the input signal s_i. Then the signal d is fed to the quantizer and adaptive quantization is performed on the difference obtained in the previous step.

One of the advantages of the structure on Fig. 2 is that within the encoder there is a decoder inside it. This ensures synchronization between encoder and decoder without requiring any additional data. The dotted lines shown in Fig.2 show the block comprising the embedded decoder.

The ADPCM value is used by the embedded decoder to update the inverse quantiser, which in turn produces a dequantized version d_q of the difference d. To simplify the speech compression process a fixed predictor has been used instead of an adaptive predictor, which significantly reduces the amount of data memory and instruction cycles required. A weighted average of the last six dequantized difference values and the last two predicted values are used by the adaptive predictor of ITU G.721 for its adjustment and updating according to the value of each input sample. At this point, new predicted sample s_r is obtained by adding the dequantized difference d_q to the predicted sample s_p. Finally, the new predicted sample s_r is saved in s_p.

D. CELP Codec (Hybrid codecs)

Usually compression methods are based on entropy coding or on source coding. The entropy coding is also called lossless coding. If entropy coding is implemented then there is used redundancy in order to decrease the amount of the data to be compressed.

In case of information filtered out to reduce the unnecessary elements, the coding is called source coding and this I s the most commonly used for compression of video and audio streams. A typical feature of source coding is the loss of information, which means that the decompressed data stream will not contain all the elements of the original information. One of the best examples of coding with losses is MPEG coding [3].

The so called Hybrid codecs attempt to operate between waveform and so called source codecs. Waveform codecs are capable of providing good quality speech at bit rates about16 kbits/s, but are of limited use at rates below this. Source codecs on the other hand can provide understandable speech at 2.4kbits/s and below, but cannot provide natural sounding speech at any bit rate.

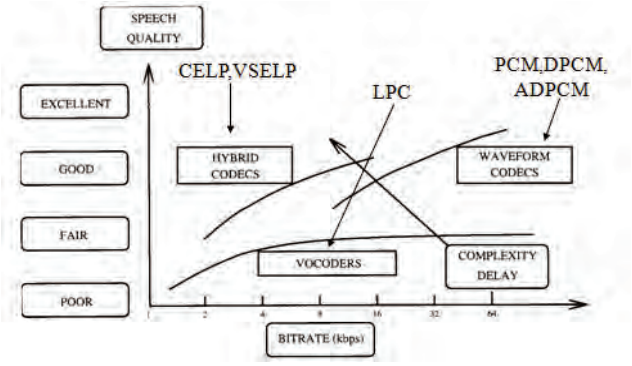


Fig. 4: Speech versus bit-rate classification of speech codecs

As can be seen from the figure 4, hybrid codecs combine techniques from both source and waveform codecs and as a result give good quality with intermediate bit rates.

The coding of this algorithm is based on analysis-by-synthesis search procedures, vector quantization (VQ) and

linear prediction (LP).

The one of the most commonly used hybrid codec type is Analysis-by-Synthesis (AS) codecs. Such coders use the same linear prediction filter model as found in source codecs. However instead of applying a simple two-state, voiced/unvoiced, model to find the necessary input to this filter, the excitation signal is chosen by attempting to match the reconstructed speech waveform as closely as possible to the original speech waveform. Thus AbS codecs combine the techniques of waveform and source codecs.

The principle of work of AS codecs is splitting the input speech to be coded into frames. For each frame parameters are determined for a filter called synthesis filter. The excitation to this synthesis filter is determined by finding the excitation signal, which minimizes the error between the input speech and the reconstructed speech. Thus the name Analysis by Synthesis means that the encoder analyses the input speech by synthesizing many different approximations to it. The basic idea is that each speech sample can be approximated by a linear combination or the preceding samples.

One more type of low bit rate coders implemented in communications is called CELP (Code Excited Linear Predictive). The CELP was first introduced by Atl and Schroder [5]. In order to achieve real-time encoding, the CELP optimisation is divided into smaller, sequential searches using the perceptual weighting function described earlier[6][7][8]. CELP is based on vector quantization. One of the most commonly used algorithm for producing good quality speech at rates below 10kbits/s is CELP.

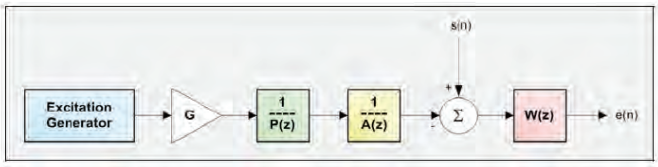


Fig. 5 Scheme of CELP encoders

On the Fig.5 is depicted a block diagram of a CELP encoder. The generator produces sequence which will be multiplied and then given to a filter called pitch synthesizing filter *1/P(z)*.

The received signal is then fed to one other filter with transfer function *1/A(z)*, where

$$H(z) = \frac{1}{A(z)} = \frac{G}{1 + a_1 z^{-1} + a_2 z^{-2}},$$

$$A(z) = 1 + \sum_{k=1}^{p} a_k . z^{-k}$$

The a_k in the formula is coefficient called the Linear Predictive Coefficient (LPC). The coefficients are determined by minimising difference between actual signal and predicted signal by the use of least square method. The variable *p* gives the order of the filter. This filter is intended to model the short-term correlations introduced into the speech by the action of the vocal tract. This kind of coding is also called Linear Predictive Coding (LPC)

The difference between the speech signal and the original speech spectrum *S(n)* is then weighted according to a subjective error criterion, *W(z)* to receive the error signal *e(n)*, which is encoded using vector quantization.

Often CELP is called a hybrid codec because it uses both waveform and source coding techniques. Unfortunately during the process of filter coefficients updates is introduced a high rate of delays. For a typical hybrid codec this delay will be of the order of 50 to 100 ms, and such a delay lead to problems. Thus, many efforts have been focused in providing a standard codec that has as bit rate 16 Kbps, while providing a quality comparable to that provided by the ADPCM 32 Kbps. The major challenge is to reduce the delay to about 5 ms. CELP needs a dedicated hardware to run in real time.

The disadvantage of CELP coding schemes is that they fail to represent the high frequencies in speech at bit rates around 6 kbit/sec or lower. For this reason, the newer CELP schemes are actually combinations of CELP and MLPC (Multipulse LPC), using a limited number of excitation pulses.

There are many papers on the latest CELP/MLPC coders which provide a deeper inside to their operation. The CELP needs a fixed bitrate of only 4.8 kbit/sec for encoding human speech, but it is worth mentioning that among all low-bitrate vocoders CELP demands the highest computational power.

The MELP vocoder described in [9] is an enhancement of CELP with a number of additional features. These features include a mixed excitation signal (i.e., a mixture of noise and pulse excitation used as input to the synthesis filter), an adaptive spectral enhancement filter, Fourier magnitude modelling of the pulse excitation, and a pulse dispersion filter. The MELP vocoder encodes 22.5 ms of speech to a 54-bit frame. Most of the parameters in a typical frame are quantized using suitable vector codebooks, and only the codebook indices are transmitted. MELP is much faster than CELP, but needs at least 4 times as much memory as CELP does.

The disadvantage of CELP coding schemes is that they fail to represent the high frequencies in speech at bit rates around 6 kbit/sec or lower. For this reason, the newer CELP schemes are actually combinations of CELP and MLPC (Multipulse LPC), using a limited number of excitation pulses.[9]

F. VSELP Codec (Hybrid codecs)

There is also one more important type codec with implementation in GSM communications called a half-rate GSM codec. It is a Vector Self-Excited Linear Predictor (VSELP) codec at bit rate of 5.6kbit/s. VSELP codec is a close relative of the CELP codec family explained in the previous chapter. A slight difference is that VSELP uses more than one separate excitation codebook, which is separately scaled by their respective excitation gain factors.

TABLE 1. PERFORMANCE AND COMPLEXITY OF ALGORITHMS

Algorithm	Bit Rate (bit/sec)	MIPS
PCM	64 k	0.01
ADPCM	32 k	2
CELP	4.8 k	16
VSELP	8 k	13.5

In Table 1 is presented comparison between different types of the already investigated algorithms with their bit rate. Implementing the original systems required several hundred MIPS (Million Instructions Per Second).

As the bit rate reduces the computational complexity increases. Much of the research being done has concentrated on reducing this load in order to facilitate implementing the algorithm on available silicon.

III. CONCLUSION

Many communication channels suffer from noise, interference or distortion due to hardware imperfections, or physical limitations. The goal of error control coding is to encode information in such a way that even if the channel (or storage medium) introduces errors, the receiver can correct the errors and recover the original transmitted information.

This report discusses algorithms such as ADPCM and CELP, MELP, VSELP and other encoding methods. It also discusses the technological advances. Speech can be coded at many levels. Lower bit-rates are achieved by imposing more constraints of the speech production mechanism are applied. The failure becomes more catastrophic as the bit rate is reduced.

REFERENCES

[1] Wai C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*, John Wiley & Sons, Inc., 2003.

[2] Kohler M.A., "A Comparison of the New 2400BPS MELP Federal Standard with other Standard Coderds, IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP'97), Munich, Germany

[3] Spanias, A.S. (1994). Speech coding: a tutorial review. In: *Proceedings of the IEEE* Vol. 82, October 1994, pp. 1541-1582.

[4] Rabiner L.R. (1994). Applications of Voice Processing to Telecommunications. In: *Proceedings of the IEEE* Vol. 82 No. 2, February 1994, pp. 174-180.

[5] Schroeder M.R. and Atal B., " Code Exited Linear Prediction (CELP): High Quality Speech at Very Low bit Rates", Proc. ICASSP-85, P. 937, Tampa, Apr. 1985

[6] Liang, C., and Xiongwel, Z. (1994). A New 1.2kb/s Speech Coding Algorithm and its Real-time Implementation on TMS320LC548. *Institute of CommunicationsEngineering, PLA University of Science and Technology*, Nanjing

[7] Kohler M.A., "A Comparison of the New 2400BPS MELP Federal Standard with other Standard Coderds, IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP'97), Munich, Germany

[8] TEXAS INSTRUMENTS (2002). *TMS320C6000 Programmer's Guide*. http://focus.ti.com/lit/ug/ spru198g/spru198g.pdf.

[9] U.S. DEPARTMENT OF DEFENSE (1998). *Specifications for the Analog to Digital Conversion of Voice by 2,400 Bit/Second Mixed Excitation Linear Prediction*

Critical Telecommunication Infrastructure Management in Express Mail Industry

Momčilo Dobrodolac[1], Dejan Marković[2], Mladenka Blagojević[3]

Abstract – **Critical infrastructure is a term used to describe items that are essential for the functioning of some system. Critical telecommunication infrastructure in express mail industry mainly refers to communication between dispatcher and courier. This type of communication on one hand is the most demanded one and on the other hand very important for service quality achievements. The purpose of this paper is to propose technical solution for this system that would lead to high quality of service.**

Keywords – **Critical infrastructure, Telecommunications, Express mail industry.**

I. Introduction

Critical infrastructure is a term used usually by governments to describe assets that are essential for the functioning of a society and economy. Speaking about Express mail industry, critical telecommunication infrastructure refers to communication between all entities participating in technological process. But the most demanded one is communication between dispatcher and courier. Therefore, the main aim of this paper is to designate crucial parameters for this telecommunication system design.

Implementation of the adequate telecommunication system between dispatcher and courier is important not only for reducing the costs and increasing the productivity but also for higher service quality achievements.

II. The Impact of Telecommunication System Dispatcher – Courier on Technological Process

In this chapter, manual and automated technological process, depending on the use of appropriate communication means, will be presented. In the first case, communication with the courier is done via mobile phone, by calling the courier or by sending him/her SMS messages. The other, automated system; uses the modern GPRS terminals.[1] Although the second case provides an opportunity for improvement of work processes in several points of the production chain, which will be discussed in the following text, it is still used in practice in a very small percentage. The reason for this phenomenon lies in the fact that this system is part of the public communication system, which does not provide an adequate level of quality network required for express and courier service.

The scheme of manual technological process is presented in Fig. 1. Technological process begins with client's giving the mail at Post Office counter or, like more common case, by calling the Call Center in order to ask for the courier to come to client's home or business address. Request of the sender is registered at the central server and forwarded to dispatcher. He/she has to determine to which courier the request should be send. With the mention request, the courier must obtain information about the sender, recipient, type of required services, some specific services, weight and shipping volume, the way of payment and others.

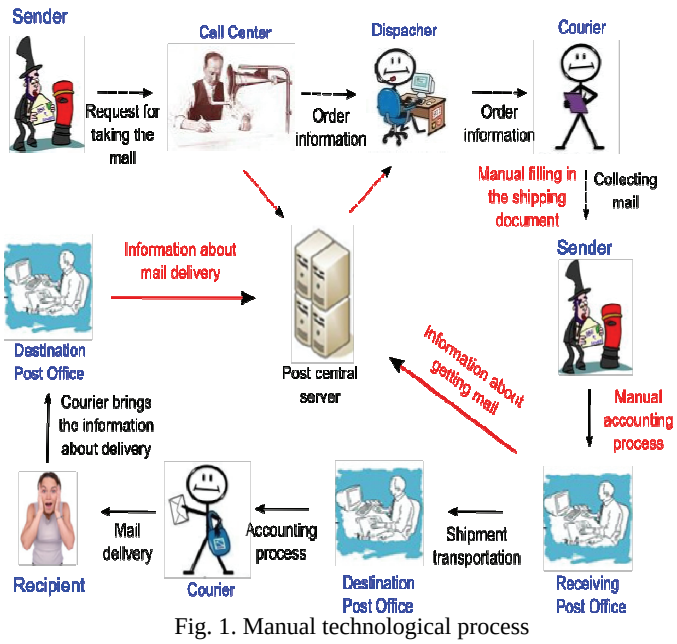


Fig. 1. Manual technological process of express mail transportation

In the manual technological process, courier receives the information via SMS or in direct conversation via mobile phones. Problems that arise in this kind of work may be various. One of them is the small capacity of the SMS message. All the necessary information can not be always conveyed in a single SMS message. Then, the cost of calls via mobile phones can significantly increase the cost of services. Also, if the conversation takes place during the drive, it results in courier's reduced security, because courier needs to note the data for following pick up. In these circumstances, it is a great possibility that an error occurs in the work.

[1]Momčilo Dobrodolac is with the Faculty of Transport and Traffic Engineering – University of Belgrade, Vojvode Stepe 305, 161904 Belgrade, Serbia, E-mail: m.dobrodolac@sf.bg.ac.rs.

[2]Dejan Marković is with the Faculty of Transport and Traffic Engineering – University of Belgrade, Vojvode Stepe 305, 161904 Belgrade, Serbia, E-mail: mdejan@sf.bg.ac.rs.

[3]Mladenka Blagojević is with the Faculty of Transport and Traffic Engineering – University of Belgrade, Vojvode Stepe 305, 161904 Belgrade, Serbia, E-mail: m.blagojevic@sf.bg.ac.rs.

When the courier arrives at the location in order to take the mail, shipping document should be filled in. The courier manually fills in the document with data obtained from dispatcher, but after they are verified by sender. The sender gets a copy of shipping document like the evidence that the shipment was submitted to the transportation. The other copies follow the mail during the transportation.

The items collected during the working day, courier brings to the Receiving Post Office. Shipments are unloaded from the vehicle and enter the sorting process. Courier has an obligation to visit the accounting worker to give the evidence about the mail that has been taken this day and to deliver all collected money, if some senders paid the postage in cash. Accounting worker has to register every item, entering the data into computer. Working in manual technological process means that each shipping document should be handled manually.

When the shipment arrives at the destination point, couriers need to charge the shipments for delivery. For this purpose, every courier gets the delivery book, representing the document where all the items to be delivered are listed. In this delivery book, the recipient puts its signature as a proof that the mail has been delivered.

After the delivery of all shipments, courier returns to the Destination Post Office, to the accounting worker. Accounting worker reads from delivery book information about delivered mail and the time of delivery and puts it into computer. This is the end of manual technology process.

What has brought the technological process to automatic one is the implementation of modern GPRS mobile terminals. They greatly facilitated phase of receiving shipment. In this case, all data received in the call center, are available to courier via GPRS terminal in electronic form. As previously mentioned, it is data about shipments, types of service, the sender, recipient, the way of payment, special services, etc. Compared to mobile phone conversation, this kind of work brings to cost savings and increased courier's security in moment of driving.

This way of technological process leads to the reduction in errors and better collection of postage. One of the numerous advantages is that the postage is automatically calculated. All data received in the call center are available to the courier during mail collection. After sender's confirmation, the postage is automatically calculated.

The next advantage of automated process that refers to collecting mail is in filling shipping document. Having in mind that all data in electronic form are available to courier over the server, there is no need for manual filling in the shipping document. Courier just has to print it using GPRS terminal. In this way time saving is made. On printed document, there is also a bar code, so there are savings as well on bar code labels. Even more importantly, the courier does not waste the time on hand writing the shipping document. This reduces the time spent on collecting location, i.e. accelerates the technological process, reduces the enterprise's costs and increases productivity.

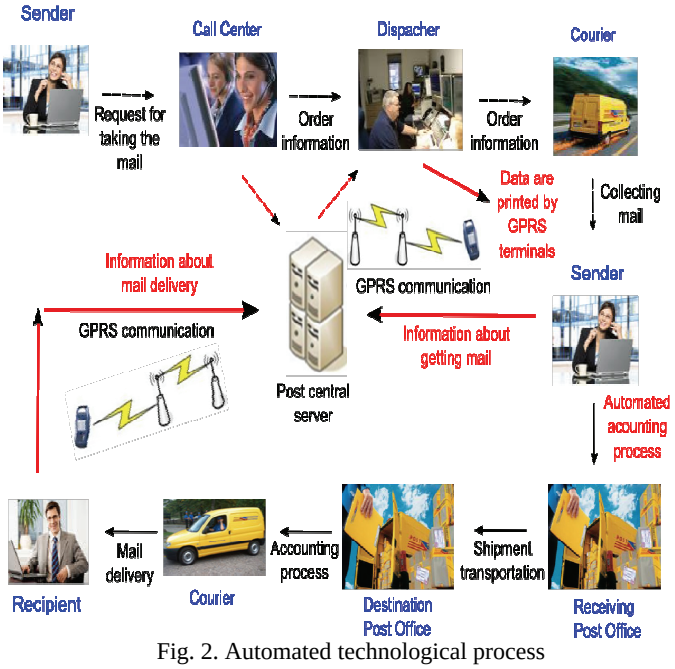


Fig. 2. Automated technological process of express mail transportation

Since the GPRS terminals are in relation to the central server, there is on-line transmission of information about collected shipment and therefore enables connection with Track and Trace system for tracking shipments.

Upon arrival in the Receiving Post Office, courier's discharge procedure is much simpler and shorter because all data are already on central server. Therefore, there is less employees working on the computer entry, resulting in a reduction of total costs.

In the delivery phase, when the GPRS terminal are used, delivery book exists only to be signed by recipient like a paper proof that the shipment is received. All the remaining necessary data is entered into the mobile terminal. Also, the delivery information is immediately transmitted to the central server, although the courier has not yet returned to the Post Office. Automatically, the data are ready for discharge before his return and accounting employee only check whether a particular assigned shipment status is correct. In this stage, there is also on-line transmission of information on the status of delivery of items, i.e. monitoring the delivery in real-time, supporting the Track and Trace system in this way.

Comparing manual and automated production process, it can be concluded that there are many advantages in using the GPRS terminal at several points of production cycle. These benefits greatly contribute to reducing costs, increasing speed of providing services, better utilization of labor, increasing the quality of services and better productivity. However, beside all these advantages, the automated system is implemented in a very small number of cases. The most common reason for this is poor availability of GPRS terminals because the system is designed to be part of a public communication network. As a potential solution to this problem, some of the Private Mobile Radio systems can be implemented. In further text, some parameters for adequate communication system design will be presented.

III. Parameters for Communication System Design

As in any other liberalized market, organizations operating in express mail industry face the key business challenges of delivering profits. Innovative communications technologies are being adopted to achieve new targets in efficiency and productivity. Yet the myriad of options available makes selection particularly difficult.

As it was investigated, the existing communication system that was part of the public network has not met the expectations. Therefore, the new communication system should be introduced. As a possible solution, Private Mobile Radio systems, also known as communication systems for closed user groups, can be implemented. In further text, some basic expectations of these systems will be introduced.

Coverage. Closed user groups require a high level of geographical coverage of their mobile communication systems. This implies that the inaccessible regions, such as the basements or distant warehouses where the couriers often collect or deliver items, should not introduce any restrictions regarding the possibility of connection. Lack of coverage significantly reduces the operational efficiency of couriers, which leads to pure quality of express mail service.

Availability. It is important that express mail industry has the access to communication services with acceptable quality level. Quality level of mobile communication systems must meet the traffic needs in peak hours. The network must be flexible enough to support communications in emergency situations such as the failure of some element of network. In order to provide the required level of quality, mobile communication system must support access to voice and data services using and combining some of the following mechanisms:

- Preemption, i.e. if all network resources are busy, the user must have the possibility of initiating a priority call. There should be the procedure that can free up resources needed for establishing this kind of call,
- The assignment and management of the frequency band dedicated to the customers in terms of congestion in the network,
- Facilitating the management in case of network congestion can be achieved by direct communication between two mobile stations without using the base stations.

Reliability and flexibility. One of the key requirements is that the infrastructure has maximum flexibility with a sufficient number of redundant elements, so that individual failures can not cause system failure. Individual components in the network must have the ability to communicate in different ways, so in case of broken link, communication can be achieved through a local base station that covers a specific operational area.

Security. Security and confidentiality are fundamental to the activities of closed user groups. The network must be protected against misuse and unauthorized access to data and resources, including expanded encryption techniques and user authentication.

In addition to the above requirements, there are some other requirements, such as a high quality audio for use in noisy environments, support for GPS positioning, full colour display for accurate display of pictures and maps.

Very important feature of every communication system is its possibility of expansion. Some users can start with small number of people in group, with low rate of communication services utilization. After some time, demands usually arise and network should be able to adjust, for example to add new users, increase coverage (geographical or in-building), work in multiple dispatcher system etc.

In order to deliver flexibility and reduce training overheads, some key elements must be considered when selecting terminals like a common user interface across mobiles and portables, one-touch functions and ease-of-configuration.[2]

In the past, several digital trunked Professional Mobile Radio systems have been developed by different manufacturers for public and nonpublic applications. In the United States these systems are known as digital *specialized mobile radio* (SMR), whereas in Europe this service is called *public access mobile radio* (PAMR). The first attempts to develop digital systems of this type date back to the second half of the 1980s.[3]

In Europe one of the most popular systems of this type are TETRA, EDACS and TETRAPOL. Most of these systems no longer play a significant role with one exception – TETRA, because manufacturers strongly promote this communication system.

Communication system iDEN holds a significant market share in the United States and many other countries.

Professional Mobile Radio systems have variety of application possibilities. They are used by Police, Military, Fire Departments, Ambulance, Transportation companies and other industries. In next section, the possible implementation in postal sector is presented.

IV. The Implementation of Professional Mobile Radio in Postal System

The importance of adequate communication system for the efficiency of technological process in postal system, especially in express mail industry is presented in the second chapter of this paper. By analyzing the situation in the postal system of Serbia it can be concluded that there are huge opportunities for improvement in the field of communications. As a potential solution some of the systems for Professional Mobile Radio can be applied.

In the process of communication system design it is necessary to take into account the economic parameter, i.e. the costs. In that sense, good characteristic of Professional Mobile Radio system is that more departments or services using communication for closed user groups may use a common infrastructure but to have a separate functioning. Thus, the introduction of such system should involve coordinating the various administrations interested in use of communication system for closed user groups. Many independent organizations (public or private) can form their own sub-

systems within a single infrastructure and thus work smoothly with the simultaneous distribution of the total investment cost.

Given that many postal systems are state-owned and that many Professional Mobile Radio users are also state organizations, such as police, fire department, ambulance, etc. in this case it is very useful to form an integral state strategy for communication development. In Figure 3 it is presented that different services can work independent within a common infrastructure.

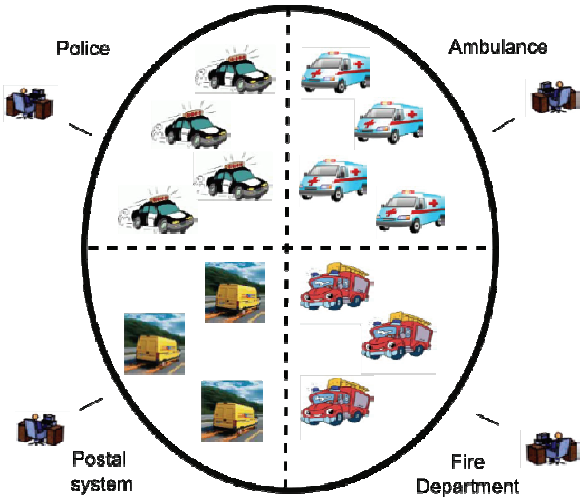


Fig. 3. Independent networks within a common infrastructure [4]

When some government plans to improve communications for their public services, it has two options available:

1. To build one or more independent networks to serve one or more groups of users or

2. To implement the required services in public mobile networks with the necessary protection mechanisms and support systems.

Regarding Serbian integral strategy, one should have in mind that TETRA communication system [5] is already implemented in Serbian Police Department. It would be useful to use the existing infrastructure as well for other services.

TETRA has from the beginning been designed as a trunked system that effectively and economically supports shared usage of the network by several organizations, yet maintaining privacy and mutual security. Virtual networking inside the TETRA network enables each organization to operate independently, but still enjoy the benefits of a large, high-functionality system with efficient resource employment. [6]

Elements of infrastructure owned by provider and terminals owned by users can be from different manufacturers, and total score results in lower costs and better equipment. One such network, which uses equipment from different manufacturers, was represented in Greece. Motorola, Nokia and Siemens were selected to provide the secure two-way radio communications system that will be used by public safety agencies during the 2004 Olympic Games in Athens. There were two TETRA networks in function. Motorola and Nokia provided the terminals, while Siemens and Motorola were responsible for communication systems with dispatcher. This new network, with over 100 base stations built in time of 10 months, during the Olympic Games served approximately 17,000 users and over 200,000 calls a day. Such examples of efficiency TETRA network can be found throughout Europe.

Speaking about the implementation of TETRA systems in Serbian Post, the use may be broader than those described in express mail service. This system could be a good support to other projects in postal system, like the project of APM (Automatic Vehicle Monitoring) or CAS (Centralized Alarm System). It is possible to set TETRA system as the basic network for transmission of information about the location and status of vehicles that are monitored by the APM system and transmission of alarm signals and data to the operational center of the CAS.

V. Conclusion

The optimization of technological process in express mail industry greatly depends on adequate communications. Efficient communication system design leads to improved quality of service and saving in materials, labor and working time. In this paper, basic expectations about communication network for postal system are presented. As one of the possibilities for adequately communication, TETRA system is introduced.

Acknowledgement

This paper is supported by Serbian Ministry of Science and Technological Development with Project 36022.

References

[1] Momčilo Dobrodolac, Dejan Marković, Bojan Stanivuković, Mladenka Blagojević, „Express Mail Service Quality Improvement Using TETRA Communication System“, XLV International Scientific Conference on Information, Communication and Energy Systems and Technologies – ICEST 2010, pp. 521 – 524, Ohrid, Macedonia, 2010.

[2] Motorola Australia, *TETRA: Enabling Critical Communications in the Oil and Gas Sector*, Motorola, Australia, 2009. (www.motorola.com.au)

[3] H.-P. A. Ketterling, *Introduction to Digital Professional Mobile Radio*, Artech House, London, 2004.

[4] Momčilo Dobrodolac, “Electronic Communications in Function of Higher Quality Achievements in Express Mail Industry”, PhD Dissertation, The Faculty of Transport and Traffic Engineering University of Belgrade, Serbia, 2011.

[5] www.tetramou.com

[6] Damm Cellular Systems A/S, *TetraFlex® V.7.4 System Description*, Damm Cellular Systems A/S, Denmark, 2010. (www.damm.dk)

Text Data-Hiding

Nenad O. Vesić[1] and Dušan J. Simjanović[2]

Abstract – **In this paper an algorithm for text data-hiding will be presented. This algorithm protects a personal data. Programs in the software package MATHEMATICA 7.0 will be given. This algorithm is an application of geometry in cryptography.**

Keywords – **curve, matrix, code, key**

I. INTRODUCTION

Group based cryptography was presented in the [1]. Those algorithms are based on the unknown groups. An algorithm which will be presented in this paper is based on the known group (group of regular matrices). In the [2] is presented text data-hiding with set of matrices, where symbols are encoded with one row of a matrix from this set. RSA algorithm [3] is algorithm which is based on large prime numbers. Preferment of the RSA algorithm [3] in this paper will be presented. An algorithm which will be presented in this paper is a combination of algorithms from [1-3]. Large prime numbers are needed in the RSA algorithm. Prime numbers are not required in the algorithm presented in this paper.

II. PRELIMINARY DEFINITIONS

The next tabular is needed for the encrypting using the algorithm which will be presented in this paper.

TABLE 1: THE CHARACTERS

r\c	1	2	3	4	5	6	7
1	0	@	#	{	}	[	%
2	1	.	"	_	:	]	?
3	A	B	C	2	-	=	!
4	D	E	F	3	\	;	/
5	G	H	I	4	×	&	≠
6	J	K	L	5	*	,	~
7	M	N	O	6	^	(	≈
8	P	Q	R	S	7	)	≅
9	T	U	V	8	+	‘	≡
10	W	X	Y	Z	9	°	␣

[1]Nenad O. Vesić is with the Faculty of Science and Mathematics, Višegradska 33, 18000 Niš, Serbia, E-mail: vesic.specijalac@gmail.com
Project: 174012, Serbian Ministry of Science
[2]Dušan J. Simjanović is with the Gimnazija „Svetozar Marković“, Branka Radičevića 1, 18000 Niš, Serbia, E-mail: dsimce@gmail.com

For the matrices $A=[a^i_j]$ and $B=[b^i_j]$ by the type $m\times n$ with integer entries, matrix

$$C=[c^i_j]=[a^i_j b^i_j] \tag{1}$$

is $*-$ ***product*** of the matrices A and B.

Let us define the following operation with matrices:

For a matrix $A=[a^i_j]$ by the type $m\times n$ with non-zero elements, the matrix:

$$\bar{A}=\left[\frac{1}{a^i_j}\right] \tag{2}$$

is $*-$ ***inverse matrix*** of the matrix A.

It is evident that following holds:

$$A*\bar{A}=\begin{bmatrix}1 & \cdots & 1\\ \vdots & \ddots & \vdots\\ 1 & \cdots & 1\end{bmatrix}.$$

The matrices whose entries are the functions of one variable (matrix curves) will be considered in this paper. Let us tell something about those matrices.

Consider the matrix:

$$\boldsymbol{M}=\boldsymbol{M}(t)=[m^i_j(t)] \tag{3}$$

by the type $p\times q$, where:

$$m^i_j:(a,b)\to\mathbf{R}$$

are continuous functions at the interval $(a,b)\subseteq\mathbf{R}$. This matrix is called the ***matrix curve*** at the interval (a,b) by the type $p\times q$.

Elements of the matrix curve could be differential functions of the order $\alpha\in\mathbf{N}$. Thoses matrix curves are α **- differential matrix curves**.

For the α - differential matrix curve $\boldsymbol{M}=\boldsymbol{M}(t)=[m^i_j(t)]$ of the type $p\times q$, the matrix curve:

$$\boldsymbol{M}'(t)=[(m^i_j)'(t)] \tag{4}$$

of the type $p\times q$ is **tangent curve** of the matrix curve M.

Matrix curve:

$$\boldsymbol{M}^{I_0}(t)=[m^{I_0}_j(t)] \tag{5}$$

of the type $1\times n$ is I_0 - **curve** of the matrix curve M for $1\le I_0\le m$.

Algorithms based on tangent vectors of the matrix curves in this paper will be presented.

III. ALGORITHM

Matrix curves with the polynomials of the degree two in this paper will be used. The tabular of the symbols (Table 1.) will be used in this article. This tabular could be arbitrary given.

Matrix curves by the type 5×3 whose entries are polynomials of the degree two encrypt a text-data in the way presented below.

1. Step

Each character is encrypted by a triplet whose coordinates are polynomials of the degree two. Only one of the polynomials has complex roots:

$$z_{1,2}=p\pm iq,$$

where p is an integer between 1 and 10. The parameter $i=p$ (real part of the root z_i) presents the row of the Table 1 where that symbol is. The parameter $j=q\ mod\ r$ where r is the number of element in the $i-th$ row of the Table 1) presents the column of the Table 1 in which that symbol is.

The text is encrypted character-by-character using the ordered set of matrices by the type 5×3. In the matrix, the row which satisfies the previous condition (only one polynomial has complex roots $z_{1,2}=p\pm iq$, $p,q\in\mathbf{Z}, p=\overline{1,10}$) generates the character. If a row does not satisfy previous condition, it generates *access character*. Any text which has $5k$ characters ($k\in\mathbf{N}$) or otherwise, will be supplemented to $5l$ characters ($l\in\mathbf{N}$) with access characters. This supplement is required for avoiding a possible brute force attack in text with standard beginning (diplomatic texts, for example).

2. Step

Let we have the ordered set $\boldsymbol{M}=\{\boldsymbol{M}_1,\cdots,\boldsymbol{M}_l\}$ of the matrix curves which encrypt text-data. System $\boldsymbol{K}=\{\boldsymbol{K}_1,\boldsymbol{K}_2,\cdots,\boldsymbol{K}_l\}=\{[(\boldsymbol{k}_j^i)_s]\}=[m_j^i(0)_s]$, $s=\overline{1,l}$, is important element for hiding.

- Let the ordered sets of matrices $P=\{P_1,P_2,\cdots,P_l\}$ and $Q=\{Q_1,Q_2,\cdots,Q_l\}$ be known like the set of matrix curves
$$\boldsymbol{B}(t)=\{\boldsymbol{B}_1,\cdots,\boldsymbol{B}_l\}=\{\boldsymbol{B}_S\},$$
$$\boldsymbol{B}_S(0)=[0]_{5\times 3},$$
which entries are the polynomials of the degree two.
- Encrypted text-data is:
$$\boldsymbol{C}=\{\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_l\}=$$
$$=\{P_S*(Q_S*\boldsymbol{M}_S+\boldsymbol{B}_S)'\},\qquad(6)$$
$s=\overline{1,l}$.
- The set $\boldsymbol{M}=\{\boldsymbol{M}_1,\cdots,\boldsymbol{M}_l\}=\{\boldsymbol{M}_S\}$ is:
$$\boldsymbol{M}=\{\boldsymbol{M}_1,\boldsymbol{M}_2,\cdots,\boldsymbol{M}_l\}=$$
$$=\{\overline{Q}_S*(\int_0^t(\overline{P}_S*\boldsymbol{C}_S)dt-\boldsymbol{B}_S)+K_S\}\qquad(7)$$
$s=\overline{1,l}$.

Text could be decrypted directly from the descriptions of the matrix curves $\boldsymbol{M}_S$ (step 1). Sets {P, Q, K, B} is private code [4].

The programs are given in the MATHEMATICA 7. 0 in the next section.

IV. PROGRAMS

A. Auxiliary programs

```
PolynomialConstantProduct[a_,p_]:
  =Module[{pcp},c0=a*p/.t->0;c1=a*Dt[p,t]/.t->0;
  c2=a*(Dt[p,{t,2}]/.t->0)/2;pcp=c2*t^2+c1*t+c0;
 pcp];


SystemStarMatrixProduct[a_,b_]:
  =Module[{ssmp=a},For[i=1,i<=Dimensions[a][[1]],i++,
    For[j=1,j<=Dimensions[a[[i]]][[1]],j++,
     For[k=1,k<=Dimensions[a[[i,j]]][[1]],k++,
  ssmp[[i,j,k]]=PolynomialConstantProduct[a[[i,j,k]],b[[i,j,k]]]]]];
 ssmp];
```

Previous two programs are necessary for hiding of a code of the text data.

```
SystemStarInverse[a_]:
  =Module[{ssi=a},For[i=1,i<=Dimensions[a][[1]],i++,
    For[j=1,j<=Dimensions[a[[i]]][[1]],j++,
 For[k=1,k<=Dimensions[a[[i,j]]][[1]],k++,ssi[[i,j,k]]=1/ssi[[i,j,k]]]]];
 ssi];
```

This program returns system b of matrices, generated with the system of matrices which hide text data. Hiding of the hidden text data with system b is encrypted text data.

```
KSetForming[a_]:
  =Module[{ksf=a},For[i=1,i<=Dimensions[a][[1]],i++,
    For[j=1,j<=Dimensions[a[[i]]][[1]],j++,
 For[k=1,k<=Dimensions[a[[i,j]]][[1]],k++,
 ksf[[i,j,k]]=ksf[[i,j,k]]/.t->0]]];
 ksf];
```

The result of this program is system of matrices which presents status of the code which hides a text for $t = 0$.

```
PossibleCodePolynomials[m_]:
  =Module[{pcp={}},For[i=1,i<=Dimensions[m][[1]],i++,
    For[j=1,j<=Dimensions[m[[i]]][[1]],j++,cc=0;nl0=0;
     For[k=1,k<=Dimensions[m[[i,j]]][[1]],k++,
   If[Discriminant[m[[i,j,k]],t]<0,nl0++;cc=k,nl0=nl0]];
   If[nl0==1,pcp=Append[pcp,m[[i,j,cc]]],pcp=pcp]]];
 pcp];
```

The result of this program is set of the polynomials from characters which are not excess characters which hide any chacacter.

```
SymbolPositions[m_,key_]:
  =Module[{sp={}},pcp=PossibleCodePolynomials[m];
    For[i=1,i<=Dimensions[pcp][[1]],i++,
     rr=-(Dt[pcp[[i]],t]/.t->0)/Dt[pcp[[i]],{t,2}];
     cc=Abs[Discriminant[pcp[[i]],t]/Dt[pcp[[i]],{t,2}]];
    If[(IntegerQ[rr])&&(1<=rr<=10)&&(IntegerQ[cc]),
    If[Mod[cc,Dimensions[key[[rr]]][[1]]]==0,
     cc=Dimensions[key[[rr]]][[1]],
      cc=Mod[cc,Dimensions[key[[rr]]][[1]]]],cc=0];
  If[cc!=0,sp=Append[sp,{rr,cc}],sp=sp]];
 sp];
```

Result of this program is the set of positions of symbols from the encrypted text. Those positions need one step more for finding of final positions. Parameter j must be found.

B. Main programs

```
CodeHiding[p_,q_,b_,m_]:
  =Module[{ch},k=KSetForming[m];hm=m-k;
  fd=SystemStarMatrixProduct[q,hm]+b;
    der=Dt[fd,t];ch=SystemStarMatrixProduct[p,der];
 ch];
```

This program hides encrypted text data.

```
CodeDehiding[p_,q_,b_,k_,c_]:
  =Module[{cd},ip=SystemStarInverse[p];iq=SystemStarInverse[q];
    fs=SystemStarMatrixProduct[ip,c];ss=Integrate[fs,t];ts=ss-b;
 fthss=SystemStarMatrixProduct[iq,ts];cd=fthss+k;
 cd];
```

The result of this program is code which was hidden.

```
DecodingPositions[p_,q_,b_,k_,c_,key_]:
  =SymbolPositions[CodeDehiding[p,q,b,k,c],key];
```

The result of this program is the final set of the positions of symbols in the tabular 1.

C. *Explanations of previous programs*

Auxiliary programs execute the operations defined in the introduction. Program *CodeHiding* hides primary code of the text. Result of the program *DecodingPositions* is set of positions of characters in the encrypted text.

V. CONCLUSION

In this paper text data-hiding algorithm was presented. The texts could be encrypted with large numbers (not necessary primes like in the RSA algorithm), like coefficients of the polynomials. There exists only one problem using encrypting/decrypting algorithm presented in this paper. The problem is to create a code.

Factorization attack is impossible because of the set of matrix curves B, while brute force attack is impossible because of the large coefficients of the polynomials which encrypts the text.

REFERENCES

[1] Alexei Myasnikov, Vladimir Shpilrain, Alexander Ushakov, *Group-based cryptography*, Birkhauser, 2008.

[2] R. Villan, S. Voloshynovskiy, O. Koval, J. Vila, E. Topak, F. Deguillaume, Y. Rystar, T. Pun, *Text Data-Hiding for Digital and Printed Documents*: *Theoretical and Practical Considerations*, http://cvml.unige.ch/publications/postscript/2006/SPIE-EI-2006-Text-Data-Hiding-paper.pdf

[3] RSA. Wikipedia [Online], http://en.wikipedia.org/wiki/RSA

[4] Private-key (or secret-key) cryptography. *Kioskea.net* [Online] http://en.kioskea.net/contents/crypto/cleprivee.php3

An Application Scenario for IPTV Transmission over WiMAX

Grigor Y. Mihaylov[1], Teodor B. Iliev[2]

Abstract – **IPTV is a system where a digital television service is delivered using Internet Protocol over a network infrastructure, which may include delivery by a broadband connection. IEEE 802.16 standard specifies the air interface of fixed broadband wireless access systems supporting multimedia services.**

Keywords – **IPTV, WiMAX, OFDM, PHY layer.**

I. Introduction

A general definition of IPTV is television content that, instead of being delivered through traditional broadcast and cable formats, is received by the viewer through the technologies used for computer networks. IPTV is typically supplied by a service provider using a closed network infrastructure. This closed network approach is in competition with the delivery of TV content over the public Internet, called Internet Television. In businesses, IPTV may be used to deliver television content over corporate LANs.

In the past, this technology has been restricted by low broadband penetration. Nowadays many of the world's major telecommunications providers are exploring IPTV as a new revenue opportunity from their existing markets and as a defensive measure against encroachment from more conventional cable television services. [5]

II. IPTV Infrastructure

A typical IPTV infrastructure consists of three major building blocks: content acquisition, content distribution and content consumption, constructed in a hierarchy of national, regional, local coverage, to consumer premises. Each part is implemented with different elements and must be able to expand when needed. Figure 1 is a simplified system diagram describing major components of a typical IPTV system. They are: [6]

• Acquisition Servers (A-Server) which encode video and add DRM or metadata;

• Distribution Servers (D-Server) which provide caching and QoS control;

• VoD (Video-on-Demand) Creators and Servers which retain a library of encoded VoD movies to provide Video-on-Demand services;

• IP Routers which route IP packets and provide fast reroute for failover;

• Residential Gateways (RG) which are IP routers for bundled services at home and Set-Top Boxes (STB) which receive video streams for TV sets.

As shown in Figure 1, there are basically two types of content sources for IPTV: broadcast and VoD. Although both can utilize the same regional or local distribution networks, they each have their own special challenges. [7]

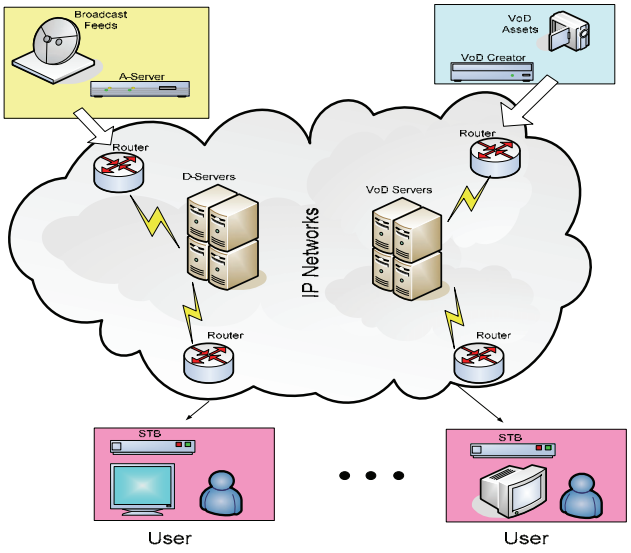


Fig 1. A simplified system diagram of an IPTV system

A. Broadcast Programs

For broadcast programs, each channel is provided by a multicast video stream. This effectively reduces the network bandwidth demand from one stream per viewer to one stream per channel. In this situation, a viewer changes the channel by leaving one multicast group and joining a different group. Switching channels in a digital environment is inherently slower than switching channels in an analog system. The delay is primarily caused by performing IGMP (Internet Group Management Protocol) processing at multicast group change, handling multiple stream (e.g. audio and video stream) synchronization, the wait for an anchor point, such as the start of a Group of Pictures (GOP), or key frame, and the lag at filling the play-out buffer. When the broadcast content is encrypted, decryption key acquisition and management may cause an additional delay. This problem is especially challenging when multiple viewers surf through many channels at the same time.

[1]Grigor Y. Mihaylov is with the Department Telecommunications, 8 Studentska Str., 7017 Ruse, Bulgaria,
E-mail: gmihaylov@uni-ruse.bg
[2]Teodor B. Iliev is with the Department of Telecommunications, 8 Studentska Str., 7017 Ruse, Bulgaria,
E-mail: tiliev@ecs.uni-ruse.bg

In addition to channel switching, there are other stream change cases which make IPTV both charming and challenging. A video server may want to composite multiple video sources to generate anew stream. One example is to replace national content or content from other regions with the content ingested locally. This is particularly useful for the information such as weather or advertisements. A more difficult scenario is to broadcast a sports event with multi-camera streams. [7]

Synthesized video from a selected viewpoint may be generated according to viewer's preference. IPTV opens the door for integrating such multi-view video or free view point TV systems. In Section IV, we will present some further discussions and existing technique solutions on seamless stream switching.

B. VoD Programs

For VoD programs, there are two scenarios: time-shifted and real-time. In the former case, video may be downloaded first and viewed later; In the latter case, video needs to be streamed to the user in real-time, which has more stringent resources requirement for VoD systems. In real-time VoD, multicast may not work, as cases when two or more users request the same movies at the same time rarely happen. Proper bandwidth provisioning is needed to guarantee the delivery of various unicast streams. Apparently, this approach imposes significant pressure on network resources.

Serving a large number of VoD requests with unicast streams can become a nightmare and any new requests may have to be turned down at some point. Actually, some service providers have to use a dedicated VoD infrastructure to deploy VoD services. [6, 7]

As VoD is moving towards "everything on demand" or "infinite content", scalability becomes a major issue for the success of IPTV. Even a dedicated infrastructure just for VoD services faces such a challenge. People seek successful stories from relevant Internet applications and P2P techniques come into sight.

C. P2P for IPTV

An effective way to release servers' workload for VoD is to use peer-to-peer techniques, in which case we assume each STB can contribute its content or part of storage to its local community. There are advantages utilizing P2P resources in managed IPTV environments comparing to related P2P applications over the Internet. [7]

- STB storage management is available to the IPTV service provider;
- Information of peer locality and upload capacity is also known;
- Conventional P2P policies, such as tit-for-tat, may not be necessary and thus P2P overhead can be greatly reduced;
- Homogeneity of peer devices minimizes the control overhead;
- Distribution servers which are located in the content delivery path can easily implement the tracker functions.

Thus new features of P2P for IPTV can bring more advantages for content distribution than conventional P2P [8]. Distinguishes Internet-based cloud model and IPTV-based physical model and analyzes the capacity and profit of peer-assisted VoD systems. With the assumption that a portion of STB storage is manageable by the service provider, researchers consider STBs as networked storage to push or preload content to peers during off-peak hours or when low network utilization occurs. An example study can be found in [2]. As storage cost continually drops, more storage can be expected from individual STBs and more P2P assisted functions can be implemented based on the networked storage system.

III. IPTV TRANSMISSION

With evolution of networks WiMAX is the best choice for delivering IPTV wirelessly.

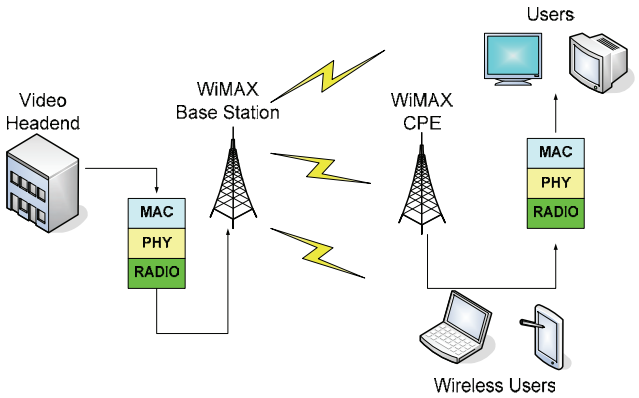


Fig. 2 System Model for IPTV Applications via WiMAX

WiMAX is essentially a next-generation wireless technology that enhances broadband wireless access. WiMAX comes in two varieties, fixed wireless and mobile. The fixed version, known as 802.16d, was designed to be a replacement or supplement for broadband cable access or DSL. A recently ratified version, 802.16e, also can support fixed wireless applications, but it allows for roaming among base stations as well. Thus, the two standards are generally known as fixed WiMAX and mobile WiMAX. The 802.16 standard is beneficial to every link in the broadband wireless chain, such as consumers, operators, and component makers. [9, 10]

In WiMAX, the medium access control layer (MAC) supports a primarily point-to-multipoint architecture, with an optional mesh topology. The MAC is structured to support multiple physical layer (PHY) specifications, each suited to a particular operational environment. [4]

IV. OFDM AND OFDMA

IEEE 802.16 will specify two flavours of OFDM systems: one simply identified as OFDM, the other OFDMA. The first aims at less challenging applications, quite short distance, eventually indoors. It employs fast Fourier transform (FFT)

size 256 — a step further from 802.11a, which uses 64 carriers [2]. All carriers are transmitted at once. The downstream data is time-division multiplexed (TDM). The upstream time frame is time-division multiple access (TDMA).

In OFDMA the higher FFT space (2048 and 4096 carriers) is divided into subchannels. They are used in downstream for separating the data into logical streams. Those streams employ different modulation, coding, and amplitude to address subscribers with different channel characteristics. In upstream the subchannels are used for multiple access. The subscribers are assigned on subchannels through Media Access Protocol (MAP) messages sent downstream. [10]

A. Subchannels

The subchannel is a subset of carriers out of the total set of available carriers. In order to mitigate the frequency selective fading, the carriers of one subchannel are spread along the channel spectrum. Figure 3 depicts the principles of division into subchannels. The usable carrier space is divided into a number of N_G successive groups. Each group contains a number of N_E successive carriers, after excluding the initially assigned pilots. A subchannel has one element from each group allocated through a pseudorandom process based on permutations, so N_G is the number of subchannel elements. For N = 2048, downstream N_G = 48 and N_E = 32, while upstream N_G = 53 and N_E = 32.

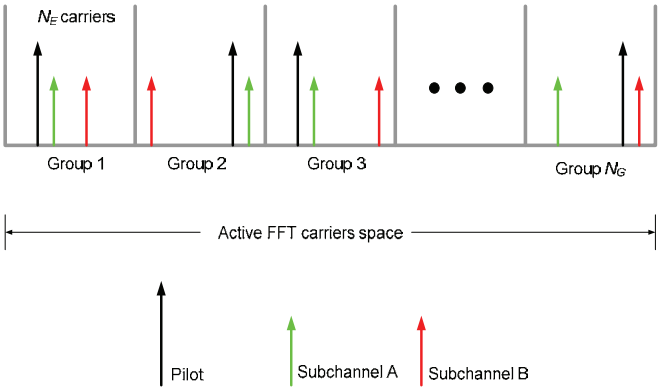


Fig. 3 Division in subchannels

In essence the principle of OFDMA consists of different users sharing the upstream FFT space, while each transmits one or more subchannels. The division in subchannels is a form of frequency-division multiple access (FDMA), where the subscriber transmits $1/N_E$ = 1/32 of the available channel bandwidth for the 2048-carrier OFDMA. A low upstream data rate is consistent with the traffic asymmetry where the streams from each subscriber add up in a multipoint-to-point regime, while downstream all the subchannels are transmitted together. So the OFDMA allows for fine granulation of bandwidth allocation, consistent with the needs of most subscribers, while high consumers of upstream bandwidth are allocated more than one subchannel. Figure 4 shows the structure of subchannels in the upstream framing.

The most important aspect of the upstream subchannels is related to coverage. A BWA system involves a high-power transmitter in the head-end and a multitude of low-cost low-transmission- power BWSUs. For the OFDMA option of N = 2048, the BWSU concentrates its power into a subchannel that has 1/32 of the channel bandwidth. For equivalent modulation and coding, this results in 15 dB premium for the upstream link budget against the downstream. For a 6 MHz channel, one subchannel has an equivalent bandwidth of 187 kHz. But this low bandwidth signal does not undergo flat fading since its 53 carriers are spread across the entire channel bandwidth.

Regarding interference, the subchannels constitute a form of frequency hopping spread spectrum (FHSS). In every group a BWSU transmits one pseudo-randomly selected carrier out of N_E possible ones. A BWSU in an interfering cell does the same type of selection, but statistically independent. The probability of collision is $1/N_E$. This is a classic scenario of FHSS with partial band jammer [9]. The hopping scenario repeats for every group in an FFT symbol. For N = 2048 there are N_G = 53 such groups. The data from carriers with low SNR is corrected through interleaving and coding.

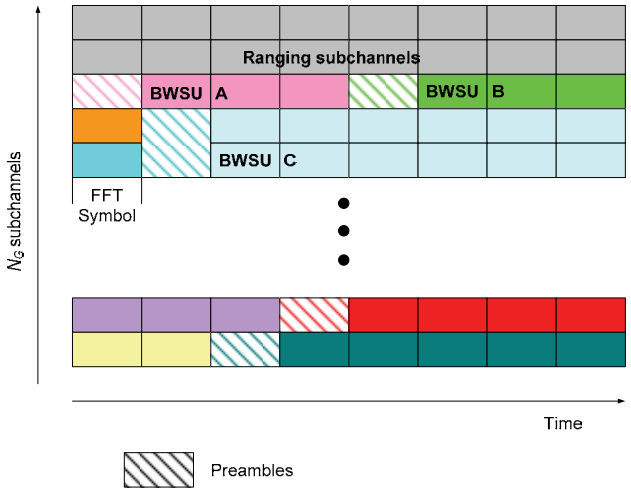


Fig.4 Upstream subchannels

The parameter that characterizes the degree of spreading in a spread-spectrum system is the processing gain, G_P. It can be expressed as a function of W, the group bandwidth R, the bit rate of one carrier and R_S, its symbol rate, by:

$$G_P = \frac{W}{R} = \frac{N_E R_S}{m R_S} = \frac{N_E}{m} = \frac{32}{m}, \qquad (1)$$

where m is the modulation density: 2 for QPSK, 4 for 16-QAM, and 6 for 64-QAM. The processing gain is important in cellular systems because it relates to the interference withstanding of the modulation and coding scheme, or the carrier-to-interference ratio in quasi-error-free operation (C/I), which is the major capacity limiting factor:

$$\left(\frac{C}{I}\right)_{atBER=10^{-6}} = \frac{RE_b}{WN_0} = \frac{1}{G_P}\left(\frac{E_b}{N_0}\right) \qquad (2)$$

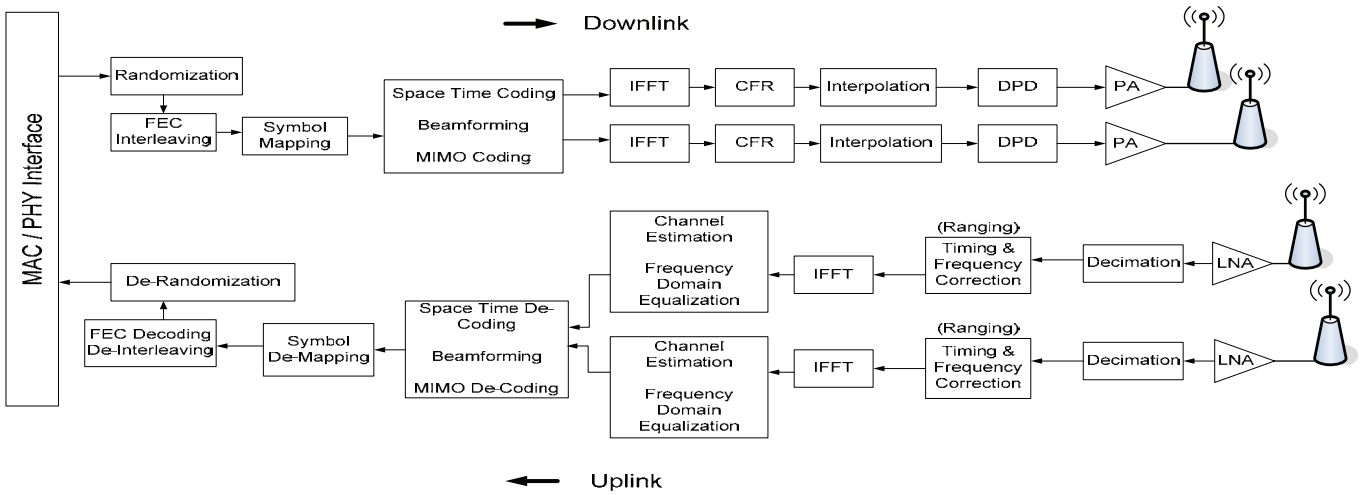


Fig.5 PHY Layer functions in a typical WiMAX base station

There is no upstream interference within the cell since its subchannels are orthogonal: each group element is used by only one subchannel.

Figure 4 illustrates the OFDMA upstream signal space. Some subchannels are reserved for physical layer (PHY) processes such as ranging, while others carry subscriber data according to the MAP allocations. Each transmission starts with the preamble. [3]

V. WiMAX Physical Layer

The WiMAX physical layer is based on orthogonal frequency division multiplexing. OFDM is an elegant and efficient scheme for high data rate transmission in a non-line-of-sight or multipath radio environment.

Apart from the usual functions such as randomization, forward error correction (FEC), interleaving, and mapping to QPSK and QAM symbols, the standard also specifies optional multiple antenna techniques. This includes space time coding (STC), beamforming using adaptive antennas schemes, and multiple input multiple output (MIMO) techniques which achieve higher data rates. The OFDM modulation/demodulation is usually implemented by performing Fast Fourier Transform (FFT) and inverse FFT on the data signal. Although not specified in the standards, other advanced signal processing techniques such as crest factor reduction (CFR) and digital predistortion (DPD) are also usually implemented in the forward path, to improve the efficiency of the power amplifiers used in the base stations. The uplink receive processing functions include time, frequency and power synchronization (ranging), and frequency domain equalization, along with rest of the decoding/demodulation operations necessary to recover the transmitted signal. [1]

VI. Conclusion

The ability to stream video over a broadband IP connection will profoundly change the video industry. With WiMAX offering high data rates to both mobile and fixed users, and the desire of users to watch real-time TV or VoD services make the implementation of IPTV over WiMAX an exciting killer application. Realization of IPTV will enable users to have VoD services as well as to subscribe whichever channel of their choice giving them a great deal of flexibility.

Acknowledgement

This work is a part of the project DMU-02/13-2009 “Design and performance study of an energy-aware multipath routing algorithm for wireless sensor networks” of the Bulgarian Science Fund at the Ministry of Education, Youth and Science.

References

[1] Mihaylov, G., Iliev, T., Hristov, G., “Analysis of the Physical Layer in IEEE 802.16(e) Standard”, ICEST 2010 Proceedings of Papers, Volume 2, pp. 505-508, Bitola, 2010.

[2] Koffman, I., Roman, V., “Broadband Wireless Access Solutions Based on OFDM Access in IEEE 802.16”, IEEE Communications Magazine, April 2002.

[3] M. Chatterjee, S. Sengupta, and S. Ganguly, “Feedback-Based real-time streaming over WiMax,” IEEE Wireless Communications Magazine, February 2007.

[4] Andrews, J., Ghosh, A., Muhamed, R., “Fundamentals of WiMAX”, Prentice Hall, 2007.

[5] Cooper, W., Lovelace, G., “IPTV.Guide”, informitv, 2006.

[6] Jiang, T., Xiang, W., Chen, H., Ni, Q., “Multicast Broadcast Services Support in OFDMA-Based WiMAX Systems”, IEEE Communications Magazine, August 2007.

[7] O’Driscoll, G., “Next Generation IPTV Services and Technologies”, John Wiley & Sons, Inc, 2008.

[8] Hrudey, W., Trajković, L., “Streaming Video Content Over IEEE 802.16/WiMAX Broadband Access”, white paper.

[9] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems, IEEE Std 802.16™-2009 (Revision of IEEE Std 802.16-2004), IEEE, May 2009.

[10] WiMAX Forum, Available online: http://www.wimaxforum.org

Comparative Analysis of the Methods of Defining the Switching Losses in Class D Audio Amplifier – part 1.

Plamen Angelov[1] Dimitar Yudov[2]

Abstract – **The aim of the proposed article is to make a comparative analysis of the switching losses in power stage Class D amplifiers. This method uses specific parameters of the output stage, whose aim is to define the output efficiency. Out of the numerical experiments we can adequately identify the switching losses of the output stage according to the type of the output transistors. The results obtained in this scientific experiment relate to type Dual N-Channel Enhancement Mode Field Effect Transistor AO4916.**

Keywords – **switching losses, switching frequency**

I. INTRODUCTION

In this article several numerical experiments are conducted to determine the output switching losses in the audio output stage class D. The methods used for analysis and evaluation are based on already known computational methods by which we identify the switching losses in the power stage transistors. These losses are defined by changing the load current. In order to maximize the accuracy of the results several studies have been conducted, such as:

- Numerical experiment to identify the switching losses at various values of the load current;
- Numerical experiment to identify the switching losses at various values of the switching frequency;
- Numerical experiment to identify the switching losses during concomitant modification of the two parameters – load current and switching frequency;
- Comparison of the results for each of the two studies according to known methods for assessment.

II. ANALYSIS OF THE METHODS FOR DETERMINING THE SWITCHING LOSSES

A point of particular interest is the simulated comparison research of the switching losses of two of the known methods [1], [2]. With reference to expression.1, we know that the output losses [5] are defined as the sum of the: switching losses + conduction losses + Gate losses.

$$P_{total} = P_{sw} + P_{cond} + P_{gate} \qquad (1)$$

where: P_{sw} - switching losses; P_{cond} - conduction losses; P_{gate} gate losses.

This means that the switching losses are only one component for defining the complex losses. The comprehensive scientific analysis will show that they can be compared using various methods for assessing the switching losses [3], [4], [5]. In order to determine these losses there are two known methods of assessment. These methods take into account parameters such as: power supply, drain current and conduction resistor. The first published method [5] argues that the value of the switching losses is determined by the expression:

$$P_{sw_1} = C_{oss} \cdot U^2{}_{dd} \cdot f_{sw} + I_d \cdot U_{ds} \cdot t_f \cdot f_{sw} \qquad (2)$$

Another expression extends the analysis and defines the same losses but in the form:

$$P_{sw_2} = [0{,}5.I_d \cdot U_{dd} \cdot (t_r + t_f) \cdot f_{sw}] + \ldots$$
$$\ldots + [0{,}5 \cdot C_{oss} \cdot U^2{}_{dd} \cdot f_{sw}] + \ldots \qquad (3)$$
$$\ldots + [K \cdot 0{,}5 \cdot Q_{rr} \cdot U_{dd} \cdot f_{sw}]$$

where: C_{oss} – output switching capacity; U_{dd} – DC power supply; f_{sw} – switching frequency; I_d – maximum drain current; U_{ds} - maximum drain-source voltage; t_r – rise time to switch on the transistor; t_f – fall time to switch off the transistor; Q_{rr}- reverse ratio output charge; K – static coefficient defined by the working temperature and value of the charge Q_{rr}.

Considering the second expression we realize that the size of the charge Q_{rr} depends on the fast diode connected in parallel between the source-drain (D-S) of the switching transistor. This diode is included in reverse direction to the terminals DS which protect the transistors from reverse voltage. Q_{rr} charge is defined for the time of obstruction of the

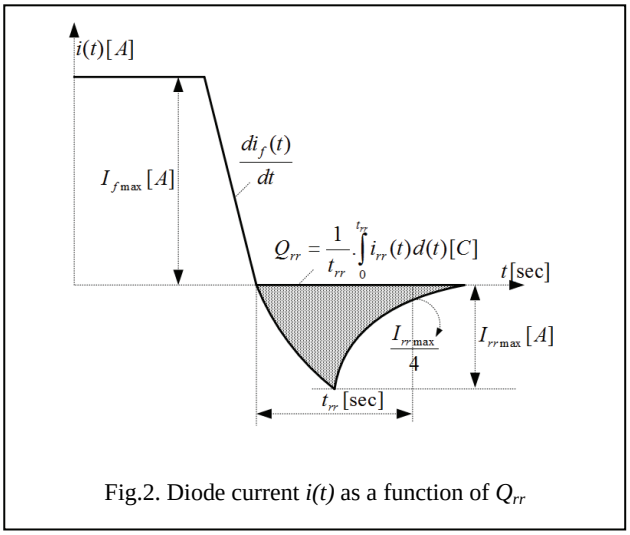


Fig.2. Diode current *i(t)* as a function of Q_{rr}

[1]Assist. Professor Plamen A. Angelov, Faculty of Computer Science and Engineering, Burgras Free University, 62 San Stefano Str., Burrgas-8001, Bulgaria, e-mail: pangelov@bfu.bg
[2] Professor Dr. Dimitar Yudov, Faculty of Computer Science and Engineering, Burgras Free University, 62 San Stefano Str., Burrgas-8001, Bulgaria, e-mail: yudov@bfu.bg

diodes t_r - Figure 2, by the expression:

$$Q_{rr} = \frac{1}{t_{rr}} \cdot \int_0^{t_{rr}} i_{rr}(t) \cdot dt \qquad (4)$$

Besides this charge, in the second expression 2, two additional parameters are introduced: coefficient K and reverse ratio output charge Q_{rr}. It is expected that these two parameters will increase the accuracy to identify the switching losses. Whether this is correct or not will be determined after the experiment defining the switching losses when changing any of the participating parameters. To conduct numerical experiments we need to determine the limits of the supply voltage. The magnitude of this voltage is determined by the expression:

$$U_{dd} = 2 \cdot \frac{\sqrt{2 \cdot P_{out} \cdot R_{load}}}{M} + U_{ds_{sat}} \qquad (5)$$

where: P_{out} – is a output power of the stage; R_{load} – load; $M=2,2.\delta=2,2.(ti/T)$ – duty cycle; $Uds_{sat}=(0,1\text{-}0.3)V$

From Eq.5. the maximum supply voltage can be determined. For this reason the first numerical experiment will determine the output voltage modification with change of the output power at constant load R_{load}.

On the other hand carrying out a numerical experiment requires pre-selection of the switching transistors. For the purposes of the scientific experiment we selected parameters for the Dual N-Channel Enhancement Mode Field Effect Transistor AO4916 with base parameters [6]: V_{ds}=30V; I_{dmax}=8.5A(Vgs=10V) $C_{oss}=190pF$; $Q_{rr}=9.2nQ$.

III. NUMERICAL EXPERIMENTS AND DISCUSSION

A. Numerical experiments to define the maximum power supply

To carry out the numerical simulation it is necessary to define the limits of the voltage supply at a constant frequency. This limitation will follow from the maximum parameters of the selected transistor voltage $V_{ds}=30V$, which will determine the maximum output power. This means that exploring the modification of the output voltage will limit the voltage supply $U_{dd}=30V$. We will make examination by conducting the numerical experiment with the following data: $R_{load}=4\Omega$; : $P_{out}=(1\text{-}30)W$; $\delta=(1\text{-}85)\%$ Applying these values to expression 5 we will get the results shown in Fig.3.

From those numerical experiments the maximum permissible value of the supply voltage $U_{dd}=30V$ is obtained at which the maximum output power is limited to the value of $P_{outmax}=15W$.

The resulting value will be used to limit the subsequent experiments. From the output power we will determine the maximum drain current by the expression:

$$I_{d\max} = \frac{2 \cdot P_{out\max}}{U_{dd}} \qquad (6)$$

Replacement the values of output power and supply voltage we obtained:

$$I_{d\max} = \frac{2 \cdot 15}{30} = 1A$$

The maximum value of the $U_{dd}=30V$, $P_{outmax}=15W$ and drain current $I_{dmax}=1A$ will be used to limit the subsequent experiments.

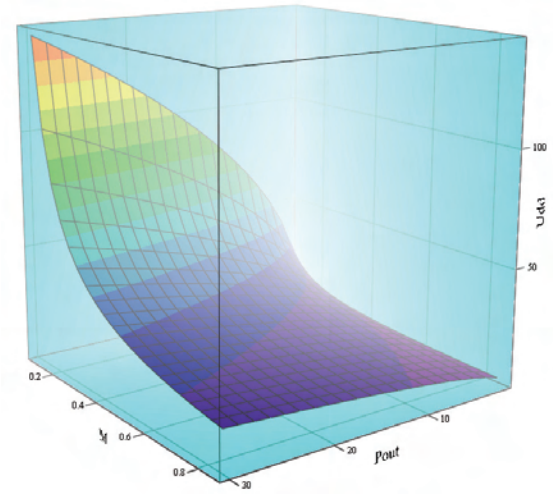


Fig.3. Numerical experiment to define the maximum output power

B. Numerical experiment to define the switching losses with different drain current

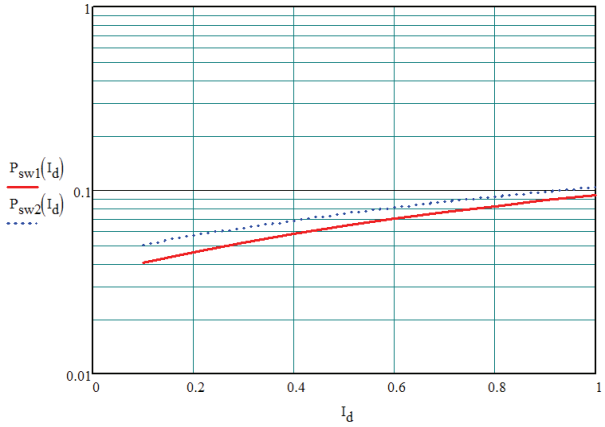


Fig.4. Numerical experiment to define the switching losses P_{sw1} and P_{sw2} with different load current

For the purposes of the scientific experiment, let's keep constant two parameters: the maximum voltage supply $U_{dd}=30V$ and the switching frequency $f_{sw}=200kHz$. Then we will determine the load current modification. The purpose of the program testing is comparison of switching losses P_{sw} on the both known methods (Expression 1 and expression 2) [5] The obtained result is shown in fig.4. In the research area there is a relatively small change of the $P_{sw1.}$ We also notice that the switching losses P_{sw1} remain relatively constant, the equivalent values show a slight discrepancy P_{sw2} resulting from a change of the load current. Additional parameters of the study are $C_{oss}=190pF$; $U_{dd}=30V$; $I_d=0,1\text{-}1A$; $t_r=t_f=10ns$

C. Numerical simulation to identify the switching losses to amend the operating frequency

In this numerical simulation it is necessary to determine the limits of the frequency at the supply voltage. To conduct the scientific experiments we choose the maximum frequency to be limited to f_{sw}=*(40-800)kHz* at a voltage supply U_{dd}=*30V*. The value of this voltage is selected by the restrictive conditions for use of a transistor. The result of the numerical simulation of two expressions (expression 2 and expression 3) to change the working frequency is shown in Fig.5.

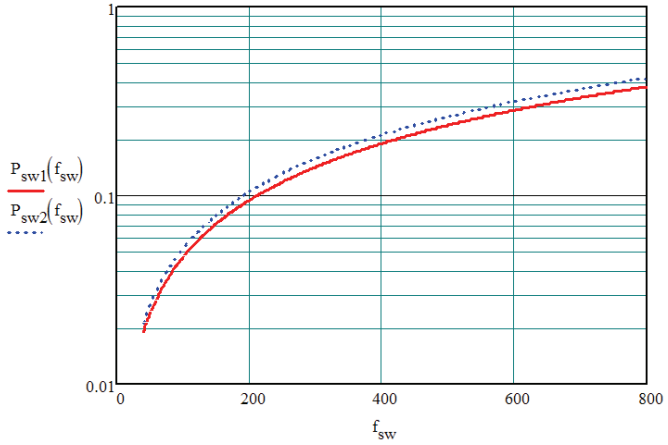


Fig.5. Numerical experiment to define the switching losses P_{sw1} and P_{sw2} with different frequency f_{sw}

From here we can define the following conclusion: While the first numerical experiment is reported in a large discrepancy in the initial values of load current, here inverse relationship is observed. At low frequency both expressions derived approximately the same result. The additional parameters of the study are: C_{oss}=*190pF;* U_{dd}=*30V;* I_d=*1A;* t_r=*10ns;* t_f=*10ns*

Examination also shows that changing the operating frequency significantly affects the growth of the switching losses. To avoid this shortcoming I recommend working frequencies below 400kHz.

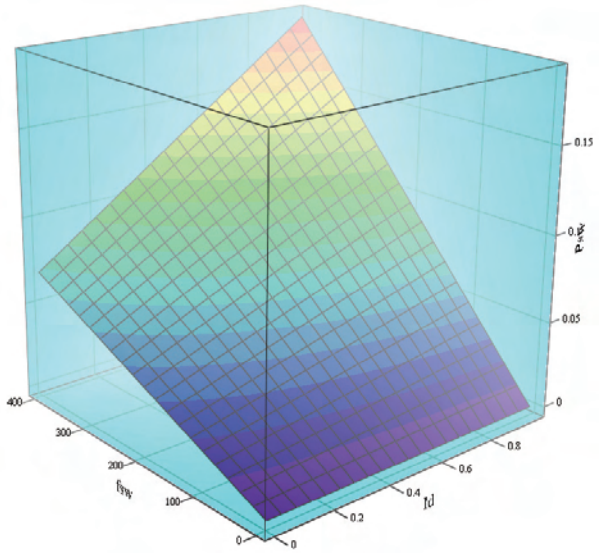


Fig.6. Numerical simulation of the switching losses P_{sw1} with different switching frequency f_{sw} and load current

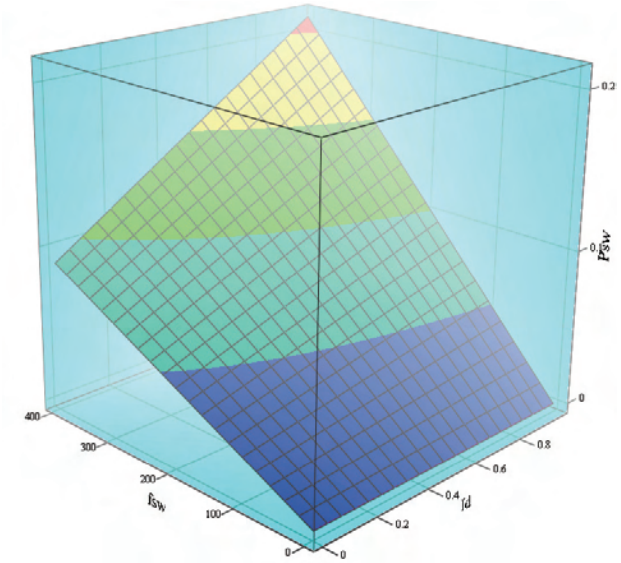


Fig.7. Numerical simulation of the switching losses P_{sw2} with different switching frequency f_{sw} and load current

D. Numerical simulation to identify the switching losses for the simultaneous amendment of the two parameters - Load current and operating frequency

For conducting the research we should comply with the restrictive conditions of a few experiments. This means that the variation in load current will be I_d=*(0,1-1)A*, while the frequency should be amended in the range f_{sw}=*(40-400)kHz*. These results are shown in fig.6. and fig.7 On fig.6 are depicted the switching losses estimated at the first expression, until fig.7 shows the result of expression 2. The result obtained in both studies clearly shows the same variation of switching losses regardless of the chosen method. It is striking low divergence of results in minimum values of output current and maximum operating frequency.

E. Numerical simulation to identify the switching losses by changing the voltage supply

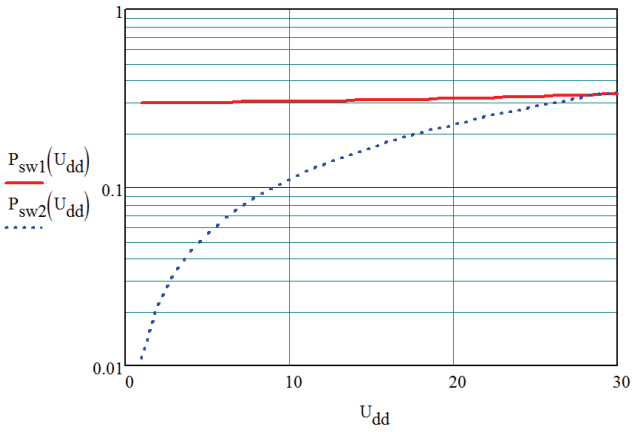


Fig.8. Numerical simulation of the switching losses P_{sw1} and P_{sw2} with parameter the voltage source.

For the purposes of the scientific experiment let us change the voltage supply varying by the value U_{dd}=1-30V. On the other hand the switching frequency is f_{sw}=200kHz. The purpose of the numerical simulation is the comparison of the switching losses P_{sw} on the both familiar mathematical expressions [1], [2].

The programmable result is shown in fig.8. The study shows a significant discrepancy between the results for P_{sw1} and P_{sw2} At low values of this voltage the results for P_{sw2} suggest low switching losses while the value of power around U_{dd}=30V readings are approaching a significant. Additional parameter of the numerical experiment is: C_{oss}=190pF; U_{dd}=30V; I_d=0,1-1A; t_r=t_f=10ns. The conclusion is that within a range of the study values obtained for the switching losses differ significantly. Out of the experiments we observe a point of intersection of characteristics about voltage U_{dd} = 30V, but they are not allowed to be criteria for the assessment.

F. Numerical simulation to identify the switching losses by changing the switching frequency

For the proper implementation of this numerical simulation it is necessary to set limits on the frequency at constant voltage. To conduct scientific experiments the limits of the operating frequency are consistent with practical PCM digital audio systems in the operation range f_{sw}= 40-800kHz In this case the power supply voltage is kept constant U_{dd}=30V.

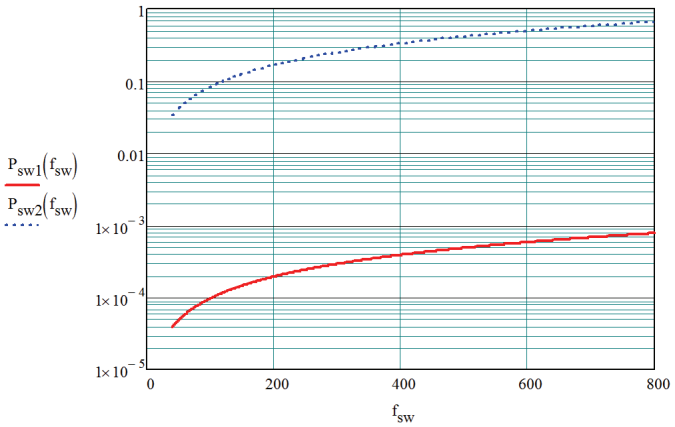


Fig.9. Numerical simulation of the switching losses P_{sw1} and P_{sw2}

The exact value is selected from the data of the first experiment. The result obtained by the numerical simulation of expression 2 and expression 3 with changing the switching frequency is shown in fig.9.

From the result obtained we notice significant discrepancy of the results. From there we can define the following conclusion: The first numerical experiment shows that the reported values for P_{sw2} are lower or equal to P_{sw1}, while the second switching losses P_{sw2} are reported as significantly lower. Additional parameter of the numerical experiment is: C_{oss}=190pF; U_{dd}=30V; I_d=1A; t_r=10ns; t_f=10

IV. Conclusion

- In this numerical simulation it is necessary to determine the limits of the frequency at the supply voltage. In order to conduct the scientific experiments we choose the maximum frequency to be limited to fsw=(40-800)kHz at a voltage supply of Udd=30V;
- While the first numerical experiment is reported in a large discrepancy in the initial values of load current, here we observe inverse relationship. At low frequency both expressions derived approximately the same result.
- There are minimal differences at low currents Id, but they can not be criteria for the assessment of the wall range;
- The outcomes of the research suggest that the results of both methods can be compared over a limited drain current Id> 0,2 A. This current is equal to one fifth of the maximum, which is defined by the moshtnosts output and supply voltage.

References

[1] Eric Gaalaas Class D Audio Amplifiers: What, Why, and How Analog Devices 2007;

[2] Richard.G.Ruehs Motorola High audio power amplifier with short circuit protection Motorola 1985;

[3] National Semiconductor LM4651 & LM4652 Overture™ Audio Power Amplifier 170W Class D Audio Power Amplifier Solution 2000;

[4] Jorge Cerano “Class D Audio Amplifier Performance Relationship to MOSFET Parameters” IRF 2007;

[5] Jun Honda & Jonathan Adams “Class D Audio Amplifier Basics” IRF 2007;

[6] Dual N-Channel Enhancement Mode Field Effect Transistor with Schottky Diode - AO4916 – Alpha & Omega Semiconductor, Ltd.

Method for Paths' Optimization during Path Recovery in MPLS Network

Veneta Aleksieva[1]

Abstract – **MPLS(Multi Protocol Label Switching) is being used in many corporate networks and public infrastructures and as a backbone technology of many Autonomous Systems. Many mission critical applications require better resilience than that provided by the current Internet routing convergence process. During path recovery in MPLS networks large numbers of packets may be dropped. This paper presents a method, which overcomes part of this problem by optimizing paths during path recovery in MPLS network.**

Keywords – **MPLS, LSP, backbone networks**

I. INTRODUCTION

MPLS(Multi Protocol Label Switching) networks are currently evolving towards an universal and convergent network, capable of flowing multiservice traffic as voice, data and video over the same IP based infrastructure. In a real situation in most of MPLS networks is used a physical trace of fiber optic. But sometimes, this fiber cut can cause all the traffic in the fiber to totally interrupt, which is equal to at least tens of Gbps capacity or sometimes even up to hundreds of Gbps capacity. The loss of this huge amount of traffic can bring a significant impact on our economy. Thus, network protection and survivability is of paramount to today's telecommunication networks. This is the main reason for applying of MPLS conception of LSP priorities.

MPLS uses Label Switching Paths (LSPs) priorities. The purpose of them is to mark some traffic as more important than others and allow them to use resources from less important LSPs (pre-empt the less important LSPs). This makes it possible for an important LSP to be established along the most optimal path for this LSP, regardless of existing reservations, if those reservations have a lower priority than this LSP. When LSPs need to reroute, important LSPs have a better chance of finding an alternate path the lower priority LSPs. Best effort traffic that does not need the same treatment in the network, can be mapped to low priority LSPs and higher priority LSP can pre-empt those low priority LSPs if it becomes necessary.

II. RELATED WORKS

The recovery of the MPLS network is based on the algorithm that is applied in order to detect the faults and to route the data flow in an alternative path. For a MPLS based backbone network, the fault-tolerant issue focuses on how to protect the traffic of LSP against node and link failures. In IETF, two well-known recovery mechanisms (protection switching and rerouting) have been proposed, but many researchers create every year some better suggestions, which have different advantages and disadvantages [1].

When an IP packet travel on a MPLS domain, it follows a predetermined path depending on the Forwarding Equivalence Classes (FEC) [2] to which it was assigned by the ingress router. The two main approaches to determine the desired granularity for FEC and determining the paths for the Label Switching Paths (LSP) are:

- **Offline path calculation** - This way of doing path calculations can lead to optimal resource usage, predictable routing and stable network configurations, because determined paths with an off line tool without the LSRs directly participating in the process.
- **Constraint based routing** - Each LSR determines an explicit route for each traffic trunk (aggregation of traffic flows) originating from that LSR based on the bandwidth and the cost of the links and other topology state information [2].

In practice, the traffic engineer will specify the endpoints of a traffic trunk and assign a set of attributes to the trunk about the performance' expectations and behavioral characteristics of the traffic trunk, but there are two main categories of how to set up a LSP:

- **Static LSP**
- **Signaled LSP**.

Static LSP is a LSP that is manually configured via CLI or SNMP. Visiting each LSR and using network management to set the label and interface typically create this kind of LSP.

Dynamic signaling protocols have been designed to allow single routers to request the establishment and label binding to FEC for an end-to-end path. The router that needs to setup an LSP simply determines the best path through the network according to the local constraints and requests the routers in the path to establish a LSP and distribute the label binding to FEC. Configuring a new LSP, over a domain that is MPLS and signaling enabled, does not require anything beyond the configuration in the instantiating router. Signaling is a way in which routers exchange relevant information. In an MPLS network, the type of information exchanged between routers depends on the signaling protocol which is being used. At a base level, labels must be distributed to all MPLS enabled routers that are expected to forward data for a specific FEC and LSPs created. The MPLS architecture does not assume any single signaling protocol [3] and so four methods have been specified for label distribution:

• Label Distribution Protocol (LDP)[4]

[1] Veneta P. Aleksieva is with the Department of Computer Science and Engineering, Technical University of Varna, str."Studentska "1, 9010 Varna, Bulgaria, e-mail: ven7066@abv.bg

• Resource Reservation Protocol extension for MPLS (RSVP-TE)[5,6,7]

• Constrained Routing with LDP (CR-LDP)[8,9,10]

• Distributing labels with BGP-4[11]

Multiple protection routing schemes are possible. To minimize disruption and control overhead, it is used protection routing schemes that change the route from the origin LSP when it traverses a failed link before it fails. Among this class of routing reconfiguration techniques, link-based protection is the most widely used. Thus, link-based protection is good decision, but this scheme can extend to path-based protection, which can be viewed as a special case of link-based protection in an overlay topology.

III. Analysis

To recover a failure, protection of end-to-end connectivity does not need to know where the failure is. Once the two end nodes of the working path detect a network failure, they just perform the switching-over actions. Protection is carried out at the two end nodes of a working LSP. Thus, for a specific link failure, only those unaffected protection LSP can be used to protect the working path. LSP restoration also allows spare capacity sharing among different protection LSPs. The key condition to ensure full failure recovery is that a fiber link should reserve an amount of protection capacity that is maximal among all the link failure situations.

In Fig.1 is presented one example of MPLS network with LSP and its LSP restoration when link failure is occurred. Primary LSP start from LSR1 and follows LSR1-LSR2-LSR3-LSR4, but when link failure between LSR3 and LSR4 arise, packets, which travel on this path, will switch on the backup (alternative) LSP: LSR1-LSR2-LSR3-LSR7-LSR4. This backup path is created before link failure in off-line phase on the protocol, but on this link travel primary traffic.

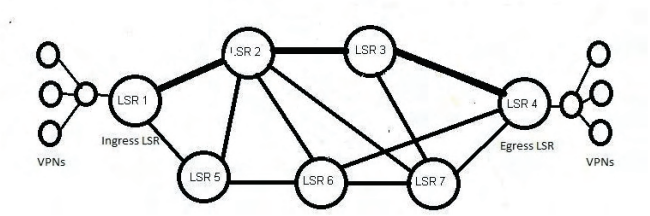


Fig. 1. MPLS network

This means that this link may be responsible for two LSPs – basic path and protection path, and they are sharing protection capacity on their overlaying links. The same links protect more than one LSP. Therefore, the condition of spare capacity sharing is that each time there is only one single network failure and only one of the working paths is recovered, because if more than one link failure arises, it leaves all the traffic on the other working path totally lost.

According to some recent studies on node failure protection with path restoration or shared backup path protection, it is found that a network that supports only single-link failure protection can essentially provide a high percentage, for example, more than 90%, of single-node failure protection without bringing in any extra protection capacity, i.e. it may use the protection capacity that is specially planned for the single-link failure restoration to recover single-node failures and more than 90% node-failure traffic flows can be recovered[13]. Typically, for a mesh network, as a backbone MPLS network, a span restorable network can have a spare capacity efficient around 50-70%, while a shared backup path protection network can achieve spare capacity efficiency around 30-40%[14].

IV. Problem Formulation

In link-based protection, the source node of a failed link reroutes the traffic originally passing through the failed link along a detour route to reach the tail node of the link. Thus, the protection routing only needs to be defined for each link that requires protection; in contrast, the base routing defines routing for each primary LSP.

If as the alternative path is on not empty link and it has own traffic, which uses temporary close to full bandwidth of link, available bandwidth may be not enough for this new repaired traffic. For example, in fig.1 primary path on link LSR3-LSR4 will switch on LSR3-LSR7_LSR4, but on link LSR7-LSR4 traffic exists in the same moment. This traffic, for which this link is the primary path, has higher priority than the new added traffic. This means, that if bandwidth is less than all traffic, some part of rerouting traffic will be lost before repairing of the original link and restoring on this link this traffic.

It is difficult to find optimal routing for alternative link, because optimal routing depends on each interface. For example, if the protocol on layer 3 is IS-IS, this route will be optimal, but if the protocol on layer 3 is OSPF- would not.

This means that for optimizing of network performance and minimizing of packet losses, when MPLS recovery occurs, must be found new algorithm, which will evaluate the behavior of MPLS recovery mechanism.

V. Solution strategy

Based on analysis, the new method for recovery must consist of two phases:

- **Off-line phase** – In this time there are optimized both routing and protection routing in the same time, using original RSVP-TE protocol. The main goal of this stage is to minimize the overload, when failure occurs.
- **On-line phase** – In this time, after failure occurs and traffic is sending on alternative (recovery) path, LSR applies protection routing as fast reroute. This gives to the MPLS network advantage, because if in the moment this alternative link has not enough bandwidth for both traffics, instead of packet losses, these packets will reroute on second alternative link, which is temporary and dynamically created in the moment.

To discover information for MPLS traffic, including Virtual Private LAN Service (VPLS) information, must enable the appropriate agents. They are different for different vendors – Cisco, Huawei, Juniper, Laurel etc. The agents that retrieve

MPLS data use either Telnet or SNMP to retrieve the data. Before enabling the MPLS agents, it must configure Telnet and SNMP access on these devices. Agents that retrieve VPLS information can retrieve large amounts of data. Enabling these agents can add significant processing time to the discovery process.

Basic algorithm of this suggestion is presented below:

When a packet arrives at a router, its next-hop is computed using the network map minus the failed links. If this next-hop would send the packet out an interface that has a failed link, then the router follows next steps for each packet:

1. to remember the failed link
2. to recompute the route using this new failure information
3. to return to step one if the new next-hop also incurs a failure or, if not, forwards the packet to its next- hop

Algorithm for each packet is:

```
Initialization:
        packet.failed links = NULL
Packet Forwarding:
        while (TRUE)
    path = ComputePath(M − packet.failed links)
    if (path == NULL)
        abort("Path is absent")
    elseif (path.next hop == FAILED)
        packet.failed links != path.next hop
    else
        Forward(packet, path.next hop)
    Return
```

Moreover, short explanation of mathematical model of this suggestion is presented below.

There are LSRs $X = \{x_i\}, i = \overline{1,n}$, which are connected with links with bandwidths $D = \{d_j\}, j = \overline{1,k}$ and cost of link $C = \{c_j\}, j = \overline{1,k}$, and $H(k) = \| h_{ij} \| i, j = \overline{1,n}, k = \overline{1,K} - classes, h_{ij}$ – intensity of K-class, which is sent from LSR_i to LSR_j in KBps. Algorithm will found LSP as queue from links $E = \{(r,s)\}$ with throughput $\{\mu_{rs}\}$ and dispersion of flows for all classes $F(k) = [f_{rs}(k)]$ that cover all traffic from each class $H(k)$ and minimize number of packets' losses CLP_k, used minimal cost of network. The main goal is:

$$\min_{E\{\mu_{rs}\}} C_{\Sigma}(M) = \Sigma \sum_{(r,s)\in E} C_{rs}(\{\mu_{rs}\}) \quad (1)$$

And main condition is:

$$CLP(\{\mu_{rs}\};\{f_{rs}\}) \le CLP_k \quad (2)$$

In [15] authors found mathematical expression for CLP_{rs} :

$$CLP_{rs} = P_k = \left[\sum_{k=0}^{n_{rs}} \left(\frac{f_{rs}}{\mu}\right)\frac{1}{k!} + \left(\frac{f_{rs}}{\mu}\right)^{n_{rs}} \frac{1}{n_{rs}} \sum_{K=1}^{N} \left(\frac{f_{rs}}{n_{rs}\mu}\right)^{k}\right]^{-1} \left(\frac{f_{rs}}{\mu}\right)^{n_{rs}} \frac{1}{n_{rs}!} \sum_{K=1}^{N} \left(\frac{f_{rs}}{n_{rs}\mu}\right)^{N_{rs}} \quad (3)$$

Based on (3) average probability of packets' losses in entire MPLS network from K-class is:

$$CLP_k = 1 - \prod_{(r,s)} \left(1 - CLP_{rs}(\mu_{rs}; f_{rs})\right) \quad (4)$$

The main goal of this algorithm is to optimize LSP, in order to minimize packet losses during path recovery process, but it does not reduce time for LSP recovery.

Administrator may affect on the choice of primary LSP, when use bandwidth, priority, administrative weight and attributes and affinity. Configurations on primary and backup paths are presented below:

```
ip rsvp bandwidth <B>                                                    (5)
        where B=75% by default and it is bandwidth of interface
tunnel mpls traffic-eng <s> {H}                                          (6)
        where s=[0;7] setup priority, H=[0;7] holding priority, 0 –high
        priority , by default s=7,H=7
mpls traffic-eng administrative-weight <M>                               (7)
        where M=[0;2^32-1]- metric, which overwrite IGP metric
tunnel mpls traffic-eng path-selection metric {te|igp}                   (8)
        where igp is by default and it is used when channel has delay
mpls traffic-eng attribute-flags <0x0-0xffffffff>{ mask<0x0-0xffffffff>} (9)
tunnel mpls traffic-eng path-option 1 explicit name straight
tunnel mpls traffic-eng oath-option 2 dynamic                            (10)
```

This configuration shows that in LSRs is possible to define proper static path, but when link failure occurs, LSR will find dynamically new one. This will work, if in ingress router is configured fast reroute with command:

```
tunnel mpls traffic-eng fast-reroute                                     (7)
```

But on the protected link:

```
tunnel mpls traffic-eng backup-path <backup-tunnel>                      (8)
```

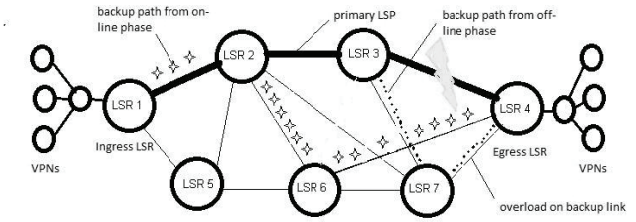


Fig. 2. Example of LSPs in MPLS network

On Fig.2 is presented one example of this algorithm. When link between LSR3 and LSR4 is failed, primary LSP is failed to, and recovery mechanism will switch traffic on backup path. Then it is calculated throughput on basic traffic on this link. For example, if there are two traffics, each of them with average capacity 4.92GB and free bandwidth on reserved link is 10GB, but 4.92GB from this link are already used for high priority traffic. This means, that 4.92GB basic+4.92GB new= 9.84GB, but only 7GB (70% from 10GB) are borrowed about this class of traffic. This is with 2.84GB more than the capacity of the link and they must be rerouted dynamically on different link, because otherwise they would be lost. To overcome this problem, part of the traffic with low priority is

sent back to the LSR2, and LSR2 recalculate new path to the LSR4, without link LSR7-LSR4, where is the problem with overflow. The new dynamic path for this traffic will be LSR1-LSR2-LSR6-LSR4. When the link between LSR3 and LSR4 is repaired, traffic will travel on the primary path LSR1-LSR2-LSR3-LSR4.

VI. Conclusion and Future work

MPLS (Multi protocol Label Switching) is being used in many corporate networks and public infrastructures and as a backbone technology of many Autonomous Systems. This is a connection oriented technology that arises to palliate the problems that current networks have related to speed, scalability and traffic engineering.

A traditional traffic engineering algorithm computes an effective base routing that optimizes a network metric, such as minimizing congestion cost or maximum link utilization. Then, a protection routing is derived from this method, for example, through fast rerouting. While simple and well studied, this traditional approach can easily result in serious network congestion and performance unpredictability under failures.

In this paper first it is formally defined the problem of overflow after the LSP recovery process and then is explained reasons for its challenging. After that it is introduced the key ideas of algorithm, which overcomes packets' losses during the LSP recovery and finally it is given one example of this suggestion.

For the future work this algorithm will be implemented in MPLS module in Network Simulator 2 to be possible to compare with well known recovery schemes in MPLS networks all qualitative and quantitative parameters, because it is important to find in each of them relationship among failure rate of routing paths, the repaired time for finding of protection/alternative path and number of availably alternative paths, number of packets' losses etc.

Acknowledgement

The work presented in this paper was supported within the project BG 051PO001- 3.3.04/13 of the HR Development OP of the European Social Fund 2007-2013

References

[1] Aleksieva V. P., Comparison Studies on Path Recovery Schemes in MPLS Network, ICEST'10, Macedonia, Bitola 2010, vol.2, p.497-500
[2] RFC 3031
[3] L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas
[4] "LDP Specification (RFC 3036)",2001, http://rfc-3036.rfc-list.net/
[5] B. Jamoussi, L. Andersson, R. Callon, R. Dantu, L. Wu, P. Doolan, T. Worster, N. Feldman, A. Fredette, M. Girish, E. Gray, J. Heinanen, T. Kilty, A. Malis "Constraint-Based LSP Setup using LDP (RFC 3212)",2002, http://rfc-3212.rfc-list.net/
[6] L. Andersson, G. Swallow
[7] "The MPLS Working group decision on MPLS signaling protocols"RFC 3468, 2003
[8] L. Berger, Y. Rekhter "Generalized MPLS Signaling – Implementation Survey", draft-ietf-ccamp-gmpls-signaling-survey-preview-3a,2002
[9] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow,"RSVP-TE: Extensions to RSVP for LSP Tunnels (RFC 3209)",http://rfc-3209.rfc-list.net/, 2001
[10] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin,"Resource ReSerVation Protocol (RSVP) (RFC2205)",http://rfc-2205.rfc-list.net/, 1997
[11] D.Awduche, A.Hannan, X.Xiao,"Applicability Statement for Extensions to RSVP for LSP-Tunnels (RFC 3210)", http://www.ietf.org/rfc/rcf3210.txt, 2001
[12] Y. Rekhter, E. Rosen, "Carrying Label Information in BGP-4 (RFC 3107)", http://rfc-3107.rfc-list.net/, 2001
[13] http://www.network-protection.net/node-failure-protection/
[14] http://www.network-protection.net/shared-backup-path-protection-sbpp/
[15] www.foibg.com/ibs_isc/ibs-11/ibs-11-p12.pdf

Simulation Objects in Distributed Environment

Hristo Valchanov [1]

Abstract – **Simulation is a modern approach used for modelling large complex systems and understanding their behaviour. The parallel discrete event simulation (PDES accelerates the modeling process by distributing it among a number of processors. Local area networks are an available platform for PDES. The distributing of the simulation entities over the processors is very important for the performance of the simulation process. This paper presents an approach for mapping the simulation objects over distributed environment - network of workstations.**

Keywords – **PADS, PDES, Mapping, Simulation objects.**

I. INTRODUCTION

PDES accelerates the modeling process by distributing it among a number of processors. With PDES, the modeled system is presented as a set of sub-systems simulated by a number of simulation objects (SO). The SO communicate with one another by exchanging time-stamped messages for occurring events. The simulation correctness requires that the events be processed in the order of their occurrence in time. Special synchronizing protocols are used to ensure the right order of processing [1].

The simulation objects may be implemented as separate independent processes. Such implementation, however, is ineffective from the point of view of the high system overhead on switching of the processes context by the operating system (OS). On the other hand, the communications between the processes in a same computer node is implemented by the OS IPC messages mechanism and has approximate complexity as the inter-computer network communications. By these reason the speed of simulation is largely reduced. Simulation effectiveness improvement can be achieved by aggregating the simulation objects into a cluster. Each cluster will perform as an independent process within a computer node. Its purpose is to carry out scheduling of its simulation objects and ensure communication with the other clusters within the network.

This paper presents an approach for mapping of the SO into clusters over distributed simulation environment based on a network of workstations.

II. RELATED WORK

Distribution of SO over computing nodes is very important for the efficiency of the overall process of simulation. Numerous systems for PDES [3] provide such control of distribution, which require the user to map manually SO to the appropriate physical processors. This approach appears to be inefficient for simulation of models, containing many SO with high intensity of interaction. For such models it is very difficult for the user to determine the interaction between the components of the model, as well as to estimate the optimal configuration of distribution of SOs to the physical processors. There is a need to automate the process of mapping the appropriate SO to the computing nodes.

The selection of a method for distribution of the components of the simulated system into groups (clusters) has an important role for the efficiency of the overall process of simulation. One method is to distribute SO equally into clusters. This method is easy to implement and it is based on the equalization of the computing overhead in separate nodes. However, such mapping does not take into consideration the interaction between SO in the simulated system. This fact has very great influence over the simulation performance in case of simulation of complex dynamical system over distributed environment.

Another method is based on a representation of the simulated system as a graph, which is to be distributed by means of algorithms for graphs partitioning. By this method, the vertices of the graph represent the individual components of the real system, while the edges of the graph represent the interaction between the components. The edges have assigned weights that representing the amount of communication between the components. Applying algorithms for graphs partitioning results in division of the graph into relatively equal parts, thus minimizing the total communication between these parts. The method is also relatively easy to implement, because the problem of graph partitioning is well known in the graphs theory [4].

An important problem for this method is the manner of building of the graph of interaction between SO. A possible solution is to include a specific analysis within the compiler of the simulation language. This method allows generation of the information for the interaction between instances of simulation classes during the stage of compiling. This analysis, however, would be extremely complicated to implement because it is necessary to consider the dynamics of exchange between the components of the simulated system. In case of a dynamic creation of instances of the SO the complexity will increase.

Another solution is based on the *critical path analysis* in the process of simulation [3]. The key concept of this method is that if the graph of the parallel program execution (sequence of events) is known, then the critical path gives the least possible time for execution of simulation. The analyzers, presented in the literature [3], require completion of the whole process of sequential simulation for carrying out the analysis of the critical path (i.e. *post-mortem* analysis). This is applicable to sequential simulation, executed within

[1]Hristo Valchanov is with the Computer Science & Engineering Dept. at Technical university of Varna. 1, Studentska Str., 9010, Varna, Bulgaria, E-mail: hristo@tu-varna.bg.

reasonable period of time. In case of simulation models, involving large number of SO and events such requirement will result in too long execution. At the same time, such analyzers require amounts of memory and disc space proportional to the number of the simulated events and communication operations.

III. THE APPROACH

Our approach to the building of a graph of interaction of SO is based on preliminary sequential simulation, combined with dynamic analysis of the interactions between the components of the simulation model. The approach consists of two phases. During the first phase the graph of interactions between SO is created. The second phase includes partitioning the formed graph into clusters.

A. Formation the graph of interactions

For the formation of the graph of interactions we use an experimental sequential simulator with integrated analyzing component [5]. Its purpose is to analyze the interaction between SO in the preliminary execution of a simulation program. Significant difference between analyses described in the literature and presented method is that it focuses solely on the interaction between SO and not on the sequence of simulation events.

As a result of the first phase the AC builds an interaction graph $G(V,E)$ between SO, where $V=\{v_i\}, i=1..N$ is the set of graph vertices. This set corresponds to the set of the simulation objects $O=\{o_i\}, i=1..N$. The set of events $E=\{e_i\}, i=1..M$ is represented as edges of the graph and each edge introduces the interaction between the SO. Two vertices v_k and v_q are connected with edge e_{kq} if the corresponding SO O_k and O_q exchange messages about happened events during the simulation.

Formation of the graph of interactions $G(V,E)$ may be presented as a two-steps process: during the first step, the communication pairs between SO are formed, and during the second step information about the exchange of messages between them is collected. The grounds in this regard are the following. Upon generating simulation models, the user initially forms on the basis of specific language structures, the logical topology of the connections between SO. After starting the simulation, communication exchange between SO begins, reflecting the scheduling of the simulation events. While the number of events may vary during the whole process of simulation, formation of communication pairs is made yet at the beginning. The primary task is to determine the time T_{CP}, required for the formation of all communication pairs (P_{CP}) during the process of sequential simulation. After formation of P_{CP}, the process of sequential simulation (T_{sim}) continues for a specified period of time - $T_{end}=\lambda * T_{CP}$ during which data about the amount of communications are collected. The value of the parameter λ indirectly determines the accumulation of information exchange between communication pairs.

The detailed explanation of this phase is presented in [5].

B. Partitioning the graph of interactions

Formally the problem of graph partitioning is defined as follows: a graph $G(V,E)$ with set of vertices V and set of edges E is given. Let V be divided into k subsets $V_1,V_2,...,V_k$ such that:

1. $V_i \cap V_j = \Phi, \forall i \neq j$, where Φ is an empty set;
2. $\bigcup_{i=1}^{k} V_i = V$;
3. $|V_i| = \frac{|V|}{k}$;
4. The number of edges connecting vertices from different subsets is minimal.

The conditions 1 and 2 determine the division of the graph with number of vertices $|V|$ into k non-overlapping sub-graphs. The condition 3 determines that the number of vertices in individual sub-graphs must be equal.

For partitioning of the formed graph G the combinatorial multilevel-based method named *Multilevel k-way* is applied [4]. The choice of the algorithm *Multilevel k-way* is based on the following considerations. First, it incorporates the optimization criterion, very appropriate for simulation in distributed computing environment. The criterion is minimization of the general communication exchange between the computing nodes. Compared to other methods of study, it allows precise graph partitioning at comparatively low computing expenses. Secondly, the algorithm is implemented on the basis of the library METIS [4], which is available as an open source and enables the use of API functions in the applications. The execution of the algorithm *Multilevel k-way* on the formed graph generates a map of SO distribution by computing nodes.

Fig. 1 displays the process of distribution of SO between computing nodes.

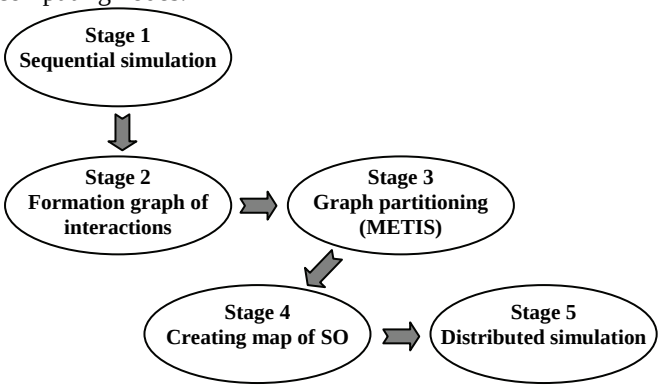


Fig. 1. Process of SO mapping

After completion of the sequential simulation (stage 1), data are generated for the formed graph of interactions $G(V,E)$

and these data serve as input data for the METIS program (stages 2 and 3). The result is a file, containing the map of distribution of SO between the relevant computing nodes (stage 4). Once the map is generated, the distributed simulation may start (stage 5). When a new SO must be created during the simulation, the run-time system uses information from this map. As a result, the SO is created on a particular node.

C. Experimental evaluation and results

Experimental studies has been carried out on the basis of the benchmark test PHOLD [1], [2] using SIMOPAL distributed simulation environment [6], Fig2. This test is widely used for assessment of the distributed simulation performance.

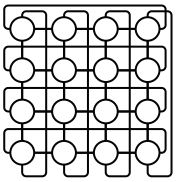


Fig. 2. PHOLD example (N=16)

The test model contains N objects, connected in 2D toroidal network and E events, exchanged between the objects. The dependence has been studied between the number of communication pairs P_{CP} and the number of the processed events during sequential simulation.

Experiments are carried out in three groups. For the first group, the process of sequential simulation is waited until completion. For the second group of experiments simulation is terminated upon achievement of 90% of the total duration T_{sim}. For the third group the proposed method is applied, whereas simulation is terminated upon reaching the time $T_{end} = \lambda * T_{CP}$ (these times correspond to the processing of a specified number of events). Experiments are carried out at different values of λ.

Comparative assessment is made by the following methods - for each computing node is formed a set Ω of SO, which are assigned to it as a result of the complete process of simulation, and the set $\tilde{\Omega}$, containing SO, assigned to it as a result of the proposed method,

$$\bigcup_{j=1}^{n} \Omega_j = \bigcup_{i=1}^{n} \tilde{\Omega}_i = N, \text{ where} \qquad (1)$$

N is the number of all SO in the model, and **n**- the number of computing nodes.

For each set $\tilde{\Omega}_i, i = 1..n$ is determined the maximum ratio $\Psi_i^{\max}, i = 1..n$ (in percentages) of the coincidences of its components with each set $\Omega_j, j = 1..n$. For each group of distribution by nodes is determined the average maximum ratio

$$\overline{\Psi}_i = \frac{\sum_{j=1}^{n} \Psi_j^{\max}}{i} [\%], i = 1..n \qquad (2)$$

The aim is that on the basis of experimental studies λ is determined at which the value $\overline{\Psi}_i, i = 1..n$ will be highest.

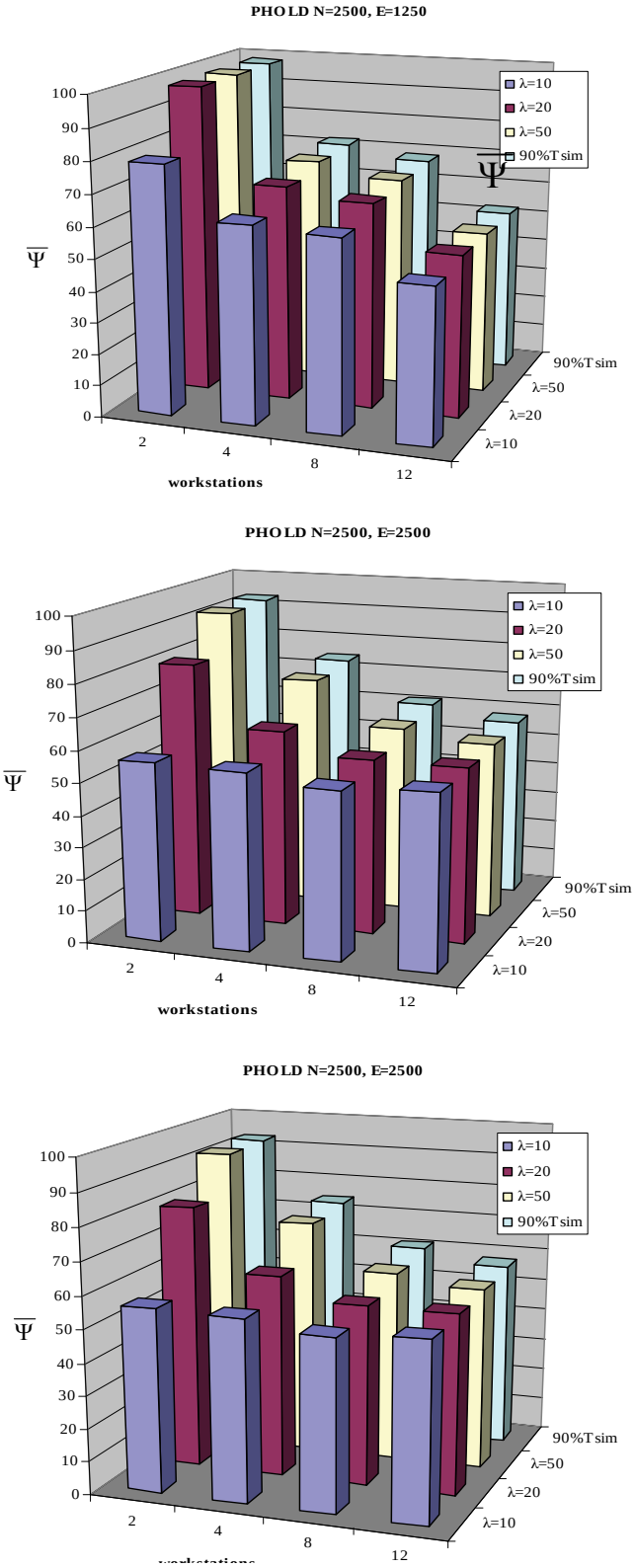


Fig. 3. Obtained results

Fig.3 shows the results from carried out experimental studies. Experiments are carried out by increasing the number of simulated events E. As a number of objects we use N=2500 because with that value the highest performance for network with 12 workstations is achieved. Comparative assessment is made with regard to the results, obtained upon achievement of 90% of the total duration of simulation T_{sim}.

As seen from the charts, with increasing number of simulated events we observed some reduction in the percentage of coincidences. This is due to the fact that a larger number of events needed more time to accumulate information about the communication exchange between communication pairs.

Increase the number of computing nodes **n** also reduces $\overline{\Psi}$. This is a normal consequence of increasing the number of communication channels between communication pairs. It should be noted that this reduction is lower when simulation uses large number of computing nodes. That indicates the correctness of *Multilevel k-way* choice in terms of its resistance to growing the size of the simulation.

As λ increases, it is logical that coincidence of sets $\overline{\Psi}$ increases too, due to the increase of information about communication exchange between simulation objects. Very important result is the level of coincidences upon achievement of 90% of the total time for sequential simulation T_{sim}. As it is evident from the graphics, regardless of the number of computing nodes and the number of simulated events, this proportion is within very close limits to the results in the cases of $\lambda > 20$. On this basis we may assume that choosing λ with values higher than 20, will allow obtaining distribution of SO over the computing nodes which are good enough upon the initial start of distributed simulation

IV. Conclusion

The mapping of simulation objects to processors is extremely important characteristic of systems for PDES to balanced load and inter-processor communication. Most of the above mentioned approaches leave this task to programmers or require completion of the whole process of sequential simulation for carrying out the analysis.

In this paper we proposed a new approach to partitioning of a graph of interaction of SO. Its key concept is based on preliminary sequential simulation, combined with dynamic analysis of the interaction between the components of the simulation model. Significant difference between analyses described in the literature and presented method is that it focuses solely on the interaction between SO and not on the sequence of simulation events The experimental results show that the proposed approach has efficiency for simulation models, characterized by high dynamics of scheduled events between SO.

The purpose of a future work will be to study the period, necessary to prolong the process of sequential simulation after reaching the required limit of communication pairs P_{CP}. Another direction will be a study of possibilities for integration of the proposed approach with dynamic load-balancing of SO between processors during the simulation.

Acknowledgement

The work presented in this paper was supported within the project BG 051PO001-3.3.04/13 of HR Development OP of the European Social Fund 2007-2013.

References

[1] J. Banks, J. S. Carson, B. L. Nelson, *Discrete-Event System Simulation* (5th ed.), Upper Saddle River, Prentice Hall (2009).

[2] R. Fujimoto, "Parallel Discrete Event Simulation", Communications of the ACM, vol. 33, pp. 41-52, 1990.

[3] Z. Juhasz, S. Turner, M. Gerzson. "A Trace-based Performance Prediction Tools for Parallel Discrete Event Simulation", Proc. Applied Informatics, vol. 3, pp. 338-343, 2002.

[4] K. Schloegel, G. Karypis, V. Kumar, "Parallel *Multilevel* Algorithms for Multi-constraint Graph Partitioning", LNCS, vol.1900, pp. 296—310, 2000.

[5] H. Valchanov, N. Ruskova, D. Genov, N. Nikolov, "Partitioning Parallel Discrete Event Simulation", Proc. of CompSysTech08, pp. IIIA.5-1 IIIA.5-6, 2008.

[6] H.Valchanov, N.Ruskova, T.Ruskov. "Distributed Simulation over Network of Workstations", Proc. of CompSysTech 2006, V.Tarnovo, IIIB.25-1, IIIB25-6, 2006.

Reduction of Large Integers by Random Modulus in Public-Key Cryptosystems

Plamen Stoianov[1]

Abstract – **Public-key cryptography is often considered to be too computationally expensive for devices if not accelerated by cryptographic hardware. The most asymmetric cryptographic algorithms used modular operationals $X = A^E \bmod M$ for large integers. These operations determine the data processing speed. The paper presents algorithm for calculating modular reduction without division and multiplication. These operations replaced with rotation and subtraction.**

Keywords – **Public-key cryptosystems, modular reduction, pre-calculation, integer arithmetic.**

I. INTRODUCTION

The need for information security has grown steadily over the years. Users require protection of information from unauthorized access and alteration. Essential tool for achieving these objectives is the use of cryptography. In simplified terms, there are three types of data in encryption technology. The first is plaintext, which is unencrypted data. Encrypted data is referred to as ciphertext. The third is a key, one or more of which is required for encryption and decryption. These tree types of data are processed by an encryption algorithm. Cryptology can be split into two areas of activity, namely cryptography and cryptanalysis. Cryptography is the study of the methods used for encrypting and decrypting data. Goal of cryptanalysis is to develop methods and tools for the revealing of cryptographic systems and evaluate their security.

Modern cryptographic algorithms are generally based on Kerckhoff's principle. This principle says that the entire security of an algorithm should be based only on the on the secrecy of the key, and not on the secrecy of the cryptographic algorithm. The opposite of Kerckhoff's principle is the principle of security by concealment. With this principle, the security of a system is based on the idea that a would-be attacker does not know how the system works. Up to now, every system based on this principle alone has been broken, usually in a very short time[6].

Cryptographic techniques are fundamental to the implementation of security services and may be divided into two classes: symmetric-key and public-key cryptography.

Symmetric-key cryptography requires a single secret key that is used for both encryption and decryption hence the designation 'symmetric'. The exchange of this secret key forms part of the key management problem, that is concerned with the secure distribution of key to the communicating parties. The two types of symmetric-key algorithms are block ciphers and stream ciphers. Block ciphers operate on a block of data while stream ciphers encrypt individual bits. Block chippers are typically used when performing bulk data encryption and the data transfer rate of the connection typically follows the encryption/decryption throughput of the implemented algorithm. The most widely used symmetric cryptographic algorithm (know as Feistel's ciphers) are Triple DES, AES, IDEA etc [11].

A major advance in cryptography came in 1976 with the publication by Diffie and Helman (New Directions of Cryptography) of the concept of public-key cryptography. This new concept that would revolutionize cryptography as it was known at the time. The primary feature is that it removes the need to use a single key for encryption as well as decryption. Pair of matched keys is used, termed 'public' and 'private' keys. The public part of the key pair can be distributed publicly without compromising the security of the private key, which must be kept secret by the receiver. A message encrypted with the public key can only be decrypted with the corresponding private key. The key management problem is greatly simplified by the use of public-key cryptosystems.

Most public-key cryptosystems used today are based on the difficulty of factorizing large integers as well as the difficulty to compute the discrete logarithm of a large integer. The implementation of these public-key cryptosystems requires modular exponentiations.

II. OVERVIEW OF ALGORITHMS FOR MODULAR REDUCTION

The operational speed of public-key cryptosystems is largely determined by the modular exponentiation operation of the form $X = A^E \bmod M$ where X is the remainder, A is the base, E is the exponent and M is the modulus. The modular multiplication operation is accomplished using two steps. It first computes a large-integer multiplication step followed by a modular reduction step. The required modular exponentiation is computed by a series of modular multiplications [8]. The RSA cryptosystem uses modular arithmetic algorithms with large integers in the range of 512 to 2048(more than 600 decimal digits) bits.

The RSA cryptosystem, named after its inventors Rivest, Shamir and Adleman, is the most widely used public-key cryptosystem[4]. Its very simple operating principle is based on the arithmetic of large integers. The two keys are generated from two large prime numbers [10]. The encryption and

[1]Plamen Stoianov is with the Technical University of Varna, Telecommunications Department, Studentska 1 , Varna, Bulgaria, E-mail: pl63@abv.bg.

decryption processes can be expressed mathematically as follows:

encryption: $y = x^e \bmod n$

decryption: $x = y^d \bmod n$

where x = plaintext
y = ciphertext
e = public key
d = private key
n = p.q =public modulus
p,q = secret prime numbers

Before being encoded, the plaintext block must be padded to the appropriate block size, which varies in the RSA algorithm according to the length of the key used. Encryption itself is performed by exponentiation of the plaintext followed by a modulus operation. The result of this process is the ciphertext. This can only be decoded if the private key is known. The decryption process is analogous to the encryption process. The security of the algorithm is based on the difficulty of factoring large numbers. It is quite easy to compute the public modulus from the two prime numbers by multiplication, but it is very difficult to decompose the modulus into its two prime factors, since there is no effective algorithm for this operation. Way to increase the speed of the RSA algorithm is to use the Chinese Remainder Theorem. Prerequisite for using the CRT is that both of the secret prime number p and q are known, which means that it can only be used for decryption [3].

A basic operation in public-key cryptosystems is the modular reduction X=AmodM of large numbers. An efficient implementation of this operation is the key to high performance. In many cases the modulus M is fixed. The fact that M is constant makes it feasible to precompute some values ahead of time which typically results in avoiding divisions and replacing them by multiplications [9].

The Classical, Barrett and Montgomery algorithms are well known modular reduction algorithms for large integers used in public-key cryptosystems. Each algorithm has its own unique characteristics resulting in a specific field of application.

Classical algorithm is a formalization of the ordinary t-n step pencil and paper method, each step of which is the division of a (n+1)-digit number M by the n-digit divisor M, yielding the one-digit quotient Q and n-digit remainder R. Each remainder R is less than M, so that it can be combined with the next digit of the dividend into the (n+1)-digit number Rb+(next digit of dividend) to be used as the new X in the next step[7]. The algorithm is as follows:

Input : A = $\sum_{i=0}^{t-1} a_i b^i$, M = $\sum_{i=0}^{n-1} m_i b^i$

Output : X = $\sum_{i=0}^{n-1} x_i b^i$ = A mod M

1. X $\leftarrow$ A
2. While X $\geq$ M b^{t-n} do X $\leftarrow$ X - b^{t-n}
3. for i = t-1 to n-t+1 step -1 do

if $r_i = m_{n-1}$ then q = b – 1

else q = r_i b+ r_{i-1} div b^{t-n} m_{n-1}

3.2 While q(m_{n-1}b+ m_{n-2}) > $a_i b^2 + a_{i-1} b + a_{i-2}$
do q $\leftarrow$ q – 1

3.3 X $\leftarrow$ X - q Mb^{i-n}

3.4 if X < 0 then X $\leftarrow$ X + Mb^{i-n}

Step 3.2 can be modified to :

q m_{n-2} > ($a_i b + a_{i-1} - q m_{n-1}$)b + a_{i-2} .

Since $a_i b + a_{i-1} - q m_{n-1} < m_n$, this step can be done in two multiplications (plus one comparison of two-digit numbers). Thus this algorithm requires n(n+2) multiplications and n divisions for 2n-bit dividend [8].

P. Montgomery introduced an efficient algorithm for modular multiplication without explicitly carrying out the classical modular reduction step[5]. By representing the residue classes modulo m in a nonstandard way, Montgomery's method replaces a division by m whit a multiplication followed by a division by a power of b. The m-residue with respect to R = b^k of an integer x < m is defined as xR mod m. The Montgomery reduction of x is defined as x R^{-1} mod m, where R^{-1} is the inverse of R modulo and is the inverse operation of the m-reside transformation. It can be shown that the multiplication of two m-residues followed Montgomery reduction is isomorphic to the ordinary modular multiplication. The rationale behind the m-residue transformation is the ability to perform a Montgomery reduction x R^{-1} mod m for 0 $\leq$ x < Rm in almost the same time as a multiplication. If x is the production of two m-residues, the result is the m-residue of the remainder, and the remainder itself is obtained by applying one additional Montgomery reduction. Instead of computing all of t at once, one can compute one digit t at a time, add $t_i m b^i$ to x, and repeat[7]. This change allows the computation of

$m'_0 = -m_0^{-1}$ mod b instead of m'. The algorithm is as follows:

for i=0 ; i < k ; i++ do {
t_i = (X* m'_0) mod b
x = x + $t_i m b^i$ }
x = x div b^k
if (X $\geq$ m) then
x = x - m

Barrett reduction was inspired by fast division algorithm that multiply the reciprocal of the divisor to emulate division. This reduction technique is advantageous in a modular exponentiation where many reductions are performed with the same modulus[1]. It was the first approach to perform reduction without explicitly using the division step in the loop. P.Barrett introduced the idea of estimating the quotient Q = A div M with operations that either are less expensive than a multiprecision division by M [2]. The estimate for Q' of A div M is obtained by replacing the floating-point divisions in

Q = $\lfloor (A/b^{2k-t})(b^{2k}/M)/b^{t} \rfloor$ by integer division

Q' = ((Adiv b^{2k-t}) μ)div b^{t} where $\mu = \left\lfloor \frac{b^{2k}}{M} \right\rfloor$

The number of multiplications and the resulting error is more or less independent of t. The best choice for t, resulting in the least number of operations and the smallest maximal error is t=k+1. The algorithm is follows:

Input : A = $\sum_{i=0}^{2k-1} a_i b^i$, M = $\sum_{i=0}^{k-1} m_i b^i$

Output : X = $\sum_{i=0}^{k-1} x_i b^i$ = A mod M

1. Pre-calculation

1.1 $\mu = \left\lfloor \frac{b^{2k}}{M} \right\rfloor$

2. Calculation of the quotient

2.1 Q $\leftarrow \left\lfloor \frac{A}{b^{k+1}} \right\rfloor \mu$

2.2 Q' $\leftarrow \left\lfloor \frac{Q}{b^{k+1}} \right\rfloor$

3. Compute the remainder

3.1 $R_1 \leftarrow$ A mod b^{k+1}

3.2 $R_2 \leftarrow$ (Q'*M) mod b^{k+1}

3.3 $R \leftarrow R_1 - R_2$

4. Correction of the result

4.1 if R < 0 then R + b^{k+1}

4.2 while R $\geq$ M do R $\leftarrow$ R - M

III. PROPOSED ALGORITHM

For computing X=AmodM without division and multiplication the following algorithm is suggested:

Input : A = $\sum_{i=0}^{t-1} a_i 2^i$, M = $\sum_{i=0}^{n-1} m_i 2^i$, t > n

Output : X = $\sum_{i=0}^{t-n-1} x_i 2^i$ = A mod M

1. Pre-calculation

1.1 X $\leftarrow$ A mod 2^n , S $\leftarrow 2^n$ - M

1.2 While X $\geq$ M do X $\leftarrow$ X – M

1.3 While S > M do S $\leftarrow$ S – M

2. Computation X

2.1 for i = n ; i < t ; i++ do {

2.2 if a_i = 0 then step 3

2.3 X $\leftarrow$ X + S

2.4 if X $\geq$ M then X $\leftarrow$ X – M

3. Correction S

3.1 S $\leftarrow$ S + S

3.2 if S $\geq$ M then S $\leftarrow$ S – M }

4. return (X)

Step 1 involves pre-calculated of X and S. The expression of A may be written in the following way:

A = $\sum_{i=0}^{t-1} a_i 2^i$ mod M = ($\sum_{i=0}^{t-n-1} a_i 2^i * 2^n$ + $\sum_{i=0}^{n-1} a_i 2^i$) mod M =

$\sum_{i=0}^{t-n-1} a_i 2^i$ mod M * S + X

where X = $\sum_{i=0}^{n-1} a_i 2^i$ mod M and S = 2^n mod M

In algorithm RSA t $\leq$ 2n because always A < M^2

In step 2, the current bit a_i is checked and if it is =1 the current value of X is corrected.

Step 3 is related preparation of S for the next cycle i+1

If the checking of i is performed before step 3 , the calculation will be reduced by time for thelast preparation of S.

In base b>2 the following algorithm is suggested:

Input : A = $\sum_{i=0}^{t-1} a_i b^i$, M = $\sum_{i=0}^{n-1} m_i b^i$, t > n

b = 2^k

Output : X = $\sum_{i=0}^{t-n-1} x_i b^i$ = A mod M

1. Pre-calculation

1.1 X $\leftarrow$ A mod b^n , S $\leftarrow b^n$ - M

1.2 While X $\geq$ M do X $\leftarrow$ X – M

1.3 While S > M do S $\leftarrow$ S – M

2. Computation X

2.1 for i = n ; i < t ; i++ do {

2.2 for j = 0 ; j < b ; j++ do {

2.3 if a_{i+j} = 1 then X $\leftarrow$ X + S

2.4 if X $\geq$ M then X $\leftarrow$ X – M

3. Correction S

3.1 S $\leftarrow$ S + S

3.2 if S $\geq$ M then S $\leftarrow$ S – M }}

4. return (X)

When is selected base b > 2 , he number of the external cycles is reduced.

When there is a larger bulk of operation memory it is possible to reduce the operating time. In step 1 is calculated:

$S_k \leftarrow b^{n+k}$ - M for k=0 to b-1

Calculated values for S_k are used to calculate X $\leftarrow$ X + S without multiplication in step 2. In this case, step 3 is outside of the internal cycle

IV. CONCLUSION

In the known algorithms for modular reduction pre-calculations are carried out in order to change the module to b^2 of 2 for faster processing of blocks of data. These calculations involve multiplication and division of large integers. The proposed algorithm uses only elementary operations of rotation, addition and subtraction without division and multiplication.

The algorithm can be employed in applications using microcontrollers with smaller computing capabilities without hardware multipliers. In addition, the efficiency and reliability of the algorithm is higher when processing small amounts of data due to the elementary pre-calculations. Therefore, it can be used to exchange session keys for symmetric algorithms.

REFERENCES

[1]]. W. Hasenplaugh, G. Gaubatz, V. Gopal,"Fast Modular Reduction". IEEE Symposium on Computer Arithmetic pp.225-229, 2007

[2] P. Barrett. "Implementing the Rivest Shamir Adleman public-key encryption algorithm on a standart digital signal processor". Advances in Cryptology – CRYPTO'86, pp.311-323, 1987

[3] T.R. Rao. "Aryabhata Remainder Theorem : Relevance to public-key crypto algorithms". Symposium on Cryptography and Information Security. Japan 2005.

[4] R. Rivest, A. Shamir, L. Adleman. "A method for obtaining for digital signatures and public-key cryptosystems". CACM, vol.21, pp.120-126, 1978

[5] P. Montgomery. "Modular multiplication without trial division". Mathematics of Computation, vol.44, pp.519-521, 1985.

[6] R. Wolfgang, E. Wolfgang, "Smart card handbook" 3rd edition, November 2003.

[7] A. Bosselaers, R. Govaerts and J. Vandewalle, "Comparison of three modular reductions", Advances in Cryptology – Crypto '93 (LNCS 773), Springer-Verlag, pp. 175-186, 1994.

[8] Chia-Long Wu. "Fast modular multi-exponentiation using modified complex arithmetic". Applied Mathematics and Computation, pp.1065-1074, 2007.

[9] N. Gura, A. Patel, A. Wander, H. Eberle, S. Shantz, "Comparing Elliptic Curve Cryptogrraphy and RSA on 8-bit CPUs" Proceedings of CHES'2004. pp.119-132

[10] D. Boneh, H. Shacham, :Fast variants of RSA". Crypto Bytes, vol. 5, No 1, pp. 1-9,2002

[11] T. Wollinger, J. Guajardo, Ch. Paar,:Cryptography in Embedded Systems:An Overview". Proceedings of the Embedded World 2003, Design & Elektronik, Germany, 2003, pp. 735-744

Session PO2:

PO2 - RADIO COMMUNICATIONS, MICROWAVE TECHNIQUE AND ANTENNAS

Comparative Performance Studies of Laboratory WPA IEEE 802.11gPoint-to-Point Links

José A. R. Pacheco de Carvalho[1],Hugo Veiga[2], Nuno Marques[3]
Cláudia F. F. P. Ribeiro Pacheco[4], António D. Reis[5]

Abstract – **Wireless communications using microwaves are increasingly important, e.g. Wi-Fi. Performance is a very important issue, resulting in more reliable and efficient communications. Security is equally very important. Laboratory measurements are made about several performance aspects of Wi-Fi IEEE 802.11gWPApoint-to-point links. A contribution is given to performance evaluation of this technology under WPA encryption, using available wireless routers from Linksys (WRT54GL). Detailed results are presented and discussed, namely at OSI levels 4 and 7, from TCP, UDP and FTP experiments, permitting measurements of TCP throughput, jitter, percentage datagram loss and FTP transfer rate. Comparisons are made to corresponding results obtained for, mainly, open links. Conclusions are drawn about the comparative performance of the links.**

Keywords –**WLAN, Wi-Fi, WPA Point-to-Point Links, IEEE 802.11g, Wireless Network Laboratory Performance.**

I. INTRODUCTION

Wireless communication technologies have been developed using electromagnetic waves in several frequency ranges, propagating in the air. It is the case of e.g. Wi-Fi and FSO, whose importance and utilization have been increasing.

Wi-Fi is a microwave based technology providing for versatility, mobility and favourable prices. The importance and utilization of Wi-Fi has been growing for complementing traditional wired networks. It has been used both in ad hoc mode and in infrastructure mode. In this case an access point, AP, permits communications of Wi-Fi electronic devices with a wired based LAN through a switch/router. In this way a WLAN, based on the AP, is formed. Wi-Fi has reached the personal home, where a WPAN allows personal devices to communicate. Point-to-point and point-to-multipoint configurations are used both indoors and outdoors, requiring specific directional and omnidirectional antennas. Wi-Fi uses microwaves in the 2.4 and 5 GHz frequency bands and IEEE 802.11a, 802.11b, 802.11g and 802.11n standards [1]. As the 2.4 GHz band becomes increasingly used interferences increase. There is a large base of installed equipments working in this band. The 5 GHz band has been receiving considerable attention, although absorption increases and ranges are shorter.

Nominal transfer rates up to 11 (802.11b), 54 Mbps (802.11 a, g) and 600 Mbps (802.11n) are specified. CSMA/CA is the medium access control. Wireless communications, wave propagation [2,3] and practical implementations of WLANs [4] have been studied. Detailed information has been given about the 802.11 architecture, including performance analysis of the effective transfer rate where an optimum factor of 0.42 was presented for 802.11b point-to-point links [5]. Wi-Fi (802.11b) performance measurements are available for crowded indoor environments [6].

Performance evaluation is a fundamentally important criterion to assess the reliability and efficiency of communication. In comparison to traditional applications, new telematic applications are specially sensitive to performances. Requirements have been pointed out, such as: 1-10 ms jitter and 1-10 Mbps throughput for video on demand/moving images; jitter less than 1 ms and 0.1-1 Mbps throughputs for Hi Fi stereo audio [7].

Wi-Fi security is very important. Microwave radio signals travel through the air and can be easily captured by virtually everyone. Therefore, several security methods have been developed to provide authentication such as, by increasing order of security, WEP, WPA and WPA2. WEP was initially intended to provide confidentiality comparable to that of a traditional wired network. A shared key for data encryption is involved. The communicating devices use the same key to encrypt and decrypt radio signals. The CRC32 checksum used in WEP does not provide a great protection. However, in spite of its weaknesses, WEP is still widely used in Wi-Fi communications for security reasons, mainly in point-to-point links. WPA implements the majority of the IEEE 802.11i standard [1]. It includes a MIC, message integrity check, replacing the CRC used in WEP. Either personal or enterprise modes can be used. In this latter case an 802.1x server is required. Both TKIP and AES cipher types are usable and a group key update time interval is specified.

Several performance measurements have been made for 2.4 and 5 GHz Wi-Fi open [8-10] and WEP links [11], as well as very high speed FSO [12]. It is important to find the effects of WPA encryption on link performance. Therefore, in the

[1]José Pacheco de Carvalho is with the Remote Detection Unit and the Physics Departmentat the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: pacheco@ubi.pt.

[2]Hugo Veiga,and [3]Nuno Marques arewith the Remote Detection Unit and the Informatics Centre at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal,E-mails:hveiga@ubi.pt, nmarques@ubi.pt.

[4]Cláudia Pacheco is with the Remote Detection Unit at the University of Beira Interior, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal, E-mail: a17597@ubi.pt.

[5]António Reis is with the Remote Detection Unit and the Physics Department at the University of Beira Interior, and with the Department of Electronics and Telecommunications/Institute of Telecommunications, at the University of Aveiro, 3810 Aveiro, Portugal,E-mail: adreis@ubi.pt.

present work new Wi-Fi (IEEE 802.11 g) results arise, using personal mode WPA, through OSI levels 4 and 7. Performance is evaluated in laboratory measurements of WPA point-to-point links using available equipments. Comparisons are made to corresponding results obtained for, mainly, open links.

The rest of the paper is structured as follows: Chapter II presents the experimental details i.e. the measurement setup and procedure. Results and discussion are presented in Chapter III. Conclusions are drawn in Chapter IV.

II. Experimental Details

The measurements used Linksys WRT54GL wireless routers [13], with a Broadcom BCM5352 chip rev0, internal diversity antennas, firmware DD-WRT v24-sp1-10011 [14] and a 100-Base-TX/10-Base-T Allied Telesis AT-8000S/16 level 2 switch [15]. The wireless mode was set to bridged access point. This was not possible to achieve with the firmware from the manufacturer. In every type of experiment, interference free communication channels were used. This was checked through a portable computer, equipped with a Wi-Fi 802.11 a/b/g adapter, running NetStumbler software [16]. WPA personal encryption was activated in the APs, using AES and a shared key composed of 9 ASCII characters. The experiments were made under far-field conditions. No power levels above 30 mW (15 dBm) were required, as the access points were close.

A laboratory setup was planned and implemented for the measurements, as shown in Fig. 1. At OSI level 4, measurements were made for TCP connections and UDP communications using Iperf software [17], permitting network performance results to be recorded. For a TCP connection, TCP throughput was obtained. For a UDP communication with a given bandwidth parameter, UDP throughput, jitter and percentage loss of datagrams were determined. TCP packets and UDP datagrams of 1470 bytes size were used. A window size of 8 kbytes and a buffer size of the same value were used for TCP and UDP, respectively. One PC, with IP 192.168.0.2 was the Iperf server and the other, with IP 192.168.0.6, was the Iperf client. Jitter, which indicates the smooth mean of differences between consecutive transit times, was continuously computed by the server, as specified by RTP in RFC 1889 [18]. The scheme of Fig. 1 was also used for FTP measurements, where FTP server and client applications were installed in the PCs with IPs 192.168.0.2 and 192.168.0.6, respectively.

The server and client PCs were HP nx9030 and nx9010 portable computers, respectively, running Windows XP. They were configured to maximize the resources allocated to the present work. Batch command files were written to enable the TCP, UDP and FTP tests. The results were obtained in batch mode and written as data files to the client PC disk. Each PC had a second network adapter, to permit remote control from the official IP Unit network, via switch.

III. Results and Discussion

The access points were configured for IEEE 802.11 g with typical nominal transfer rates (6, 9, 12, 18, 24, 36, 48, 54 Mbps). Measurements were made for every fixed transfer rate. In this way, data were obtained for comparison of the laboratory performance of the links, measured namely at OSI levels 1 (physical layer), 4 (transport layer) and 7 (application layer) using the setup of Fig. 1. For every nominal fixed transfer rate, an average TCP throughput was determined from several experiments. This value was used as the bandwidth parameter for every corresponding UDP test, giving average jitter and average percentage datagram loss.

At OSI level 1, noise levels (N, in dBm) and signal to noise ratios (SNR, in dB) were monitored and typical values are shown in Fig. 2 and Fig. 3 for WPA and open links, respectively.

The main average TCP and UDP results are summarized in Table I, both for WPA and open links. In Fig. 4 polynomial fits were made to the 802.11 g TCP throughput data for WPA links, where R^2 is the coefficient of determination. A fairly good agreement was found between the WPA data and the data for open links. Also, both data agree fairly well with those from recent WEP measurements. In Figs. 5-7, the data points representing jitter and percentage datagram loss were joined by smoothed lines. It was found that the best jitter performances are, by descending order, for open links, WEP and WPA. Increasing security encryption was found to degrade jitter performance. Concerning percentage datagram loss data (1.4 % on average) no significant sensitivities were found to link type.

At OSI level 7 we measured FTP transfer rates versus nominal transfer rates configured in the access points for IEEE 802.11 g, as in [11]. The average results thus obtained are summarized in Table I, both for WPA and open links. In Fig. 8 polynomial fits are shown to 802.11 g data for WPA links. The results show the same trends found for TCP throughput.

Generally, except for jitter, the results measured for WPA links were found to agree, within the experimental errors, with corresponding data obtained for WEP and open links.

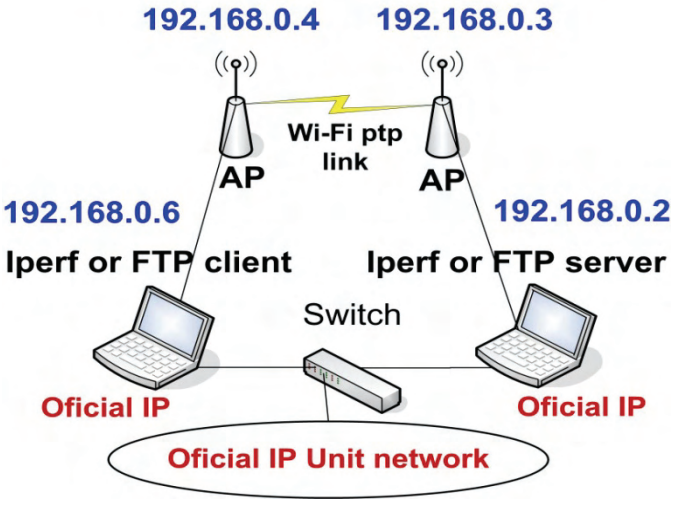


Fig. 1- Laboratory setup scheme.

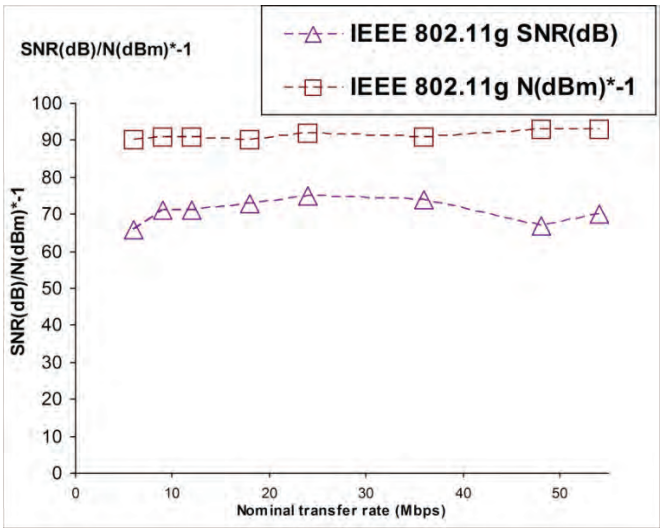


Fig. 2- Typical SNR (dB) and N (dBm); WPA links.

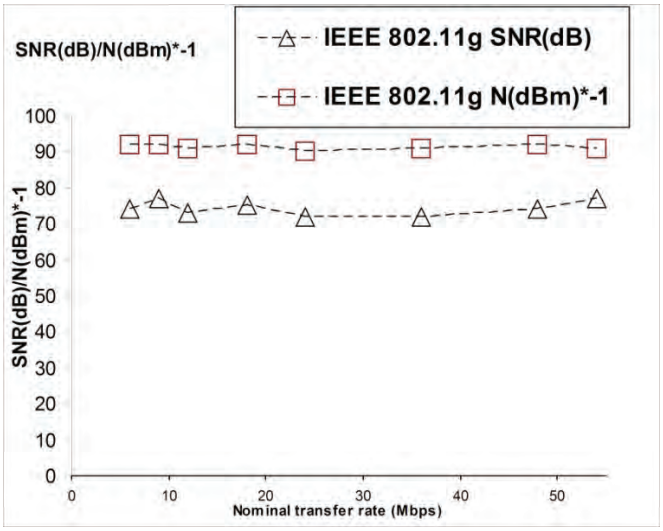


Fig. 3- Typical SNR (dB) and N (dBm); open links.

TABLE I
Average Wi-Fi (IEEE 802.11 g) results; WPA and Open links.

Link type	WPA	Open
TCP throughput (Mbps)	14.1 +-0.4	13.9 +-0.4
UDP-jitter (ms)	2.2 +-0.1	1.2 +-0.1
UDP-% datagram loss	1.2 +-0.1	1.6 +-0.1
FTP transfer rate (kbyte/s)	1527.0 +-45.8	1508.3 +-45.2

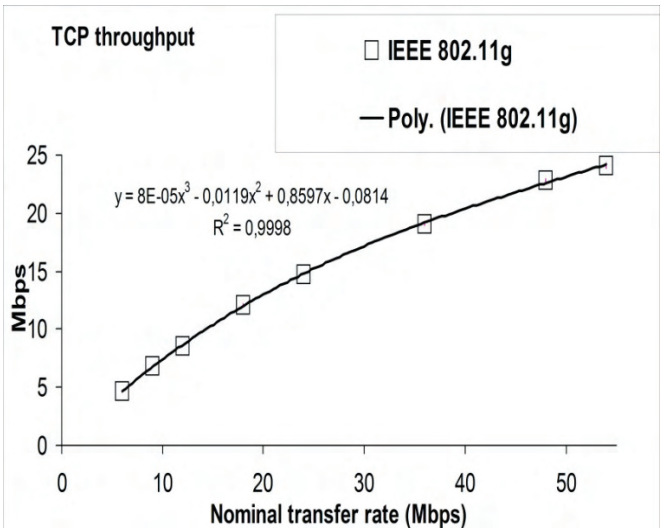


Fig. 4- TCP throughput versus technology and nominal transfer rate; WPA links.

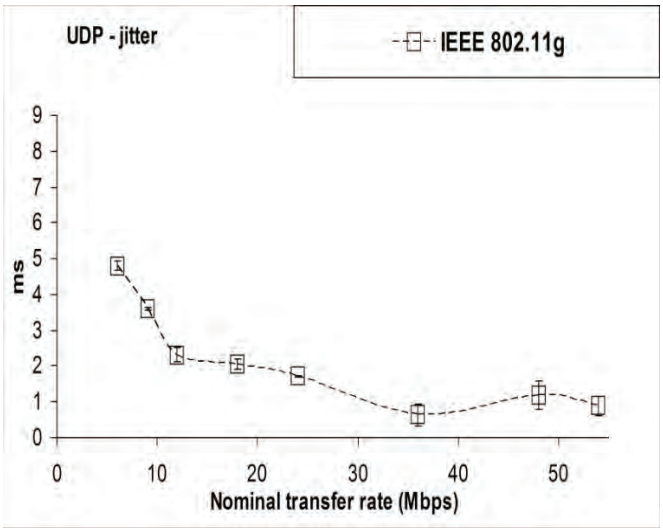


Fig. 5- UDP - jitter results versus technology and nominal transfer rate; WPA links.

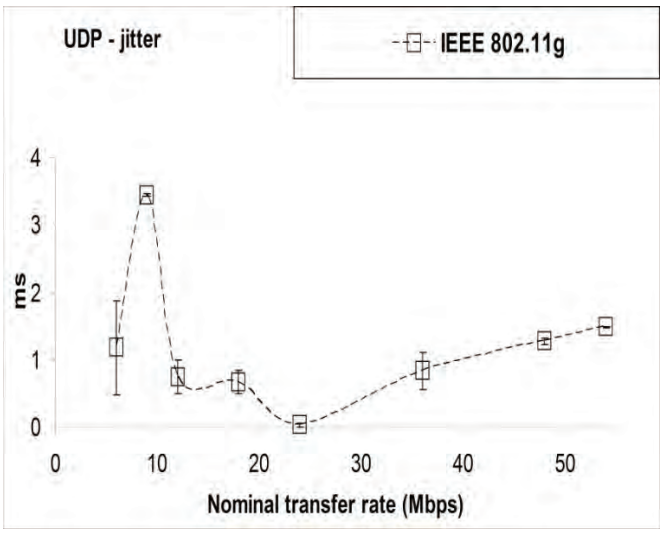


Fig. 6- UDP - jitter results versus technology and nominal transfer rate; open links.

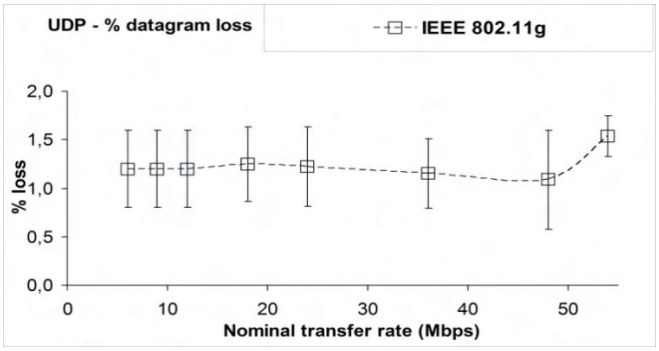


Fig. 7- UDP – percentage datagram loss versus technology and nominal transfer rate; WPA links.

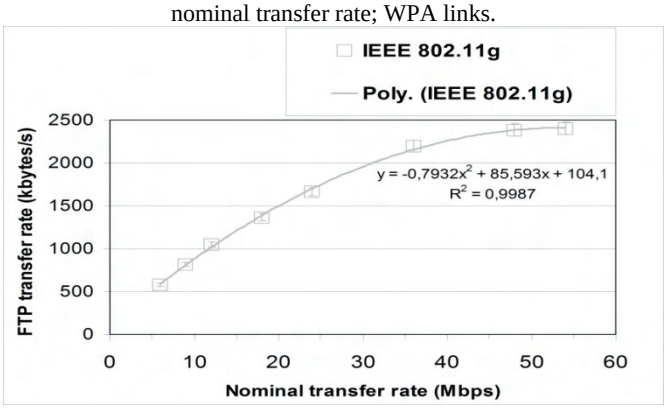


Fig. 8- FTP transfer rate versus technology and nominal transfer rate; WPA links.

IV. CONCLUSION

A laboratory setup arrangement has been planned and implemented, that permitted systematic performance measurements of available wireless equipments (WRT54GL wireless routers from Linksys) for Wi-Fi (IEEE 802.11 g) in WPA point-to-point links.

Through OSI layer 4, TCP throughput, jitter and percentage datagram loss were measured and compared for several link types. The average TCP throughput data were found to agree fairly well for WPA, WEP and open links. Concerning jitter, it was found that the best jitter performances are, by descending order, for open links, WEP and WPA. Increasing security encryption was found to degrade jitter performance. Concerning percentage datagram loss, no significant sensitivities were found to link type.

At OSI layer 7, FTP performance results have shown the same trends found for TCP throughput.

Additional performance measurements either started or are planned using several equipments and experimental conditions, not only in laboratory but also in outdoor environments involving, mainly, medium range links.

ACKNOWLEDGEMENT

Supports from Universidade da Beira Interior and FCT (Fundação para a Ciência e a Tecnologia)/POCI2010 (Programa Operacional Ciência e Inovação) are acknowledged.

REFERENCES

[1] Web site http://standards.ieee.org Web site; IEEE 802.11a, 802.11b, 802.11g, 802.11n, 802.11i standards.

[2] J. W. Mark, W. Zhuang, *Wireless Communications and Networking*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2003.

[3] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed., Prentice-Hall, Inc., Upper Saddle River, NJ, 2002.

[4] W. R. Bruce III, R. Gilster, *Wireless LANs End to End*, Hungry Minds, Inc., NY, 2002.

[5] M. Schwartz, *Mobile Wireless Communications*, CambridgeUniversity Press, 2005.

[6] N. Sarkar, K. Sowerby, "High Performance Measurements in the Crowded Office Environment: a Case Study", In *Proc. ICCT'06-International Conference on Communication Technology*, pp. 1-4, Guilin, China, 27-30 November 2006.

[7] E. Monteiro, F. Boavida, *Engineering of Informatics Networks*, 4th ed., FCA-Editor of Informatics Ld., Lisbon, 2002.

[8] J. A. R. Pacheco de Carvalho, P. A. J. Gomes, H. Veiga, A. D. Reis, "Development of a University Networking Project", in *Encyclopedia of Networked and Virtual Organizations*, Goran D. Putnik, Maria Manuela Cunha, Eds. Hershey, PA (Pennsylvania): IGI Global, pp. 409-422, 2008.

[9] J. A. R. Pacheco de Carvalho, H. Veiga, P. A. J. Gomes, C. F. Ribeiro Pacheco, N. Marques, A. D. Reis, "Wi-Fi Point-to-Point Links- Performance Aspects of IEEE 802.11 a,b,g Laboratory Links", in *Electronic Engineering and Computing Technology, Series: Lecture Notes in Electrical Engineering*, Sio-Iong Ao, Len Gelman, Eds. Netherlands: Springer, 2010, Vol. 60, pp. 507-514.

[10] J. Pacheco de Carvalho, P. Gomes, H. Veiga, C. Pacheco, N. Marques, A. Reis, "Measurements of Performance in Laboratory IEEE 802.11 b, g Point-to-Point Links", *Proc. Measurement 2009 – 7th International Conference on Measurement*, pp. 182-185, Smolenice Castle, Slovakia, May 20-23, 2009.

[11] J. A. R. Pacheco de Carvalho , H. Veiga, Cláudia F. F. P. Ribeiro Pacheco, Nuno Marques, A. D. Reis, "TCP, UDP and FTP Equipment Performances in Laboratory Wi-Fi IEEE 802.11a WEP Point-to-Point Links", *Proc. ICEST 2010 – Proc. XLV International Scientific Conference on Information, Communication and Energy Systems and Technologies*, pp. 153-156, Ohrid, Republic of Macedonia, 23-26 June, 2010.

[12] J. A. R. Pacheco de Carvalho, N. Marques, H. Veiga, C. F. Ribeiro Pacheco, A. D. Reis,"Experimental Performance Evaluation of a Gbps FSO Link: a Case Study", *Proc. WINSYS 2010- International Conference on Wireless Information Networks and Systems*, pp. 123-128, Athens, Greece, 26-28 July , 2010.

[13] Web site http://www.linksys.com; WRT54GL wireless router technical data.

[14] Web site http://www.dd-wrt.com; DD-WRT firmware.

[15] Web site http://www.alliedtelesis.com; AT-8000S/16 level 2 switch technical data.

[16] Web site http://www.netstumbler.com; NetStumbler software.

[17] Web site http://dast.nlanr.net; Iperf software.

[18] Network Working Group, "RFC 1889-RTP: A Transport Protocol for Real Time Applications", http://www.rfc-archive.org

Analysis and Optimizationof Linearly Polarized, Rectangular, Microstrip Line-Fed 3GHz Patch

Nebojša Vojnović[1]

Abstract **– In this paper, the main characteristics of a single microstrip patch antenna – it's efficiency and bandwidth, are defined.A dependence of these characteristics on the antenna physical dimensions is investigated, and some suggestions for antenna construction are listed.Various combinations of dielectric type and height have been used during the analysis, having the working frequency set at3GHz.**

Keywords **– antenna bandwidth, antenna efficiency, optimization, single microstrip patch**

I. INTRODUCTION

Microstrip patch antennasand antenna arrays are widely used in many of the today's applications where such parameters as the antennasize, weight, cost, ease of manufacturing and aerodynamic properties are very important [1]. Microstrip technology implies an unbalanced structure consisting of a dielectric slabon whose one side a thin layer of metallization is printed in a desired shape, while the other side is completely covered with metallization.These antennas are simple and cheap to manufacture using the modern printed circuit technology. They can be mounted on planar and non-planar surfaces, and when the desired shape is set, are very stabile in terms of resonant frequency, polarization, input impedance and radiation pattern. The biggest flaws of single-layered microstrip antennas and antenna arrays are, in fact, low efficiency, high Q-factor, poor scanning capabilities, existence of surface waves and a very narrow bandwidth in the order of a fraction of a percent or a few percent, at best.Patch antennascan be made in various shapes, they can be fed in various ways and made to have different polarizations.

In this paper, rectangular, linearly polarized, microstrip line-fed 3GHz patches are considered, as shown in Fig. 1. A patch element is defined with several dimensions:

L - antenna length,

W - antenna width,

L_{slit} - the length by which the feeding line is inserted into the patch (inset feed), and

W_{slit} - the width of the gap between the feeding line and the patch.

It is known that the antenna length (L) defines it's resonant frequency. With the L_{slit} parameter a patch can be successfully matched to a given characteristic impedance of the feeding transmission line. The input impedance of a patch can also be manipulated with the W parameter, while theL_{slit} parameter remains constant[2]. Breaking the antenna efficiency into it's factors and the analysis of each factor individually, using the method of moments, has shown that each factor peeks at a different patch width [3].Theoretical calculations of the microstrip patch characteristics depending on the surface current distribution [4] and using the method of moments [5], have given results similar tothoseobtained by this analysis.

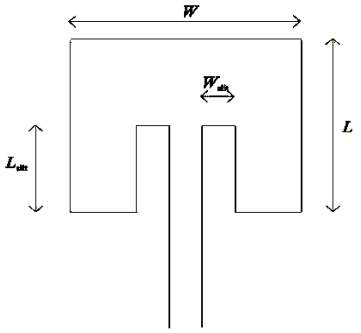


Fig. 1 -Structure considered in this paper

This paper is intended to give a more detailed perspective on the influence of the $W_{slit}$and$W$parameters on the overall antenna characteristics[6]. During the analysis, various combinations of dielectric type and height have been used, having the working frequency set at 3GHz.The program package WIPL-D Pro [7] was used for the simulation.

II. INFLUENCE OF THE W AND W_{SLIT} PARAMETERS ON THE ANTENNA CHARACTERISTICS

A. The parameters analyzed

In the following text the analysis of the influence of the W and W_{slit} parameters on the efficiency, bandwidth, resonant frequency and the needed position of the inset feed, are given.The influences of the$W_{slit}$and$W$parameters on the antenna characteristics were tested separately. The analysiswas done on two types of dielectricsubstrate and for two heights each, using the same working frequency of 3GHz.The Rogers (ε_r=3.38, $tg\delta$=0.0027) andFR4 (ε_r=4.5, $tg\delta$=0.02) dielectric substrates were used. The heights used for the Rogersdielectric were0.508mmand0.813mmand for the FR41mmand1.5mm.The metallization thickness was set to have a constant value of 17μm.Single patch models were constructed for two dielectric types and for two heights each (four sets of parameters, altogether). The length of the patch was initially taken to have the exact same value as the half-

[1]Nebojša Vojnović, Institut IMTEL komunikacijea.d., Bulevar Mihaila Pupina 165b, 11070 Belgrade, Serbia E-mail: nebojsav@insimtel.com

wavelength in the substrate. It should be mentioned that the starting models were chosen to have a square shape, $W=L=W\_nom$, and the W_{slit} parameter was chosen to have the exact same value as the width of the50Ω transmission line at the given frequency, substrate type and height,due to the simplicity of the production. The width of the transmission line was calculated using the impedance calculator which is embedded in the WIPL-D program package.Those starting models were then optimized using the optimizer which is embedded in the WIPL-D Pro program package.In this way, by optimizing the length of the antenna, models operating at 3GHz were obtained. Those models had to be further optimized in order to have the appropriate input impedance. The starting value for theL_{slit} parameter,was taken to be 25% of the antenna length.The models were then experimentally optimized, with the precision of 0.5%, in terms of matching (attaining the lowest possible s_{11} parameter at the given frequency or at some frequency very close to the given one)by changing the position of the inset feed line. It was found that the optimal L_{slit} parameter has a value of about 15÷33% of the antenna length, and depends on the dielectric type used and it's thickness. In this way, the optimal patch elements were attained, and the WandW_{slit} parameters were afterwards altered (in steps of 20%, from 80% to 160%).Besides, the functional dependence of antenna electrical characteristics on these physical parameters was tested and the wanted relation was found.

B. The estimation of bandwidth and efficiency method

The bandwidth for each optimized model was calculated using the -10dBlevel of the s_{11} parameter.The reflection coefficient was found in several close points (at severalclose frequencies).The obtained results were then used for the curve fitting by means of interpolation. The antenna efficiency is defined as the ratio between the power that the antenna radiates and the power that is used for it's feeding. Also, it is possible to define the antenna efficiency using the definition of it's losses. If we assume that it is possible to construct an antenna using only the perfect dielectric and conducting materials, then we can be certain that the constructed antenna will be 100% efficient.Bearing in mind that all the created models have losses, we need to choose one representative parameter and compare it with the same parameter of the referent model (antenna with no losses).In this way we can estimate the antenna efficiency.The parameter used for this comparison was the radiation pattern maximum in units [8].

C. The analysis results

The summary of all of the results of the optimized models, is given in Table I.Those results refer to the starting models ($W=L=W\_nom$,andW_{slit} equals the width of the50Ω transmission line, $W_{slit}=W_{slit}\_nom$) at the frequency 3GHz.

TABLE I

THESTARTING MODEL CHARACTERISTICS AFTER THE OPTIMIZATION

Substrate type	Thickness[mm]	Gain [dB]	Efficiency[%]	BW [%]
Rogers	0.508	4.25	58.00	0.63
Rogers	0.813	5.10	70.68	0.83
FR4	1	0.98	30.59	2.00
FR4	1.5	1.88	40.38	2.33

When two models constructed on the same dielectric having different thicknesses are observed, it can be noted that the models on the thicker dielectric have larger gain, are more efficient and have a wider bandwidth.Comparison of the two models built on different substrates with similar heights (Rogers on 0.813mm and FR4 on 1mm), shows that using the FR4 dielectric lowers the antenna gain and efficiency, but widens it's bandwidth.The widening of the bandwidth when lossier structures are used, can be explained using the Q-factor, which is defined with a relation

$$Q = \frac{f_0}{\Delta f} \tag{1}$$

where the$f_0$stands for the resonant frequency, and$\Delta f$is the bandwidth. Also, we have the following relation

$$Q \sim \frac{energy\ in\ the\ system}{losses\ in\ the\ system} \tag{2}$$

Therefore the bandwidth is proportional to the system losses and the results from the Table I are in order.

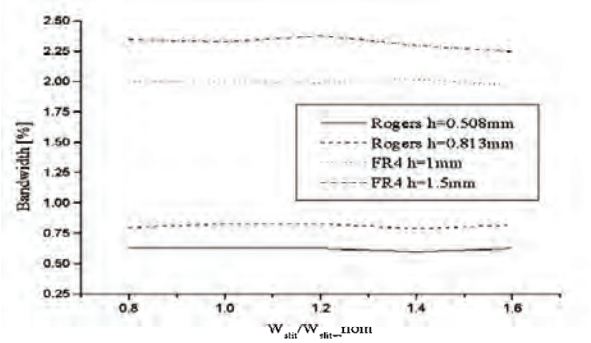
The effects of the W_{slit} parameter on the antenna bandwidth, needed position of the inset feed, efficiency and the resonant frequency are given in the Figs. 2a-2d, respectively.Both types of dielectric with two heights each, were used, having the frequency set at3GHz.

The bandwidth of the antenna is not influenced by theW_{slit} parameter (Fig. 2a).It is noticeable that using thicker substrates or the lossier ones results in wider bandwidths, which was expected.

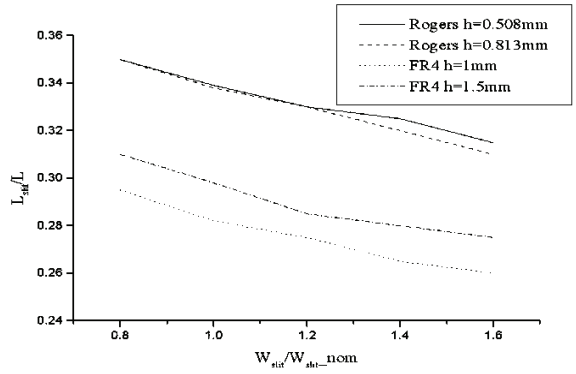
The change of the W_{slit}parameter, degrades the antenna matching, and reoptimization of the position of the inset feed is needed in order to match the antenna input impedance to the characteristic impedance of the 50Ω transmission line.The dependence of the new inset feed depth is, in general, linear and with equal slope in all considered cases (Fig. 2b).In the case of the Rogers dielectric, the substrate thickness doesn't affect the depth of the inset feed in a significant way.On theFR4 dielectric, lesser inset feed depth is needed compared to the Rogers case.

Antenna efficiency isn't affected by the change of theW_{slit} parameter.Higher efficiency can be obtained by using thicker dielectric slab or by using the substrate with lower losses(Fig. 2c).

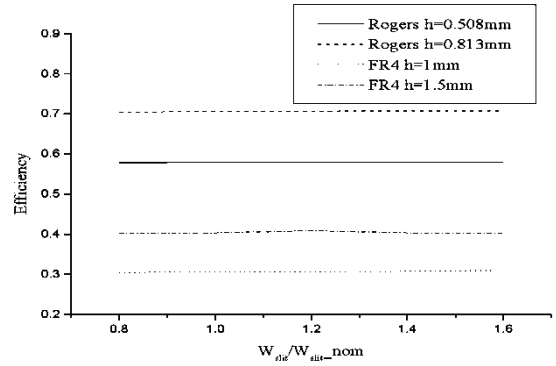
The increase of theW_{slit}parameter slightly increases the antenna resonant frequency (Fig. 2d). The slopes of the resulting curves aren't the same in all the cases. Increasing the dielectric thickness and higher substrate losses both result in raising the curve slope.



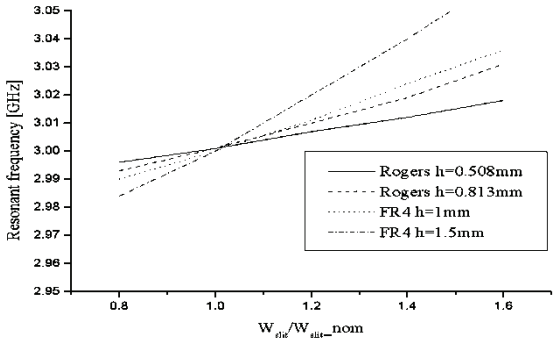
(a)



(b)



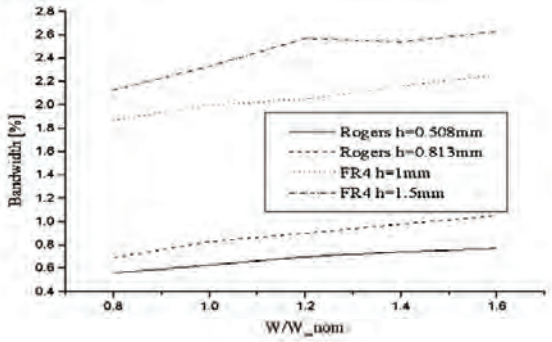
(c)



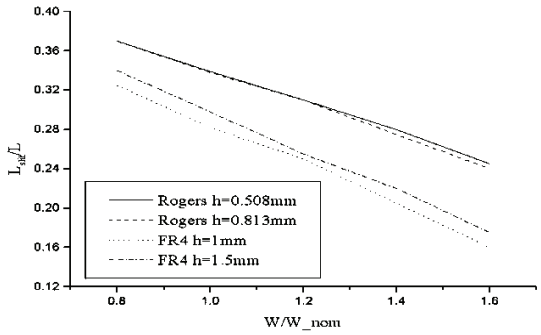
(d)

Fig. 2 -The influence of the W_{slit}parameter on the bandwidth (a), needed position of the inset feed (b), efficiency (c) andresonant frequencyof the antenna (d)

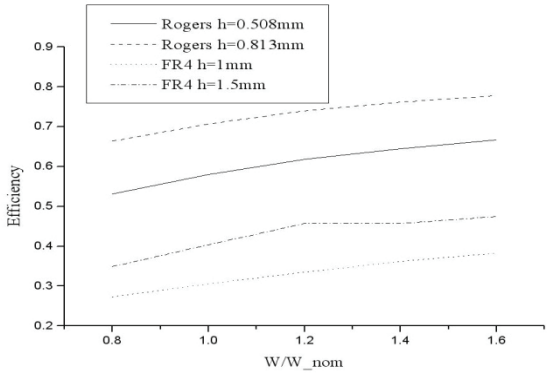
The effects of the *W* parameter on the antenna bandwidth, needed position of the inset feed, efficiency and the resonant frequency are given in the Figs. 3a-3d, respectively.Both types of dielectric with two heights each, were used, having the frequency set at 3GHz.



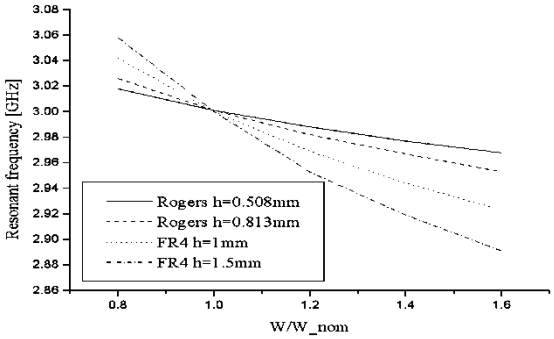
(a)



(b)



(c)



(d)

Fig. 3 - The influence of the *W* parameter on the bandwidth (a), needed position of the inset feed (b), efficiency (c) and resonant frequency of the antenna (d)

The bandwidth of the patch raises along with it's width (Fig. 3a). The slopes of the resulting curves are approximately the same, which means that the relative change of the bandwidth isn't affected by the substrate type. The change of the FR4 slab thickness from 1mm to 1.5mm results in a larger change in bandwidth than in the case of the Rogers substrate and the 0.508mm and 0.813mm thicknesses. On the Rogers dielectric, with the thickness of 0.508mm, the increase of the *W*parameter of 20%,results in increasing the bandwidth from 0.03% to 0.07%. For the slab thickness of 0.813mm,the range of bandwidth increase is from 0.07% to 0.14%. If we look at the FR4 , 20% increase in the antenna width results in an increase of bandwidth from 0.09% to 0.24% for the dielectric thickness of 1mm, and the bandwidth increases from 0.05% to 0.13% when a 1.5mm thickness is used.

As it was the case with the W_{slit} parameter, changing the antenna width also degrades the patch matching. A reoptimization of the depth of the inset feed is needed in order to match the input impedance of the patch to the characteristic impedance of the 50Ω transmission line. With the increase of the width of the patch, the needed depth of the feed position decreases approximately linearly (Fig. 3b). On the Rogers dielectric, the influence of the substrate thickness on the position of the inset feed is lesser than in the FR4 case. When using the FR4 dielectric, smaller depths of the inset feed are needed than in the Rogers case.

The patch width affects it's efficiency in such a way that it increases approximately linearly along with the antenna width (Fig. 3c).The Rogers dielectric is much more efficient than the FR4, which was expected due to lower dielectric loss tangent. Raising the patch efficiency can also be done by increasing the dielectric height.When using the FR4 substrate, change in dielectric thickness from 1mm to 1.5mm results in a lesser change in efficiency than in the Rogers case, using the thicknesses of 0.508mm and 0.813mm. On the Rogers dielectric 0.508mm thick, increase of 20% of the*W*parameter, increases the bandwidth from1.8% to 4.8%, andfrom 1.6% to 4.3% when using the 0.813mm thickness. If we look at the FR4 dielectric, 20% increase of the antenna width results in a bandwidth increase of 0% to 5.5% on the 1mm thickness, andfrom2.1% to 3.3% on the 1.5mm thickness.

With increasing the *W*parameter, lowering the antenna resonant frequency is achieved (Fig. 3d).The slope of the resulting curve depends on the type and height of the dielectric used.Namely, higher substrate thickness results in higher curve slope. Also, using the higher losses dielectric leads to a greater curve slope.On the Rogers dielectric0.508mm thick,that slope is around -0.06, and on the 0.813mm thick substrate the same slope is around -0.09. On the FR4 dielectric 1mm thick, the slope is -0.15, and on the1.5mmthick dielectricthe slope is -0.21.

III. Conclusion

In this paper, the method for printed antenna modeling is given, which allows fast and efficient overview of change of relevant antenna parameters depending on the antenna geometry and characteristics of the dielectric material used.This is necessary due to a constant need for changing the antenna shape and it's place of installation in various devices, and to avoid the situation where the antenna dictates the shape of the final user product.In this way, the device designers have complete control, because it is always possible to adjust the antenna shape and position of it's installation to the shape of the device.For these kinds of applications, a fast and reliable way of printed antenna modeling, is needed.

The influences of the*W*andW_{slit}parameters on the rectangular, single-layered, linearly polarized, microstrip line-fed 3GHz patch antennas, are observed.

Future work should include the analysis of microstrip patch antennaoperating on the second resonant frequency, on it's bandwidth. This concept is still not enough investigated, and the published papers on similar subjects indicatethat this kind of analysis could yield good results.

Acknowledgement

The author would like to thank prof. Branko Kolundzija from the School of Electrical Engineering in Belgrade, for his aid in writing this paper.

References

[1] C.A. Balanis, *Antenna Theory - Analysis and Design*, New York,John Wiley & Sons, Inc., 1997.

[2] M. T. Chryssomallis and C. G. Christodoulou, "Methods of Controlling the Input Impedance of a Microstrip Patch Antenna", Antennas and Propagation Society International Symposium, vol. 4, pp.2470-2473,Orlando, USA, 1999.

[3] S. Yong, J. T. Bernhard, “Investigation Into the Relationship Between Microstrip Patch Width and Efficiency”, IEEE Antennas and Wireless Propagation Letters, vol. 8, pp.1171-1174, 2009.

[4] P. Perlmutter, S. Shtrikman, and D. Treves, “Electric Surface Current Model for the Analysis of Microstrip Antennas with Application to Rectangular Elements“, IEEE Trans., Antennas Propag, vol. AP-33, no. 3, pp. 301-311, Mar. 1985.

[5] D. M. Pozar, “Considerations for Millimeter Wave Printed Antennas“, IEEE Trans., Antennas Propag, vol. AP-31, no. 5, pp. 740-747, Sep. 1983.

[6] N.Vojnović, *Analysisof microstrip antenna planar arrays* - graduate paper, Belgrade, 2008.

[7] WIPL-D Pro - 3D EM Simulator v6.4, Belgrade: WIPL-D d.o.o., 2007.

[8] D. M. Pozar and B. Kaufman, “Comparison of Three Methods for the Measurement of Printed Antenna Efficiency“, IEEE Trans., Antennas Propag, vol. 36, no. 1, pp. 136-139, Jan. 1988.

Upstream Design Considerations in HFC/CATV Systems

Oleg Borisov Panagiev[1]

***Abstract* – In this paper are proposed circuit solutions and methodology to balance the Upstream channel, as well as expressions for determining the output signals of cable modems, the total noise in the channel and the signal to noise ratio. The results are shown in block diagrams, graphs and tables.**

***Keywords* – HFC/CATV, cable modems, C/N.**

I. Introduction

Today a typical HFC/CATV network allocates a large amount of capacity to Downstream transmission with the spectrum ranging between 47(85)–862 MHz. With most of the spectrum is dedicated to TV channels (analog, digital, VoD and maybe HDTV), only a small amount of unallocated bandwidth remains [1].

The Upstream spectrum, too, is limited - available bandwidth is 5–30(65) MHz, out of which the lower frequencies cannot be used due to noise. This bandwidth is required to service network monitoring, interactive TV (digital), VoD, cable telephony and cable modems [2], [3].

II. Structural Scheme and Mathematical Relationships

Subject of the present development are the design considerations for the Upstream (return path) in HFC/CATV networks and in particular the part of the fiber node (FN) to the subscribers' equipments (Fig.1). The cable amplifiers (CA) are two-way and some of them are with active, another with passive return channel. This is defined out of the results received for the parameters of transmitted signals by sizing the return path. The passive Taps and Splitters have a different attenuations and a number of tap/split outputs. In supertrunk lines are used single optical fibers, in trunk lines – trunk coaxial cables, in subtrunk lines - drop (distribution) coaxial cables from the F11/RG11 series and in subscriber lines – drop (distribution) coaxial cables from the F6/RG6 series. Subscriber equipments are: TV receiver, cable modem, phone/Fax and set-top-box. For noise reduction in return channel between the respective output of the passive devices (directional coupler or splitter) and the subscriber equipments is installed a Return Step Attenuator (RSA). RSA suppresses the signals in return path (including the noise).

The value of RSA is defined according to the level of output signal from the cable modem, the distance to the home amplifier and tap loss of the passive device to the subscriber. This way the signals in return path for every subscriber, served from this amplifier, are evened. If there are subscribers, which do not use services from the return path, instead of RSA is installed a high frequency filter (HPF) with a frequency band 85-862MHz. This reduces the penetration of noises and ingress of the subscribers' equipments, which do not use interactive services [4].

All reviews and analysis are made for a frequency band of return path from 5-65MHz. The proposed mathematical expressions are universal for both cable distribution networks, built in the apartment buildings and between family houses.

A. Calculation of the output level of cable modem

To be reduced the noise and ingress in return path is

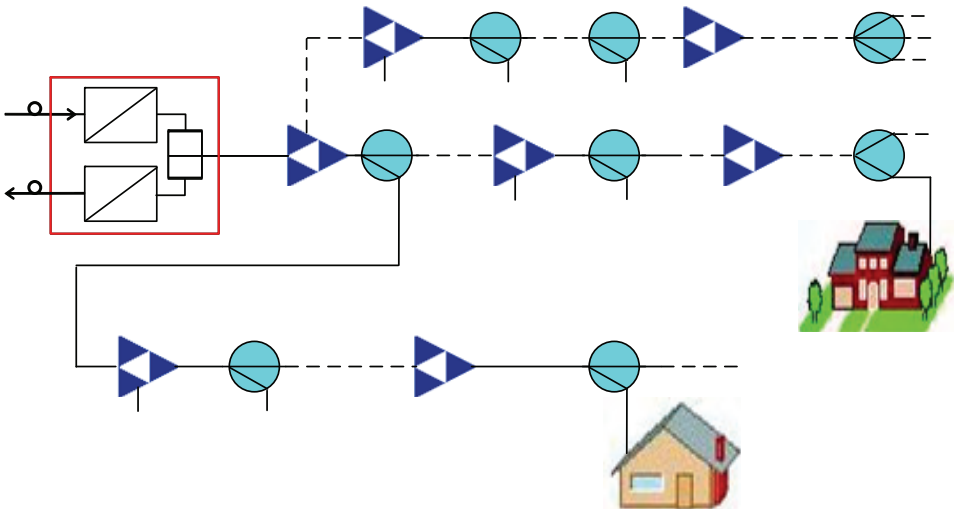


Fig.1. HFC/CATV cable distribution network

[1]Oleg B. Panagiev is with the Technical University of Sofia, Bulgaria, E-mail: olcomol@yahoo.com.

necessary the level of signal at the input points of the amplifiers and the fiber node to be equal (80dBμV) – points 1 to 12 from Fig.1. For this purpose are calculated the cable modems' output levels U_{CM} in a way that, by indicating every source of attenuation in the line between subscriber and amplifier/fiber node, on its input (for return path) will be received the necessary level 80dBμV. The common formula for making the calculations is:

$$U_{CM,k} = U_{in,i} + \sum_{i=1}^{M} G_{Up,i} - U_{in,FN} - \sum_{j=1}^{J} a_{PD,j} - \ell_{TK}.a_{TK} - \ell_{STK}.a_{STK} - \ell_{SSK}.a_{SSK}, \quad (1)$$

where $U_{CM,k}$ is the output level of k-th cable modem (k=1, 2, 3, 4, 5 …) in dBμV; $U_{in,i}$ is the input level of i-th cable amplifier mounted nearest to k-th cable modem, [dBμV]; $U_{in,FN}$ is the input level of fiber node, [dBμV]; $G_{Up,i}$ is the gain of upstream path components in i-th cable amplifier, [dB]; $a_{PD,j}$ - attenuation of passive devices in the line between cable modem and fiber node (tap and through loss), j is the number of passive component, [dB]; ℓ_{TK}-full length of trunk coaxial cable [m]; ℓ_{STK}- full length of subtrunk coaxial cable [m]; ℓ_{SSK}- full length of subscriber coaxial cable [m]; a_{TK} – attenuation of trunk coaxial cable for 1m at 65MHz, [dB/m]; a_{STK} – attenuation of subtrunk coaxial cable for 1m at 65MHz, [dB/m]; a_{SSK} – attenuation of subscriber coaxial cable for 1m at 65MHz, [dB/m].

On Fig.1 with a dotted line is marked the possibility of existence of other amplifiers and passive devices, as 1, 2, 3…12 are input points of the amplifiers and fiber node.

B. Calculation of the aggregate noise

Main passive devices in the cable distributive network, which are used for tap/splitting of the signals transmitted in downstream (forward) path, are directional couplers (Taps) and Splitters (Fig.1). Through them is realized also the transmitting of signals in upstream (return) path, as in this case they take the functions of sumators. There is a correlation between the Tap/Splitter value and the susceptibility of a subscriber location introducing noise and ingress to the network. Conceptually it can be illustrated by focusing on the isolation between the subscriber and the network provided by the Tap/Splitter. For example: the 4 dB Tap/Splitter provides only 4 dB of isolation between the subscriber and the network - whereas the 30dB Tap provides 30dB of isolation. The smaller value Tap will inherently allow more noise and ingress into the network.

On Fig.2 are given the graphical symbols of Tap and Splitter, as well the ports, their parameters and the levels of signals and noises. Based on those, below are presented mathematical relationships for calculating of the aggregate noise at every point of the cable distribution network.

1. Calculation of the signal at the output of the passive component:

a) when the signal at the output of the Tap comes from the direct line:

$$C_{0,out}^{(T)} = C_{0,in} - a_{thru}, \text{ [dBμV]}; \quad (2)$$

b) when the signal at the output of the Tap comes from the side/drop/tap lines (i=1÷ 4):

$$C_{i,out}^{(T)} = C_{tap,i} - a_{tap,i}, \text{ [dBμV]}; \quad (3)$$

c) when the signal at the output of the Splitter comes from the direct lines (i=1÷ 4):

$$C_{i,out}^{(S)} = C_{in,i} - a_{thru}, \text{ [dBμV]}. \quad (4)$$

2. Calculation of the noise at the output of the passive component:

a) when the noise at the output of the Tap comes from the direct line:

$$N_{0,out}^{(T)} = N_{0,in} - a_{thru}, \text{ [dBμV]}; \quad (5)$$

b) when the noise at the output of the Tap comes from the side/drop/tap lines (i=1÷ 4):

$$N_{i,out}^{(T)} = N_{tap,i} - a_{tap,i}, \text{ [dBμV]}; \quad (6)$$

c) when the noise at the output of the Splitter comes from the direct lines (i=1÷ 4):

$$N_{i,out}^{(S)} = N_{in,i} - a_{thru}, \text{ [dBμV]}. \quad (7)$$

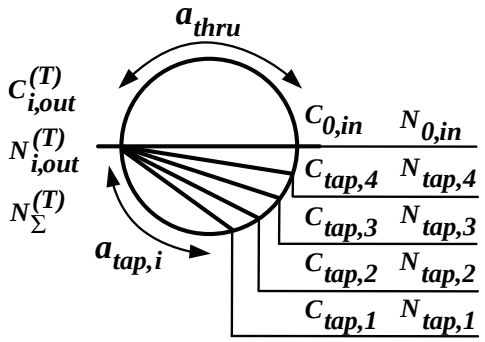
3. Calculation of the aggregate noise at the output of the passive component:

a) for Tap:

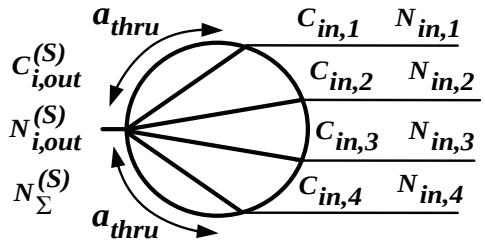
$$N_{\Sigma}^{(T)} = 20\lg[\sum_{i=0}^{4} 10^{(N_{i,out}^{(T)}/20)}], \text{ [dBμV]}; \quad (8)$$

b) for Splitter:

$$N_{\Sigma}^{(S)} = 20\lg[\sum_{i=1}^{4} 10^{(N_{i,out}^{(S)}/20)}], \text{ [dBμV]}. \quad (9)$$



a) Tap



b) Splitter

Fig.2. Graphic symbols of passive devices

III. CALCULATION OF U_{CM}, C/N AND C/N_{Σ}

The practical application of the above shown mathematical relationships is described below for cases in which are used different passive devices for branching of the signal to the subscribers (Figs.3 and 4). Considered are cable distribution networks for subscribers, which live in family houses. Proposed are different options, for making a comparison of the network quality and recommendations for their application.

Cable amplifiers are with built amplifier for return path in order to be provided a level of signal 80dBμV in the input of every one of them. This way are compensated the losses in the coaxial cables and the passive devices between every two amplifiers. The distances from the subscriber equipment (set-top-box) to the respective directional coupler (Tap) are equal. This will allow the comparison of the respective parameters of signals not only for the subscribers from one branch, but also for those from both branches (upper and lower), Figs.3 and 4.

The maximal distance in the upper branch between CA and subscriber (test point 24 and test point 1/2/3) is 185m, and in the lower one it is 140m. The proposed example architecture of the cable distribution network allows the number of subscribers to increase from 13 to 100, if Taps with four tap ports and four port Splitters are used.

On Figs.3 and 4 are presented the architectures of a cable distribution network with and without Return Step Attenuator (RSA). RSA are wrapped directly to port "tap" of a Tap and to port "out" of a Splitter. The value of RSA by balancing of return path must be chosen so, that the level of noise (N, N_{Σ}) in direction to the amplifier reduces itself. If at any test point this level is higher from the noise level at a previous test point, it is necessary the value of the attenuation at the step attenuator to be increased. The results of the analysis for the signal's level alternation, carrier-to-noise ratio and carrier-to-aggregate noise at different test points are given in Table 1, Figs.5 and 6.

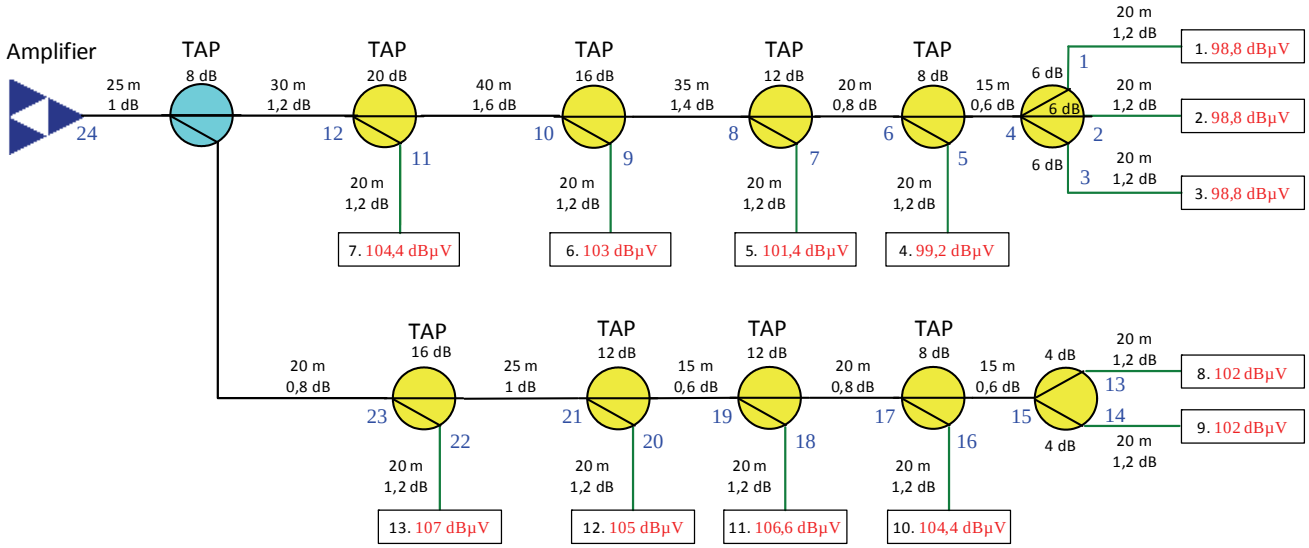


Fig.3. Architecture of cable distribution network without Return Step Attenuator

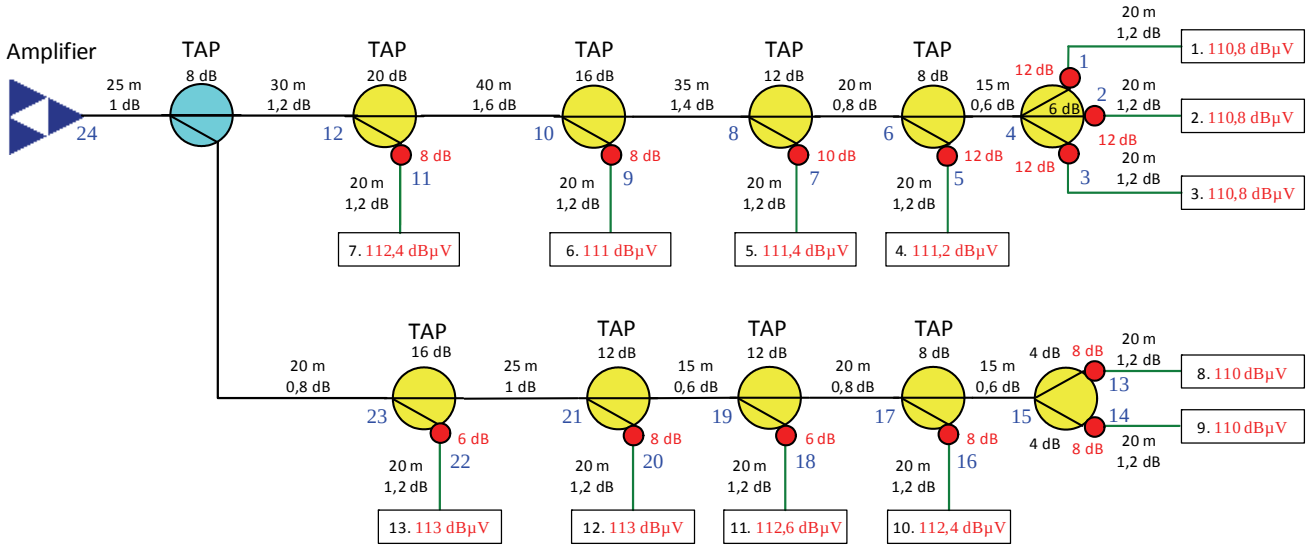


Fig.4. Architecture of cable distribution network with Return Step Attenuator

TABLE 1

Test point №	without RSA			with RSA		
	C dBµV	*C/N* dB	C/N_{Σ} dB	*C* dBµV	*C/N* dB	C/N_{Σ} dB
1	97,6	67,6	-	109,6	79,6	-
2	97,6	67,6	-	109,6	79,6	-
3	97,6	67,6	-	109,6	79,6	-
4	91,6	-	58,06	91,6	-	70,6
5	98	68	-	110	80	-
6	90	-	55,66	90	-	67,66
7	100,2	70,2	-	110,2	80,2	-
8	88,2	-	54,17	88,2	-	65,82
9	101,8	71,8	-	109,8	79,2	-
10	85,8	-	53,1	85,8	-	64,24
11	103,2	73,2	-	111,2	81,2	-
12	83,2	-	52,3	83,2	-	63,09
13	100,8	70,8	-	108,8	78,8	-
14	100,8	70,8	-	108,8	78,8	-
15	96,8	-	64,8	96,8	-	72,78
16	103,2	73,2	-	111,2	81,2	-
17	95,2	-	62	95,2	-	69,99
18	105,4	75,4	-	111,4	81,4	-
19	93,4	-	60,32	93,4	-	67,92
20	103,8	73,8	-	111,8	81,8	-
21	91,8	-	58,65	91,8	-	66,32
22	105,8	75,8	-	111,8	81,8	-
23	89,8	-	57,52	89,8	-	64,97
24	80	-	48,5	80	-	57,96

In the defining of signal levels at the test points in the drop lines is used equation:

$$C_i = U_{CM,k} - \ell_{SSK} . a_{SSK} \text{ , [dBµV], where} \quad (10)$$

i= 1, 2, 3, 5, 7, 9, 11, 13, 14, 16, 18, 20 and 22 or for our case:

$$C_i = U_{CM,k} - 1{,}2 \text{ , [dBµV].} \quad (10a)$$

IV. CONCLUSION

From Fig.5 is to be seen, that the levels of signal for a specific test point in the direct line with and without RSA are the same. The levels of signal in the drop line for a specific test point are different (the level of signal at a given test point by using a RSA is higher from this one at the same point but without RSA and with the value of the attenuation of a RSA).

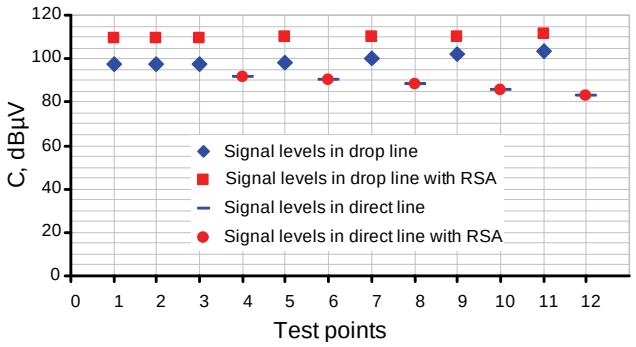
From Fig.6 is to be seen, that the values of *C/N* and C/N_{Σ} are increasing by using of RSA, as for test point 24 it is C/N_{Σ}=57,96dB. This value is with around 9,5dB higher from the value of C/N_{Σ}=48,5dB in an architecture, in which is not used a step attenuator.

REFERENCES

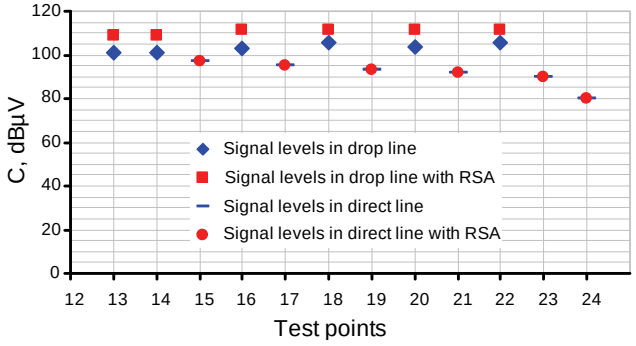
[1] Symmetrical Services over HFC Networks. White Paper, Xtend Networks Inc., pp. 1-7, January 2003.

[2] L. Jordanova, D. Dobrev. "Improvement of the HFC system reverse path performance", ELECTRONICS' 2005, 21-23 Sept., Sozopol, pp.156-161, 2005.

[3] K. Koitchev, K. Angelov and S. Sadinov, *Design of interactive cable television networks*, Ex-Press, Gabrovo, 2010.

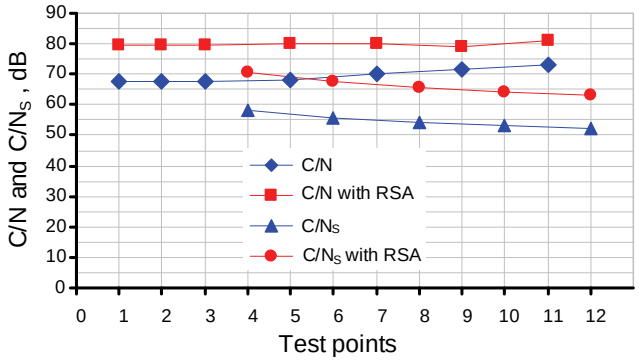


a) Upper branch

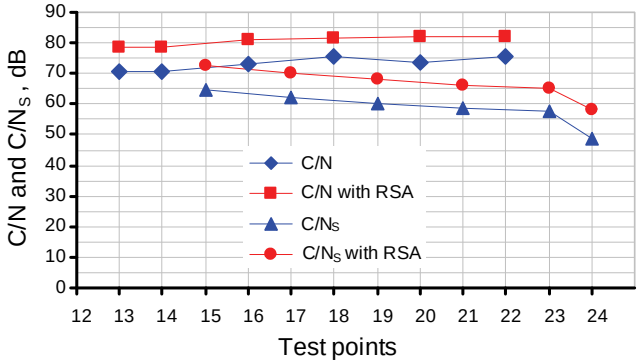


b) Lower branch

Fig.5. Signal levels



a) Upper branch



b) Lower branch

Fig.6. *C/N* and C/N_{Σ}

[4] Product Application Note: ARSA Return Step Attenuator. Arcom Labs, Syracuse, NY, 2008.

The Effects of Multiple Reflection in Conducted RF Measurements

András Fehér[1] and Szilvia Nagy[2]

Abstract – **During conducted measurements the reflection is one of the most varyingcomponents of the measurement uncertainty. For studying the cable reflections, in this paper we make a simple computer model and compare its output with measured results.**

Keywords– **Reflection, frequency dependency of velocity factor, measurement, measurement uncertainty.**

I. INTRODUCTION

Well known formula [1] for the solution of telegrapher's equations is

$$v(x,t) = |A| \cdot \cos(\omega \cdot t - \beta \cdot x) \cdot e^{-\alpha \cdot x} + |B| \cdot \cos(\omega \cdot t + \beta \cdot x) \cdot e^{\alpha \cdot x} \quad , \tag{1}$$

Here,A and B are the complex amplitudes of the forward and reflected waves in the transmission line, andα, βare the real and imaginary parts of the propagation constant

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \, . \tag{2}$$

In this case R, L, G and C are the per-unit-length parameters of the transmission line.

The velocity of the waves in the guide also depends on the cable parameters as

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{L \cdot C}} = \frac{1}{\sqrt{\varepsilon \cdot \mu}} = \frac{c}{\sqrt{\varepsilon_r \cdot \mu_r}} \, . \tag{3}$$

Well known is the effect of the reflection if the load impedance ($Z_L$in Fig.1) is not matched with waveguide impedance (Z_0 in Fig.1).The complex voltage reflection coefficient (Γ) is

$$\Gamma = \frac{B}{A} = \frac{Z_L - Z_0}{Z_L + Z_0} \, , \tag{4}$$

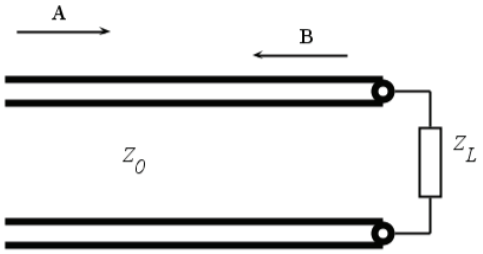


Fig. 1. Reflection at load

The characteristic impedance of the transmission line is expressed as [2]

$$Z_0 = \frac{V^+}{I^+} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \, . \tag{5}$$

In case of conducted power measurements the reflection causes uncertainty. In this article we made a simple model to study the effect of multiple reflections.

The first step in developing a simulation model of the studied system is to know the precise properties of the measured transmission lines.One of the most important property of the reflection simulation is the β propagation constant of the used cables.

II. STUDY OF PROPAGATION VELOCITY IN COAXIAL CABLES

As Eq. (1) shows, the waves in the lossless transmission lines havea periodicity in time and space in case of sinusoidal excitation.

$$v(x,t) = V\left(0, t - \frac{x}{v}\right) . \tag{6}$$

Measuring the propagation velocity a short circuited transmission line can be carried out as Fig. 2 shows.

The first attenuation (S_{21}) maximum belongs to the resonance frequency of the L long cable under test. The velocity factor in the cable under test is

$$Vf = \frac{\lambda_c}{\lambda_0} = \frac{2 \cdot L}{\lambda_0} = \frac{2 \cdot L \cdot f}{c} \, , \tag{6}$$

where *f* is the resonance frequency of the cable under test, and *c* is the speed of the waves in free space 299 792 458 m/sas Eq. (3) sows.

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

[2]SzilviaNAGY, Széchenyi István University, Department of Telecommunications, Egyetem tér 1, Győr, Hungary, H-9026 E-mail: nagysz@sze.hu.Web: http://ta.sze.hu

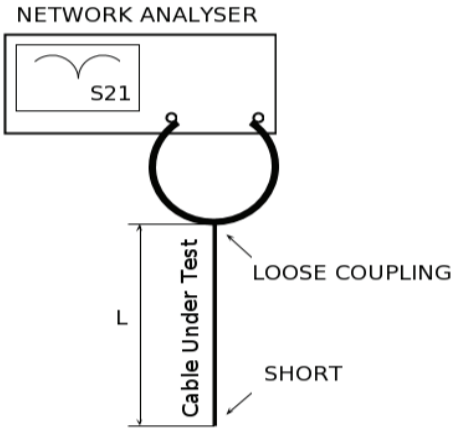


Fig. 2. Measurement setup for propagation velocity

Generally the observed frequency is determined by the length of the cable under test (*L*). Instead of applying different lengths of cables higher harmonics of the resonance can be used for determining the frequency dependence of the velocity factor and thus the dielectric constant.

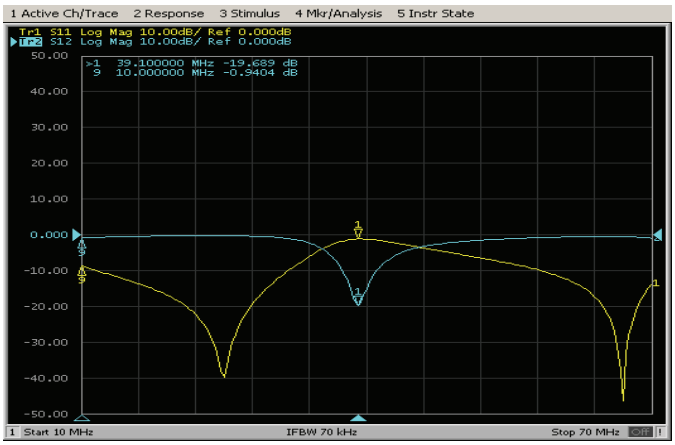


Fig. 3. The 1st resonance of the CUT (1) at 39.1000 MHz, with cable length 11.290 m. In the picture the minimum of the S_{12}can be seen at MARKER 1.

Unfortunately, for low frequency examination of the velocity factor this method needs long cable length*(L)*. The radiation of the open ended cablesintroduce more resonance disturbances, therefore it is worth to use two times longer short circuited cables for the precise measurements.

In this article the Cable Under Test (CUT) types are (1) Hirschmann KOKA 709 (75 Ω), and (2) H155 (50 Ω) low loss coaxial cables. The Fig. 3 shows the velocity factor'sfrequency dependency. The frequency dependency of ε can be calculated from Eq. (3), too.

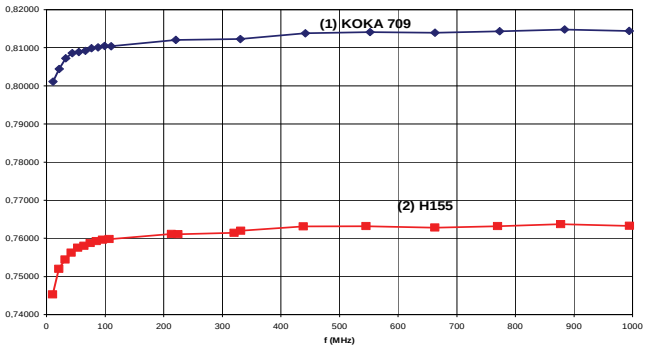


Fig. 4. Frequency dependency of velocity factor of (1) KOKA 709 and (2) H155 coaxial cables

wavelength variability have to be taken care of, at the same time. Therefore the used algorithm is

(1) seek the lowest resonance frequency (f_1),
(2) seek the next resonance frequency (f_2) near $2 \cdot f_1$,
(3) seek the n^{th} resonance frequency (f_n) near $n \cdot f_{n-1}/(n-1)$.

By using automated measurement control the algorithm can followed easily, but at manually measured values, if you want to measure at near discrete frequencies, it is difficult to identify the order number of resonances because of the frequency dependency of the cable parameters.

As Fig. 2 shows for the test we have to use loose coupling between the CUT and the measuring loop for decreasing the unwanted impedance transformation into the measured transmission line. In this case the minimumof S12 can be smaller as Fig. 3 shows.

As Fig 4 shows, from 100 MHz to 1000 MHz the variability of velocity factor is not dominant, therefore in the next model we use constant instead of it.

III. Model Elements for Multiple Reflecton of Transmission Line

A. Generator

The RF generator can be represented by itsoutput impedance (Z_g), and output voltage (V_g), at the nececcary frequency, of course. In practice V_g is calculable from the output power and Z_g, if it is matched. If the load impedance of the generator varies a lot, instead of the output power the *emf* (electromotive force) value should be used, witch is the output voltage of a generator without any load. This value is two times higher then the matched case.

In our model the generator output voltage is

$$V_g(t) = V_0 \cdot \cos(\omega \cdot t)\,. \tag{7}$$

Instead, by introducing$\boldsymbol{A_g}$ as forward complex peak amplitude, the generator voltage can be expressed as

$$V_g(t) = real(A_g \cdot e^{j\omega t})\,, \tag{8}$$

B. Transmission line

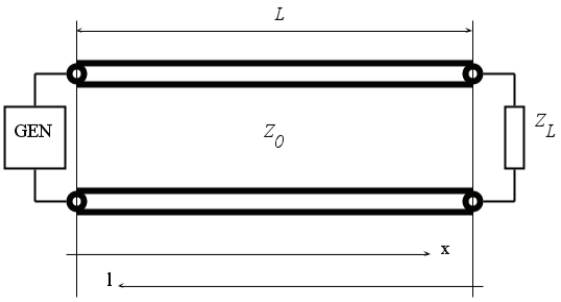


Fig. 5. Transmission line model

The $A_g$forward wavein the line at place $x$is

$$A(x) = A_g \cdot e^{-\alpha x} \cdot e^{-j\beta x}, \tag{9}$$

and

$$A(x,t) = A_g \cdot e^{-\alpha \cdot x} \cdot e^{-j \cdot \beta \cdot x} \cdot e^{j \cdot \omega \cdot t}, \tag{10}$$

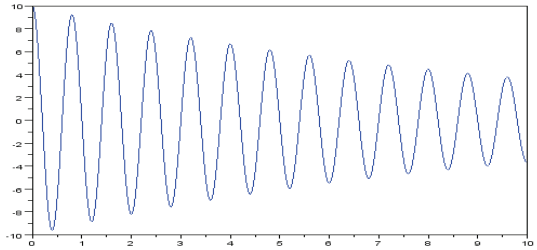


Fig. 6. The *real(A(x,t))* voltage in the transmission line (L=10m, A_g=10V, t=0, α=0.1, f=300MHz)

Observing the voltage of the transmission line in one period, its form can be seen in Fig. 7.

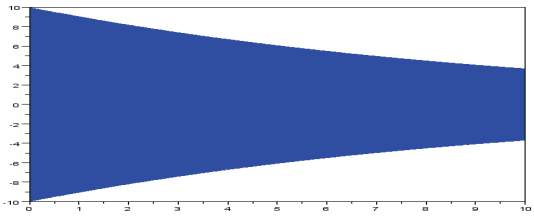


Fig. 7. Voltage in the transmission line (L=10m, A_g=10V, t=0…T, α=0.1, f=300MHz)

The B forward wave in the line at place $x$is

$$B(x,t) = B_0 \cdot e^{-\alpha l} \cdot e^{-j\beta l} \cdot e^{j \cdot \omega \cdot t} = \left[A_g \cdot e^{-\alpha \cdot L} \cdot e^{-j \cdot \beta \cdot L} \cdot \Gamma\right] \cdot e^{(x-L) \cdot \alpha} \cdot e^{j \cdot (x-L) \cdot \beta} \cdot e^{j \cdot \omega \cdot t} \tag{11}$$

The forward and reflected voltage at t=0 is shown in Fig. 8. forΓ= −1, and inFig. 9. forΓ= +1.

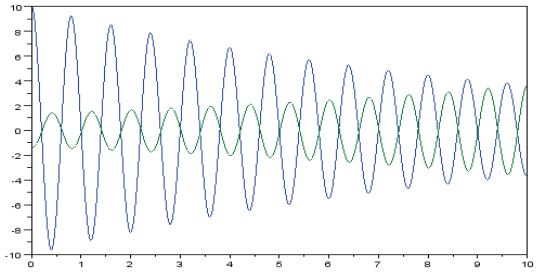


Fig. 8. Reflected voltage in the transmission line as a function of x(L=10m, Γ= −1)

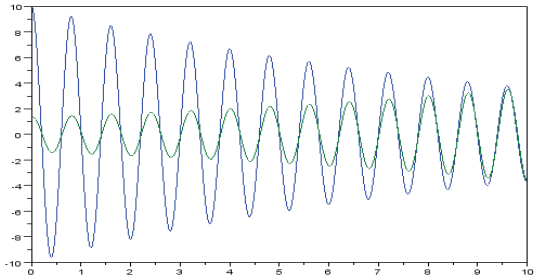


Fig. 9. Reflected voltage in the transmission line as a function of the distance x (L=10m, Γ= +1)

The voltage in the transmission line can be got by

$$V(x) = real(A(x) + B(x)). \tag{12}$$

The voltage shape in the transmission line for one period with Z_L=0 and with Z_L=∞can be seen in Figs. 10 and 11.

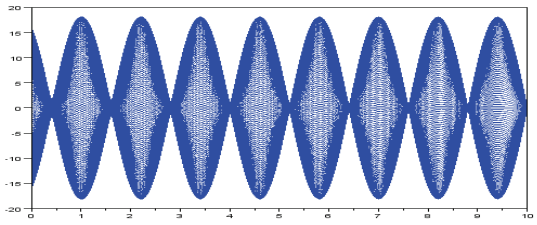


Fig. 10. Wave in the loss transmission line as a function of x (L=10m, Γ= −1, f=100 MHz)

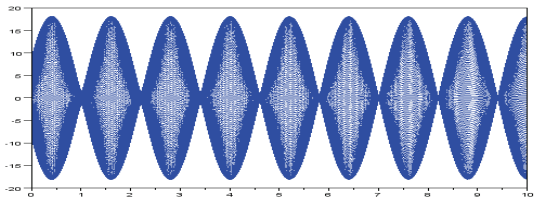


Fig. 11. Wavew in the loss transmission line as a function of x (L=10m, Γ= +1, f=100 MHz)

The difference depends on the load, as it can be seen at the above plots at the cable ends (x=10 m).

C. Transmission line steps

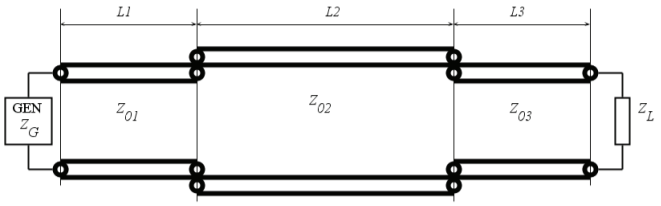


Fig. 12. Cascade coupled transmission lines

The input impedance of the 3rd transmission line in general case is [3]

$$Z_{in} = Z_0 \cdot \frac{Z_L + Z_0 \cdot th(\gamma L)}{Z_0 + Z_L \cdot th(\gamma L)}. \quad (13)$$

In case of lossless transmission lines, where α=0

$$Z_{in} = Z_0 \cdot \frac{Z_L + j \cdot Z_0 \cdot tg(\beta L)}{Z_L + j \cdot Z_0 \cdot tg(\beta L)}. \quad (14)$$

The reflection coefficient at step $Z_{01} \rightarrow Z_{02}$ is Γ_{12} from Eqs. (4) and(13). In this simple model the $Z_g=Z_{01}$, $Z_{03}=Z_L$, therefore there are no reflections at the generator and at the load.

The input voltage of the 2nd line must be equal to the voltage of the 1st line at the end. (It depends on the Γ_{12}.) At the end of 1st line the voltage is V_{2IN}

$$V_2(x_2 = 0) = real(A(x_1 = L_1) + B(x_1 = L_1)). \quad (15)$$

$$B = \Gamma \cdot A. \quad (16)$$

Therefore

$$A_2(0) = A_1(L_1) \cdot [1 + \Gamma_{12}]. \quad (17)$$

At step$Z_{02} \rightarrow Z_{03}$ the reflection is Γ_{23}, therefore the reflected value is

$$B_2(L_2) = A_2(L_2) \cdot \Gamma_{23}. \quad (18)$$

This reflected wave is reflecting at step$Z_{02} \rightarrow Z_{01}$

$$A_{2[1]}(0) = B_2(0) \cdot \Gamma_{21}. \quad (19)$$

$$A_{2*}(0) = A_2(0) + \sum_{i=1}^{\infty} A_{2[i]}. \quad (20)$$

IV. Conclusion

Generally the higher order reflections are not dominant, i.e.,

$$A_{2*}(0) \sim A_2(0) + \sum_{i=1}^{\infty} (\Gamma_{23} \cdot \Gamma_{21})^i. \quad (21)$$

The phases of the 1st, 2nd etc. reflected waves will be the same. The simulation result by SciLab [4] can be seen in Fig. 13.

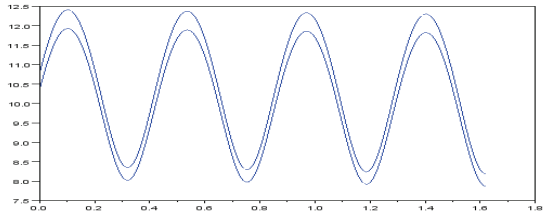


Fig. 13. $A_{20}+B_{20}$(lower curve), and $A_{2*}+B_{2*}$(upper curve) in the transmission line(L=1.62m, α=0.01, $Z_G=Z_{01}=Z_{03}$=50Ω, Z_{02}=75Ω)

Acknowledgement

This work was supported by the Bolyai Janos Research fellowship of the HungarianAcademy of Sciences and the TáMOP 4.2.1.A project of the SzéchenyiIstvánUniversity.

References

[1] M. Kuczmann, A. Iványi, *Finite Element Method in Magnetics*, Academic Press, Budapest, 2008, ISBN: 978 963 05 8649 8.

[2] Kuczmann,Finite Element Modeling of Antenna FeedingActa Technica Jaurinensis, Vol.3 No.1, 2010, pp. 127-136

[3] Istvánffy, Edvin, *Transmission Lines, Antennas, Wave propagation* (Tápvonalak, antennák, hullámterjedés), Budapest, Tankönyvkiadó, 1987.

[4] SciLab 5.3 The Free Software for Numerical Computation. http://www.scilab.org

Investigation into Filter with Hausdorff's Weighted Window Function Designed for Wideband Channels

Borislav Naydenov[1], Ginka Marinova[2] and Valentina Markova[3]

Abstract – **In this paper the results received from investigation into non-recursive digital filter are presented. Hausdorff's weighted window is used as well known Kaiser window. For the weighted window is applied expression representing the delta function approximation to algebraic polynomial with Hausdorff's values of the coefficients. Correlated interferences observed in UWB cannels are used.**

Keywords – **non-recursive digital filter, WB channel, Hausdorff's window**

I. INTRODUCTION

Digital filters with the finite impulse response (FIR) are discrete systems with one input and one output and constant in time parameters. They are characterized by strictly linear phase characteristic. This determines their wider use in practical implementations compared to filters with infinite impulse responce (IIR). Transfer function of the physical realizable FIR filter has the following form[2]:

$$H(z) = \sum_{n=0}^{N-1} h(n).z^{-n}\,, \tag{1}$$

where *H(z)* is a polynomial of z^{-1} of degree *N-1*. Thus, *H (z)* has *N-1* zeros, which can be located arbitrarily in the final z-plane. Poles are *N-1* and are located in the central point z = 0. The frequency response is trigonometric polynomial

$$H(e^{iw}) = \sum_{n=0}^{N-1} h(n)e^{-iwn}\,. \tag{2}$$

The design of FIR filters can be accomplished either by finding the coefficients of the pulse characteristics, or the determination of N samples of the frequency response. The basic approach in defining filters FIR is by crossing his infinite length pulse characteristic and thus obtain the impulse characteristic with the final length. Assuming that $H_d(e^{iw})$ is the ideal frequency response characteristic, then the corresponding sequences of the samples of the pulse characteristic has the form:

$$h_d(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(e^{i\omega})e^{i\omega n}d\omega\,. \tag{3}$$

In most cases, the ideal frequency response characteristic $H_d(e^{iw})$ of frequency selective filter is constant for different parts of the passband and stopband, with a break at the endpoints between them. The interruption determines the sequence of pulse characteristics, which have infinite lengths $h_d(n)$ and requires it to be truncated to obtain physically realizable filters. Periodic frequency characteristic in (3) can be viewed as presented by means of Fourier series, where pulse sequence $h_d(n)$ acts as Fourier coefficients. Intersection of the ideal impulse characteristic is equivalent to to study convergence of Fourier sequence. This convergence in the theory is presented as Gibbs phenomenon.

The errors within the passband and stopband are specified as δp and δs. The frequency response is allowed to fluctuate both positively and negatively within these error limits. We can translate these specifications into the decibel gain by using (4) and (5), [3].

$$a_{pass} = 20\log(1-\delta_p) \tag{4}$$

$$a_{stop} = 20\log(\delta_s)\,, \tag{5}$$

where δ_s and δ_p are the error coefficients in stopband and passband.

II. USING WINDOWS TO OBTAIN FIR FILTERS

Physically realizable FIR filter is obtained by restricting the ideal pulse characterization [2]. From a mathematical point of view, limiting the ideal pulse characteristic is equivalent to multiplication with a weight function:

$$w(n) = \begin{cases} 1, & 0 \le n \le N-1 \\ 0 & \end{cases}. \tag{4}.$$

The actual pulse characteristics are:

[1]Borislav Naydenov is with the Faculty of Electronic, Technical University - "Studentska" 1, 9010 Varna, Bulgaria, E-mail: borna@abv.bg.

[2]Ginka Marinova is with the Faculty of Computing and Automation, Technical University - "Studentska" 1, Varna, Bulgaria, E-mail: gin_kaleva@abv.bg.,

[3]Valentina Markova is with the Faculty of Electronics, Technical University – "Studentska" 1, 9010 Varna, Bulgaria,
E-mail: valliq@abv.bg.

$$h(n) = h_d(n).w(n). \quad (5)$$

Transfer function of the actual filter is obtained by convolution between the transfer function of the ideal filter and frequency response of the window.

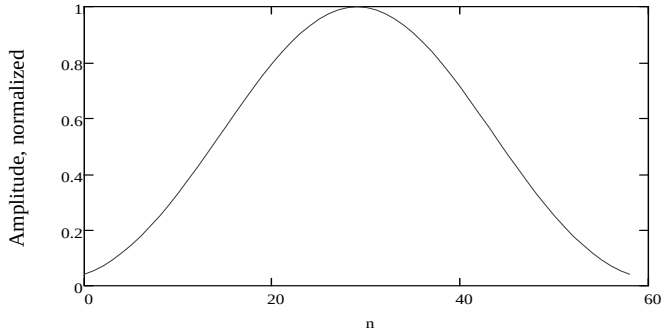


Fig. 1. Kaiser window

The frequency response of the window should be as narrow as to be able to reproduce precisely specified frequency response. With increasing the number of elements N of the transfer function width of the main leaf and the side leaf of the frequency responses of the window is decreased. By this reason, their amplitudes increased, as the area under these main and side leaves remains the same. This leads to larger fluctuations in the frequency response of the synthesis filter. In the theory of Fourier series is known Gibbs phenomenon, which defines this uneven convergence. It can be reduced by using a smooth intersection of Fourier series. With the gradual shrinking of the window to zero on each side, can reduce the height of the side leaf, which obtain by increasing the width of the main leaf and hence a wider zone at the point of sharp transition.

In practice, a variety of weight functions, which in most cases have names of their discoverers, is using. Kaiser offers a family of weight functions using modified Bessel function of first kind and zero order [2]

$$\omega(n) = \frac{I_0\left[\omega_a\sqrt{\left(\frac{N-1}{2}\right)^2 - \left[n - \left(\frac{N-1}{2}\right)\right]^2}\right]}{I_0\left[\omega_a\left(\frac{N-1}{2}\right)\right]}. \quad (6)$$

They obtained the greatest power in the main leaf when is set the amplitude of side leaves. This property makes Kaiser window (KW) nearly optimal. The Kaiser Window function is shown on Fig. 1 - $\omega_a = 5.605$.

III. Application of Hausdorff's Weight Function

In scientific publication [5] is proposed a weight function with Hausdorff's metrics. Weight function is obtained by approximation of a delta function with algebraic polynomial (7) [6], that performs the best approximation in Hausdorff's metrics in the interval [-1, 1].

$$P_m(x) = \varepsilon T_m\left(\frac{2x^2 - 1 - \alpha^2\varepsilon^2}{1 - \alpha^2\varepsilon^2}\right), \quad (7)$$

where ε means Hausdorff's distance; T_m is polynomial of Chebishev of first order and degree m; α - parameter, $\alpha\varepsilon$ - value, which determins the width of the function in the main maximum. Dependencies between the parameters of the polynomial are determined by expression (8):

$$\alpha\varepsilon = \sqrt{\frac{ch\left[\frac{1}{m}Ach\left(\frac{1}{\varepsilon}\right)\right] - 1}{ch\left[\frac{1}{m}Ach\left(\frac{1}{\varepsilon}\right)\right] + 1}}. \quad (8)$$

Weight function is obtained by Hausdorff's polynomial translation in a positive direction with a value one, definitional field is reduced to field of main maximum $[1-\alpha\varepsilon, 1+\alpha\varepsilon]$ and raise of degree 1.27 [5]:

$$w_m(x) = \left\{\varepsilon\cos\left[m\arccos\left|\frac{2(\alpha\varepsilon x - \alpha\varepsilon)^2 - 1 - \alpha^2\varepsilon^2}{1 - \alpha^2\varepsilon^2}\right|\right]\right\}^{1,27}. \quad (9)$$

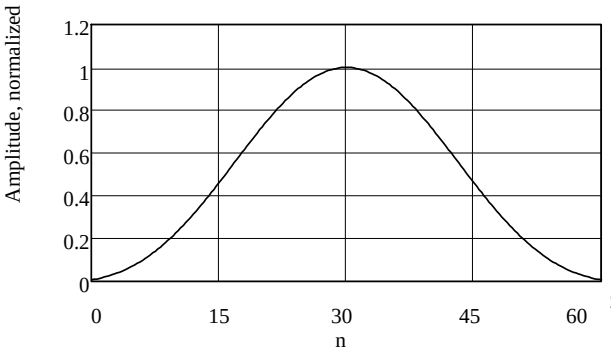


Fig.2. Hausdorff's weight function, attenuation a = 60 dB, Hausdorff's distance ε = 0.021 and parameter α = 1.693

With a mark attenuation in stopband expressed in decibels. Hausdorff's distance is determined by the dependencies:

When *a<24 dB*

$$\varepsilon = 1. \quad (10)$$

When *24 dB* $\le a <$ *50 dB*

$$\varepsilon = \frac{0.66}{\left(2.7e - 5a^2 - 8e - 4a + 1.073\right)^{a-25}}. \quad (11)$$

When *50 dB<a* $\le$ *130 dB* (Fig. 2)

$$\varepsilon = \frac{0.66}{1.1035^{a-25}}. \quad (12)$$

When *a>130 dB*

$$\varepsilon = \frac{0.66}{\left(0.0001a + 1.09\right)^{a-25}}. \qquad (13)$$

It is appropriate the degree of Hausdorff's polynomial m to be equivalented to N.

IV. Results from the Investigation into Digital Filter with FIR using Weighted Window Described by Hausdorff's Function and Kaiser Window

For the investigation into digital filter with FIR is used models of correlated interference, obtained by the avtoregresion rows. The coefficients of avtoregersion sequence are selected so as to dominate the low frequency component of interference (Fig. 3).

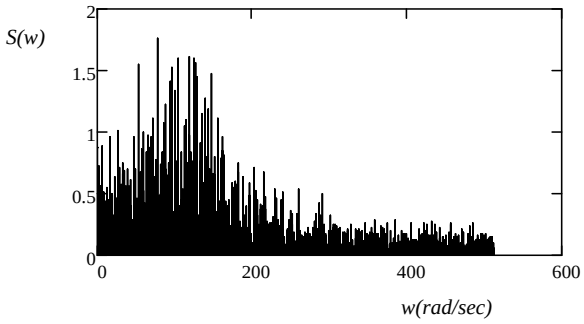


Fig.3. Spectrum of correlated interference

There are several approaches in the synthesis of filters. In investigation in this publication is set ideal impulse characteristic of the type $h_d(n) = \frac{\sin[\omega_c(n-\kappa)]}{\pi(n-\kappa)}$ [7]: where ω_c - cut-off frequency, $\kappa = \frac{N-1}{2}$ and N - order of the filter. Real pulse characteristic is obtained by expression (5). Transfer function of the filter is obtained by convolution between the transfer function of the ideal filter and frequency response of the window. When is used the rectangular window is received non satisfactory results in approximation of frequency response characteristic in the transitional area. In the stopband is observed large ripple of response. When is applied the filter processing with a rectangular window of the investigated correlated interference we received ripples expressed by (5) - a_s=7 *dB* and attenuation about 17 dB.

In these studies used lowpass filter with Kaiser window and highpass filter with Hausdorff's window (HW). For lowpass filter ω_c = 90 Hz. For the highpass filter ω_c = = 220 Hz.

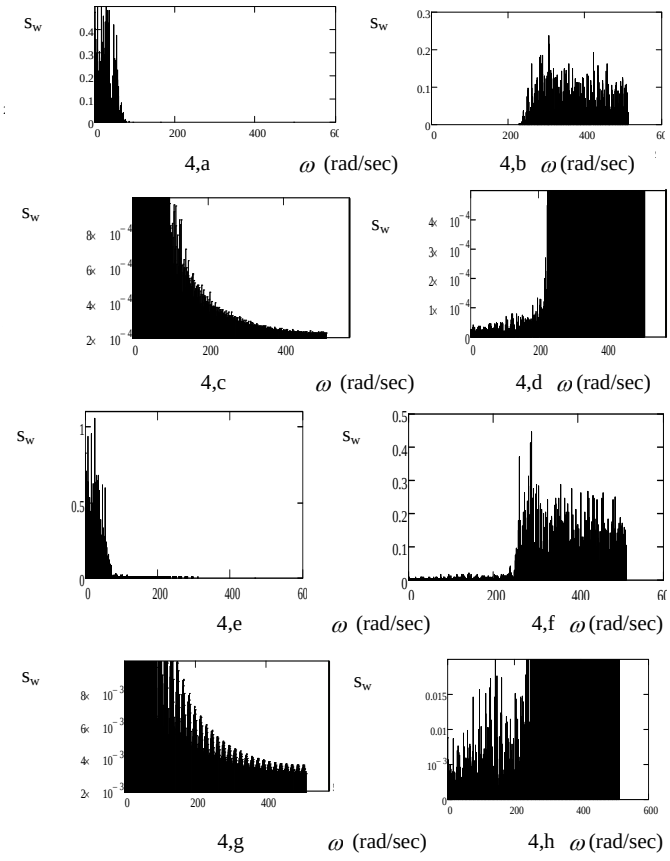


Fig. 4. Output signals of filter with KW (4a, 4c, 4e, 4g) and HW (4b, 4d, 4f, 4h). The input signals are processed with KW (4a, 4c), HW (4b, 4d) and rect. window (4e, 4f, 4g, 4h)

V. Conclusion

When used KW (Fig.4c) in stopband occurs gradually increasing of the attenuation, as the maximum attenuation is 80 dB. When using the HW (Fig. 4d) there is less change in the stopband. Attenuation is about 95 dB, but there are more ripples. In HW processing transitional area is steeper (Fig.4).

The results confirm the conclusions made in the article [5].

References

[1] Z. Nikolova, V. Poulkov, G. Iliev, K. Egiazarian. New adaptive complex IIR filters and their application in OFDM systems. J. S., Image and Video Proc. Springer. V. 4, I. 2, 2010, pp. 197-207.

[2] A. Oppenheim, R. Schafer, Digital signal processing. Prentice hall, inc., Englewood cliffs, New Jersey, 1979.

[3] Les Thede, Practical Analog and Digital Filter Design, Artech House, Inc., 2004.

[4] Р. Иванов, Цифрова обработка на едномерни сигнали, Almamater international, Габрово, 1999 г.

[5] Peter Apostolov, Application of Hausdorff's Window Function by FIR Filters Synthesis, ICEST 2008, Serbia, Nis, 25-27 June.

[6] Бл. Сендов, Хаусдорфовые приближения. С., БАН, 1979.

[7] R. Bogdanov, B. Bedzhev, Z. Tasheva, B. Stoyanov, An Algorithm for Lossless Signal Processing, Proc. of the Third IEEE W-p on IDAACS 2005, IX, 5-7, 2005, Sofia, Bg, pp.498.

Portable 3D System for Visualization and Protection of Wireless Networks

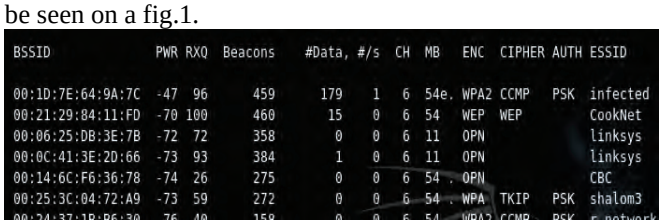
Teodor Kalushkov[1], Plamenka Borovska[2] and Georgi Todorov[3]

Abstract – **This paper describes a wave analyzing system for wireless networks, that can be easy integrated in some working simulators, using the existing or partial made models. The system can visualize real time results, in a 3D form that gives opportunity for enhancing the security and architecture of existing systems, correcting virtual models and researching real networks. The system has an enhanced possibility for detecting external intrusion tries and localization of the external equipment.**

Keywords–**signal, antenna, waves, security, 3D position.**

I. Introduction

Many of the existing systems for modeling and building a wireless networks use different 3D models, based on pure theoretical dependencies, but after real construction of these networks the experimental analysis are made usually only on the base of signal level. Sometimes it is not enough, because there are some other factors that have influence over the desired secure connection. Usually used applications give us information about MAC-addresses of users, of access points, channels of communication, used encryption method and as mentioned before about the signal strength. Such monitor can be seen on a fig.1.



BSSID	PWR	RXQ	Beacons	#Data,	#/s	CH	MB	ENC	CIPHER	AUTH	ESSID
00:1D:7E:64:9A:7C	-47	96	459	179	1	6	54e.	WPA2	CCMP	PSK	infected
00:21:29:84:11:FD	-70	100	460	15	0	6	54	WEP	WEP		CookNet
00:06:25:DB:3E:7B	-72	72	358	0	0	6	11	OPN			linksys
00:0C:41:3E:2D:66	-73	93	384	1	0	6	11	OPN			linksys
00:14:6C:F6:36:78	-74	26	275	0	0	6	54 .	OPN			CBC
00:25:3C:04:72:A9	-73	59	272	0	0	6	54 .	WPA	TKIP	PSK	shalom3

Fig.1. Wireless network monitor

If we have a 3D map of access points and users, it will be easier to detect an intrusion attempts. For building this map theory is not enough and real physical measurements should take their place.

II. Attack Actions And Defence

When hackers realize attacks, they use equipment, which is situated near the desired wireless networks. This equipment usually requires not only a computer but also directional antennas, because they have to jam the signal of the access point or user. The most popular techniques for finding wireless passwords include sending deauthorization signal and capture of four way handshake between user's computer and access point or packet injection. When network password is found, hackers can sophisticate the packets if WEP encryption is used [3]. Otherwise, if WPA/WPA2 is used, online sophistication is not so easy but remains the possibility for capturing packets and decoding them offline [2]. In these situations if access point and user computer can determine their position in a 3D environment, they can easy detect intrusion attempts. So the jamming can be detected not only as a rapid change in a signal level, but also as a position "jump" of the opposite point on the 3D map. This example is shown on fig.2.

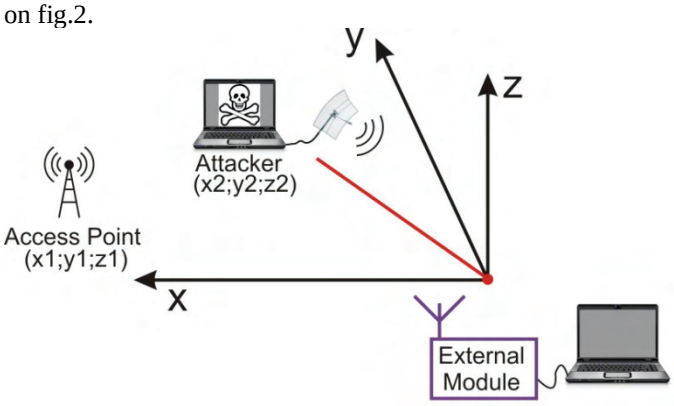


Fig.2. Intrusion detection, using 3D mapping

If we analyze the position, according signal strength and phase of the access point signal, when attacker tries to emulate access point, 3D analyzer will detect immediately change of the coordinates. Often the software protections in this field react against the intrusion with blocking the communication for a short period of time and trying to reestablish the connection afterwards. This approach is not convenient for the user and does not solve the problem permanently.

III. Main Parts And Operating Principles Of The System

In order to detect the phase and strength of the opposite side, our system should ask the access point to return a test signal which will give us a base for our calculations and orientation in the 3D space. In order to avoid latencies and complications the system will use its own unit, of antennas. In 2D space we need minimum three of them. According to our goals the system should have minimum four antennas,

[1]Teodor Kalushkov is with the Faculty of Mathematics and Informatics, St.Cyril and St. Methodius University of Veliko Tarnovo, Arch. G. Kozarev 3, V.Tarnovo 5000, Bulgaria, E-mail: teodork@abv.bg

[2]Plamenka Borovska is with the Faculty of Computer Systems and Control, Technical University of Sofia, boulevard Kliment Ohridski 8,Sofia 1000, Bulgaria, E-mail:pborovska@tu-sofia.bg

[3]Georgi Todorov is with the Faculty of Mathematics and Informatics, St.Cyril and St.Methodius University of Veliko Tarnovo, Arch.G.Kozarev 3, V.Tarnovo 5000, Bulgaria, E-mail:g.todorov@uni-vt.bg

positioned in the edges of a virtual pyramid. The distances between them should be as longer as they can be, in the range of a room, in order to reach greater accuracy. This is the other reason for choosing external antennas, despite of using existing ones. All antennas will only receive signals. It is good decision because so they can not be detected from intruders and the whole system will require less power consumption. Each antenna should contain three elements, if we want make our system compatible not only with IEEE 802.11b and g standards but also with IEEE 802.11n one.

The antennas should be connected with other external device, using equal length cables. It is required of the attenuation in the cable. If the system uses different distances between antennas and concentration device, we should integrate some kind of signal compensators or provide some extra computations and will complicate the prototype at all.

All cables can be concentrated in external device, that measures the strength and phase of all antennas, converts the signals into a digital form, calculates the coordinates and sends them to the personal computer via proper interface for post processing. External device can contain the elements, shown on fig.3.

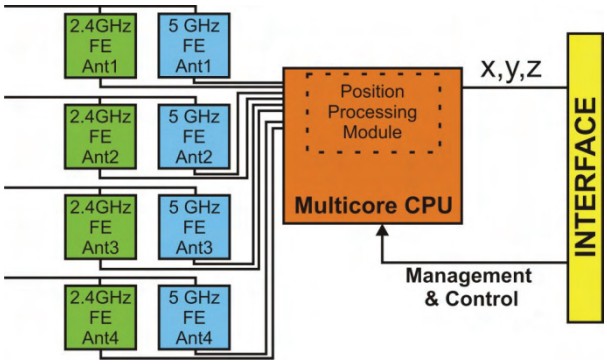


Fig.3. Block diagram of multiple antenna external module

For more secure and accurate computations the system can use a phase differential and amplitude processing from four antenas.When the signals are received from antennas, they are processed in front-end modules in RF part of the external module. These front-end modules convert the analog signal streams into digital format, which is proper for further processing. After front-end modules, data streams are processed into a multi-channel, multi-core CPU that calculates the 3D coordinates. External module outputs low-noise phase/amplitude measurements from 4 antennas. Its interface is used also for feedback control, which is applied direct to the multicore CPU.In this way it is possible to achieve full software configuration management over the real time measurements and processing.

The suggested external module can be assembled on a single small-size board. It is very important according portability of the whole system. External part is consistent also with weight and consumption limitations. Most of the modern multicore processors can scale their power consumption according to their workload. Another advantage is that duplication of hardware modules is minimized and it reduces the price.

In some applications, that use the same principles, one of the antennas is accepted as a main and the others are auxiliary [6]. It is good approach, but only if we need some extra measurements like speed, attitude, deformation of objects.In our case is better to give an account of thephase differences between all the couples of antennas as presented in a table 1.

TABLE I
COMPARISON OF THE PHASES IN ALL POSSIBLE COUPLES OF ANTENNAS – IN CASE OF FOUR ANTENNAS USED

Comparison combinations	Phase I	Phase II
1	φ1	φ2
2	φ1	φ3
3	φ1	φ4
4	φ2	φ3
5	φ2	φ4
6	φ3	φ4

Each of these combinations can be decided as a specific coordinate of the source. On the other side the strength of a signal on every antenna can be visualized as a sphere in the 3D space, and the place where all the spheres cross each other, shows the area of the signal source.

Many sources [4][5], describe wireless network propagation the same way like the light's propagation. The reason is that they both have wave nature. According to this, it is correct to explain that some phenomena present in our situation too. Diffraction, refraction, interference, and attenuation have influence over wireless waves. Practically the result from all these can be observed as a rapid change in the strength of a signal, when moving slowly in a closed area. When building and using described system, we should give account of the fact that the signal on the straight ray between two points has the biggest amplitude if the frequency modulation is used. This rule gives possibility easy to filter the desired signal from source.

When coordinates are already known, two different methods can take place for protecting the connection. First of them can be called "jump detection" and was explained above. It uses detection of rapid change in a phase and amplitude of the received signal. The second method is based on declaring coordinates for the access point area. The administrator can describe the real 3D-area of the opposite communication point and deny access from all other points. It gives high efficiency of the security level.

The coordinates from external module can be integrated into a working simulator or can be used for developing a new one. It is important to note that multicore CPU's are already used in all new computers and the scientists are trying to find new algorithms for parallelization of the processed data in order to improve the performance. Presented system can easily take advantage of multicore technologies not only in external module, because the received signals are highly parallelized by nature. Every antenna is a source of one or more data stream(s). In external module this streams form

different processing threads and are processed using the same instructions. It means that so called SIMD (singe instruction multiple data) computers are proper for the system. At the end of the system all computations present a common result – coordinates and direction of the wave source. If this result should be in a graphic form, this also consumes extra computation power. Then internal multicore CPU can help in visualizing the environment. Other proper technology that is very implemented now and is proper for this system is CUDA (compute unified device architecture). This technology gives possibility to achieve some extra performance, using GPU's resources of the computer.

IV. Practical Issues

The same method as described is used in a military device for finding enemies. It uses multiple microphones placed in different points around soldiers. If enemy shoots, is very hard to find his position if soldiers don't see the gun fire and just hear the bullets. If distance is long enough, bullets come first and then the sound from shooting weapon. The microphones record the sounds, they receive from flying bullets, send them for analysis and results trace the enemy position.

Another system that uses the same principles is used for determination of attitude of the rover based on the L1 signal carrier phase differences from multiple GPS receivers [6].

V. Conclusion

The suggested system is a hardware decision that gives a high rate enhanced security of the connections. It can be used in one side of the communication channel as well as in both sides if higher protection is required. The system is compatible with new hardware technologies and can be easy integrated into software environments. It can be used also in e-learning courses for building and exploring real wireless networks.

Acknowledgement

This paper is financed by project: "Creative Development Support of Doctoral Students, Post-Doctoral and Young Researches in the Field of Computer Science", BG 051PO001-3.3.04/13, EUROPEAN SOCIAL FUND 2007–2013. OPERATIONAL PROGRAMME "HUMAN RESOURCES DEVELOPMENT"

References

[1] Angelov M,"Cells network bases." PAN- VT, Veliko Tarnovo, Bulgaria, 2000

[2] Kalushkov T., Borovska P., "Choosing Of Wireless End-user Passwords. Dual Language Encryption Method." SAI'10 Conference, Sofia, Bulgaria, 2010

[3] Lashkari A., Mansoori M., Danesh A., "Wired Equivalent Privacy (WEP) versus Wi-Fi Protected Access (WPA)." International Conference on Signal Processing Systems, 2009

[4] Schmitz A., Rick T., Karolski T., Kuhlen T., Kobbelt L., "Simulation of RadioWave Propagation by Beam Tracing." Eurographics Symposium on Parallel Graphics and Visualization, 2009

[5] Madej P., "3D Wireless Networks Simulator- visualization of Radio Frequency propagation for WLANs." A dissertation submitted to theUniversity of Dublin, TrinityCollege, in a partial fulfilment of the requirements for the degree of Master of Science in Computer Science, 2006

[6] SimskyA.,Vander Kuylen L., Boon F., "Single-board Attitude Determination System Based on the PolaRx2@ GPS Receiver."ENC GNSS 2005, Munich, Germany, 2005

Frequency Selective Method for Measurement and Estimation of Electromagnetic Emissions

Boncho Bonev[1], Ilia Iliev[2],
Borislav Pankov[3], Kliment Angelov[4] and Vladimir Poulkov[5]

Abstract – **In this paperamethod for measurement and integral estimation of electromagnetic emissions(EME) in VHF and UHF is presented.The approach is to divide the evaluated frequency range of sub-bands and the measurement to be performed with the maximum frequency resolution. The objective is to increase the accuracy and traceability of the measurements and respective assessments.The method issuitable for on-site (ad hoc) estimations and takes into consideration the internationally standardized EME measurement requirements.Results of measurements are presented and analysed.**

*Keywords:***Electromagnetic Emission, electromagnetic compatibility, VHF, UHF, RF measurement,**

I. INTRODUCTION

The development of communication technologies during the last decades and especially of mobile communications (GSM, GPRS, UMTS, TETRA, Wi-Fi, WiMAX, and expected LTE) lead to a significant increase of the electromagnetic emissions (EME) from telecommunication sources.Electromagnetic emissions find application in many other spheres of human life, such as medicine, industrial technologies, etc., which also contribute to the overall increase of the electromagnetic background level. This maintains the interest to this topic not only from a scientific point of view, but also because of the sensitivity of the society concerning the influence of electromagnetic emissionsover human health and especially the ones from the base stations of the mobile networks.

The problems of electromagnetic compatibility (EMC) and the admissible levels of the Electromagnetic Fields (EMF) generated from different types of communication systems is permanently a topic of research and analysis. World-wide there are legal regulations and standards defining the maximum admissible levels of EME for the different frequency ranges.According to these regulations, in the process of putting into operation, commissioning, maintenance and monitoring oftelecommunication systems radiating in the radio spectrum, measurement and control of the EME is required, so that the radiation from such systems does not exceed the maximumadmissible levels.

Monitoring is one of the basic ways for control of the levels of EME from different sources in an area or region. It can be done using two basic approaches;first, by the installation of stationary sensors, orsecond, as periodic “ad hoc” measurements, andstoring in a data base and analyzing the results for the levels of the EME.

The first way gives the opportunity of performing constant monitoring and analysis of the results in predefined points, but requires significant financial resources and time for the development and building up the networkof sensors. The second is based upon periodical measurements and analysis of the electromagnetic field and its variations in predefined “hot spots”, which are of special attention and/or social interest, as for example hospitals, kindergartens, schools, universities etc.This approach allows higher accuracy of the measurements, requires less financial and time resources, but does not give the possibility of a long time continuous monitoring of the electromagnetic field and its variations at a specific point. Such measurements are performed with the help of specialized measuring equipment and for the analysis of the electromagnetic field and the estimation of its long term expected variations are applied statistical approaches. The measurements could be performed either with equipment, giving an integral estimation of the electromagnetic radiation, or with the help of a frequency selective equipment, which gives the possibility of estimating the contribution of each of the telecommunication technologies to the total level of EME in a given point or area.

In this paper a method for the measurement and estimation of EME, with the help of a frequency selective type of equipment is considered. The method is in conformance with European and national legislations for the maximum admissible levels of exposure of the population to EME in populated areas and working conditions. The specifics of the telecommunication applications emitting in the different frequency ranges are taken into consideration, as well as the resolutionof the measurement equipment. In order to increase the measurement accuracy, in the case of integral estimation of EME, a segmentation of the frequency range into sub-rangesit is proposed. The measurements must be performed with the highest frequency resolution of the equipment. Such a methodology is applicable in the VHF and UHF ranges, with are basically used by telecommunication technologies today.

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

[2]Ilia Ilievis with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski Blvd. 8, 1756 Sofia, Bulgaria, E-mail: igiliev@tu-sofia.bg.

[3]Borislav Pankov is with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski Blvd. 8, 1756 Sofia, Bulgaria, E-mail: bpankov@tu-sofia.bg.

[4]Kliment Angelov is with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski Blvd. 8, 1756 Sofia, Bulgaria, E-mail: kna@tu-sofia.bg.

[5]Vladimir Poulkov is with the Faculty of Telecommunications, Technical University of Sofia, Kliment Ohridski Blvd. 8, 1756 Sofia, Bulgaria, E-mail: vkp@tu-sofia.bg

II. MEASUREMENT AND ESTIMATION METHODOLOGY

The proposed method is in conformance with the maximum allowable levels of the intensity of the electrical fields and power density of the electromagnetic fields (EMF) in populated areasand working conditionswhich are defined in European and national legislations [1,2,3,4].According to the principles of superposition the intensity of the fields from different sources in the VHF range are summed geometrically for obtaining an integral estimation of the intensity of the field in the range. In the UHF range, it could be assumed that the sources are uncorrelated, thus the resulting energy flux could be obtained as an algebraic sum of the fluxes from all of the separate sources [5,6].

The proposed methodology is applicable for an integral estimation of the EME emissions in the VHFandUHFranges, generated by radio andalso other sources. It supposes the measurements to be done with a frequency selective type of equipment in thefar field zone of the transmitting antennas.

A. Measurements in theVHFrange

The measurements of the levels ofEME emissions in this range are performed in the integral estimation mode of the measurement equipment. The measurement is performed applying the smallest possible Resolution Bandwidth (RBW) of the equipment, but in any case it should be at least 300kHz. The time of the measurement is not less than 6 minutes. In this frequency range the major sources are the VHF-FM radio stations and TV transmitters. These technologies use a bandwidthwhich is higher or commensurable to the one used in the measurement, whichguarantees enough accuracy. This range is used also by narrowband radio stations for special services which are transmitting in the ranges of 50MHzand 160MHz, but their emitted powers are significantly lower than those of the above transmitters. Moreover, their emissions usually are not constantor they are sporadic, taking into account the type of their users. In the measurements, the frequencies for which there are maximums of the field intensity and theirabsolute values are considered.

B. Measurements in the UHF range

This range is used for a number of telecommunication technologies, including the ones from the mobile networks, which have the major contributions to the increase of the total level of the electromagnetic field. The bandwidth of the radio channels, used for the different applications varies significantly,from 200 kHzforGSM, up to 5 MHzforUMTSand8 MHz for analogue TV broadcasting. This requires the separation of this frequency range into sub-ranges, in order to increase the accuracy of the measurement, because usually the frequency selective equipment does not allow small RBW when measuring the whole frequency range. We suggest the division of sub-ranges to be performed according to the frequency intervals given in Table I.

For the determination of the RBWfor each sub-range the specifics of the basicoperating technologies are taken into consideration. Some of them are recommended in the respective documents, as for example is the case withGSM-900,GSM-1800, UMTS[1,8,9,10]. The measurement of each sub-range is performed using the integral estimation mode of the equipment and lasts not less than 6 minutes. For each sub-range the measurements are done for at least two frequencies, witha maximum of the energy flux. This is necessary in order to analyze the sources, which have biggest contribution in the resulting integral values of the energy flux.

TABLE I

No.	Frequency	RBW_{max}	Basic Technologies
1.	300 – 880 MHz	200kHz	Analog TV, DVB-T, TETRA
2.	880-960 MHz	50kHz	GSM-900
3.	960 – 1710 MHz	300kHz	Military
4.	1710 – 1880 MHz	50kHz	GSM-1800
5.	1880 – 2400 MHz	200kHz	UMTS
6.	2400 – 3000 MHz	200kHz	Wi-Fi

C. Integral estimation of the level of electromagnetic fields

The integral estimation of the total influence of all of the sources from the ranges in Table I is done using the following relation:

$$\frac{E_{VHF}^2}{E_{\max VHF}^2}+\frac{S_{\Sigma UHF}}{S_{\max UHF}}\le 1, \quad (1)$$

where $E_{\max VHF}$[V/m]and $S_{\max UHF}$[μW/cm^2] are the maximum admissible values of the EME for the frequency ranges according to the national regulations and standards.In order to determine the sum of the power density in the UHF range, the following relation is used:

$$S_{\Sigma UHF}=\sum_{i=1}^{6}S_i\ , \quad (2)$$

where$S_i$are the measured values of the energy fluxfor each of the sub-ranges according to Table I.

When people are working in the vicinity of radio emitting equipment, the maximum permissible time for staying in a given area is determined by the formula [4]:

$$T_{\max}=\frac{W_{S\max}}{S_\Sigma}, \quad (3)$$

wherethe value of W_{Smax}[μW.h/cm^2] is alsostandardized and is dependent on the type of the antenna system.$S_\Sigma$is calculated using the equation,

$$S_\Sigma=\frac{E_{VHF}^2}{120\pi}+S_{\Sigma UHF}\ . \quad (4)$$

III. MEASUREMENT RESULTS

The proposed method is tested in the measurement and estimation of the level of the EME emissions in more than 20 points of a densely populated city area. The measurements are performed with a frequency selective equipment. For comparison, integral measurement results in one of the points in the whole frequency UHF rangewith the minimum possible RBW of 1 MHz (without division into sub-ranges), are presented in Table II.

TABLE II

E_{VHF}, mV/m	$S_{300\text{-}880\ MHz}$, μW/cm^2	$S_{880\text{-}960\ MHz}$, μW/cm^2	$S_{960\text{-}1710\ MHz}$, μW/cm^2	$S_{1710\text{-}1880\ MHz}$, μW/cm^2	$S_{1880\text{-}2400\ MHz}$, μW/cm^2	$S_{2400\text{-}3000\ MHz}$, μW/cm^2	$S_{\Sigma UHF}$*, μW/cm^2	$S_{\Sigma UHF}$**, μW/cm^2
199	$5{,}89.10^{-3}$	0,0223	$1{,}33.10^{-3}$	0,0159	0,0315	$6{,}1.10^{-3}$	0,08302	0,0645

* $S_{\Sigma UHF}$ calculated from Eq. (2).
** $S_{\Sigma UHF}$ measured with RBW=1 MHz.

The results from Table IIshow, that the values calculated using equation (2) and the onesmeasured with the equipment (RBW=1MHz)have a difference of about28,8%. Similar is the result (approx. 30 %) for the rest of the measured points. This is due to the fact, that in the case of integral measurement in the whole UHF frequency range, when RBW=1MHz is fixed, many of the maximum values could be “skipped”. In this case the integral estimation of the level of the EME will be with lower accuracy, due to the fact that some of the telecommunication technologies (and especially GSM, which has the major contribution in the total level of the emissions – Fig.1) use more narrow-band frequency channels.

A chart of the relative contributions of the separate sub-ranges to the total overall level of the electromagnetic field based on the performed measurements and proposed method, of estimation, is presented in Fig.1. As it could be seen, the biggest relative contributions to the level of the EME in the UHF range have the technologies GSM-900 andGSM-1800. This is easy to explain, as the measured base stations are located in the city area and create bigger intensity of the field in the measurement points. The measurements show also, that the power density in the ranges 300-880 MHzand 960-1710 MHzis relatively constant, respectively 5-7nW/cm^{2}and 1-2 nW/cm^2.

IV. CONCLUSION

In this paper a method for measurement and estimation of the electromagnetic emissions in the “far field zone” from transmitting antennas or sources of electromagnetic fields is proposed. Using a frequency selective equipment, the methodology is applicable for the integral estimation of the electromagnetic emissions in the VHF and UHF ranges, generated from radio and other sources. A division of the measured frequency range into sub-ranges is proposed andsuggested the measurements to be performed with the maximum frequency resolution of the measuring equipment. The goal is to increase the accuracy of the measurements and a more precise tracing of the estimations to be obtained.

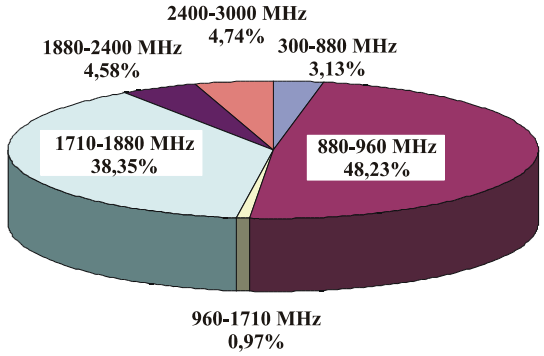


Figure 1. Relative contributions of the separate sub-bands in the UHFrange to the total level of EMF

The proposed method is approbated through the measurement and integral estimation of the EME in more than 20 points in a densely populated urban area. After summing up and analyzing the results, it is shown that:

1. The division into sub-ranges increases the accuracy of the integral estimation;
2. The biggest relative parts of the EME in the UHF range have the technologies GSM-900andGSM-1800, followed by UMTSandWi-Fi;
3. The power density in the ranges 300-860 MHzand 960-1710 MHzis relatively constant.

REFERENCES

[1] Council Recommendation of 12 July 1999 on the limitation of exposure of the general public to electromagnetic fields (0 Hz to 300 GHz) (1999/519/EC).

[2] Ordinance No. 9 (1991) for the maximum admissible levels of EMF in populated areas and determination of the hygienic zones around emitting sources. Published in the Bulgarian State Newspaper (BSN) No.35/1991. Last revised in BSN 8/2002.

[3] Ordinance No. 7 (1999) for the minimum requirements of healthy and safe conditions for work at the working places and when using working equipment. Published in Bulgarian State Newspaper (BSN) No.88/1999. Last revised in BSN 40/2002.

[4] Bulgarian National Standard 14525-90. Labor safety. Electromagnetic Radiofrequency Fields. Admissible levels and control requirements.

[5] Petrov N., B. Boychev, T. Petkov, M. Petrov, Measurement and estimation of electromagnetic emission of operating environment in ATC, XV International Science and Technical Conference “trans&MOTAUTO '08”, 18-20.09 2008, Sozopol, Bulgaria, Mashines, Technologies, Materials, iss. 8-9/2008, pp. 83-84, ISSN 1313-0226. (in Bulgarian)

[6] Gotsis A., N. Papanikolaou, D. Komnakos, A. Yalofas, P. Constantinou, Non-ionizing electromagnetic radiation monitoring in Greece, Ann. Telecommun., Vol. 63, pp. 109 - 123, 2008.
[7] Revised ECC Recommendation (02)04: Measuring non-ionizing electromagnetic radiation (9 kHz – 300 GHz). Electronic Communications Committee (ECC) within the European Conference of Postal and Telecommuni-cations Administrations (CEPT). Edition October, 2003.
[8] Mobilfunk-Basisstationen (GSM), Messempfehlung. Schweizerisches Bundesamt für Umwelt, Wald und Landschaft BUWAL und Bundesamt für Metrologie und Akkreditierung METAS, Bern, 2002.
[9] GSM measurements with the Selective Radiation Meter SRM-3000, AN_HF_1005_GSM Measurements, Narda Safety Test Solutions GmbH, 2007
[10] UMTS measurements with the Selective Radiation Meter SRM-3000, AN_HF_1007_UMTS_Measurements, Narda Safety Test Solutions GmbH, 2008

Fractal Designed Antennas Matching

Kliment Angelov[1], Boncho Bonev[2], Petko Simeonov[3] and Radoslav Tsochev[4]

Abstract – **This study describes an experimental research of a fractal designed antenna matching. Several antenna prototypes shaped as Koch's curves have been made for the performing of measurements. The optimal dimensions and the value of the antenna input reactance have been specified using the theory of planned experiment.**

Keywords – **Fractal Antennas, Koch's curve, Antenna Matching.**

I. INTRODUCTION

The Fractals are geometric objects that are not subject to ordinary rules of Euclidean geometry. Despite of their widespread deployment in nature, they receive serious scientific attention in the last quarter of last century (Benoit Mandelbrot introduced the concept of fractal in 1975). The Fractal has a number of interesting characteristics, among which the self-similarity and infinite particularity that is independent of scale and diverse spatial dimension. It is these interesting features of the fractal shaped antenna systems that lead to the interesting features of the antennas themselves. Multi-band reception / transmission are a direct consequence of self-similarity and the variable spatial dimension determines the decreasing physical size of antennas. The listed accents determine the interest of many authors in this subject [1] [2] [3] [4]. It would be of practical interest to determine the possibility of using such antenna systems in contemporary radio apparatus, and for the purpose parameters such as resonant frequencies, input impedance, etc. should be evaluated. The objective of this work is to explore the input impedance and the impedance matching of fractal antenna.

II. DESCRIPTION OF THE PROBLEM

A. Theoretical basis

The Koch's curve is one of the most typical representatives of the so-called deterministic fractals. It is obtained by a recursive process, using a generator that replaces the base of each iteration. The generator and the base are shown in Figure 1 and curves of the first three iterations - in Figure 2.

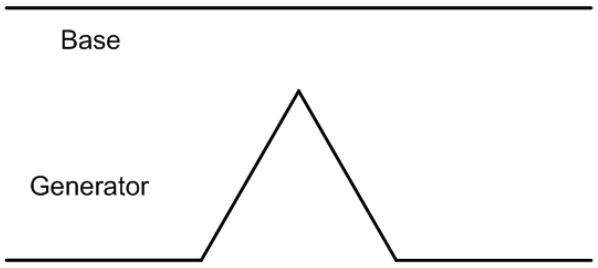


Fig. 1. Base and Generator of Koch's Curve Fractal

The fractal dimension is **ln 4 / ln 3 ≈ 1.26**. The base of the Koch's curve is a straight line, which then allows the comparison of the experimental data with those for the quarter-wave dipole.

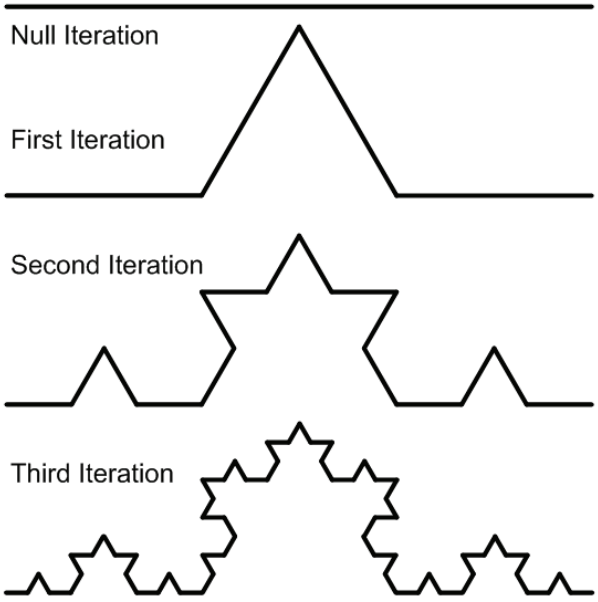


Fig. 2. Koch's Curve Fractal (initial iterations)

B. Experimental

Experimental data were obtained by testing of models of antenna with shape of curve of the first iteration shown in Figure 2. The experiments were made on the stand, whose block diagram is given in Figure 3. Through directional diverter the direct wave is separated from the reflected wave and a scalar analyzer is used to examine the value of the standing wave ratio (SWR). The dipole power occurs through

[1]Kliment Angelov is with the Technical University of Sofia, Kliment Ohridski Str. 8, 1000 Sofia, Bulgaria, E-mail: kna@tu-sofia.bg.
[2]Boncho Bonev is with the Technical University of Sofia, Kliment Ohridski Str. 8, 1000 Sofia, Bulgaria, E-mail: bgb@tu-sofia.bg.
[3]Petko Simeonov is with the Technical University of Sofia, Kliment Ohridski Str. 8, 1000 Sofia, Bulgaria, E-mail: psimeonov@tu-sofia.bg.
[4]Rodoslav Tsochev is with the Technical University of Sofia, Kliment Ohridski Str. 8, 1000 Sofia, Bulgaria, E-mail: radoslav_tsochev@abv.bg.

capacitive element (capacitor), which serves to compensate the reactive component of its input impedance.

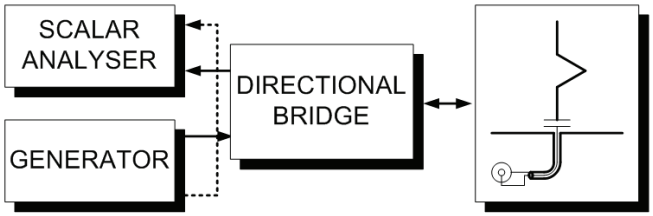


Fig. 3. Experimental Model Diagram

To achieve better matching of the antenna for a given frequency it is necessary to vary the magnitude of the capacitor and the dimension of the radiating element. Such optimization can be done using the theory of planned experiment [5]. The task comes to a full factorial experiment with two factors, ranging on three levels. Nine separate tests should be performed and the alteration of the SWR parameter within the bounds of the factorial space can be defined. This dependence is given by the expression [5]:

$$Y = b_0 + b_1.x_1 + b_2.x_2 + b_{12}.x_1.x_2 + \\ + b_{11}.x_1^2 + b_{22}.x_2^2, \quad (1)$$

where, by x_1 and x_2 the two factors that determine the alteration of the parameter Y are denoted. In the present case these are the length l of one arm of the radiating element and the value of the capacity of the capacitor C, which are presented in Figure 4.

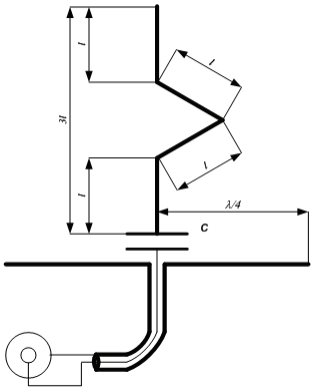


Fig. 4. Prototype dimensions

The figure shows the schematic representation of the experimental prototype. The emitting element is placed on stand, which has a horizontal metal plane made of conductive metallic plate. The dimensions of the plate are chosen to ensure a minimum distance of a quarter of the operating wavelength in all horizontal directions. Thus the antenna system of the type Ground Plane is realized; this system has a corresponding fractal copy as a dipole.

The parameter Y in formula (1) is the SWR in the supply feeder. The coefficients b, that give the weight of impact of individual factors and combinations of factors are determined by the relations [5]:

$$b_0 = \frac{5}{9}\sum_{j=1}^{9} x_{0j}.\bar{y}_j - \frac{1}{3}\sum_{i=1}^{2}\sum_{j=1}^{9} x_{ij}^2.\bar{y}_j, \quad (2)$$

$$b_i = \frac{1}{6}.\sum_{j=1}^{9} x_j.\bar{y}_j, \quad (3)$$

$$b_{ik} = \frac{1}{4}.\sum_{i=1}^{2}.\sum_{j=1}^{9} x_{ij}.x_{kj}.\bar{y}_j, \quad (4)$$

$$b_{ii} = \frac{1}{2}.\sum_{i=1}^{2}\sum_{j=1}^{9} x_{ij}^2.y_j - \frac{1}{3}\sum_{j=1}^{9}\bar{y}_j, \quad (5)$$

III. Results

For the purposes of the experiment the resonant frequency of the antenna is considered to be 434 MHz; this frequency is included in one of the ISM bands within the band of decimal waves (UHF). The order of the size of the antenna and the value of the capacity of the capacitor are determined by a rough preliminary experiment. Based on these initial data, the planned levels of ranging of the factors are chosen and the relevant values are listed in Table 1.

TABLE I

Vary Levels	x_1	x_2
-1	52,5 mm	1,2 pF
0	55,0 mm	2,4 pF
+1	57,5 mm	3,6 pF

Nine experimental tests were performed with different combinations of values for the factors. The values of SWR for each case have been measured. The results are shown in Table 2.

TABLE II

i / j	x_1 / l, mm	x_2 / C, pF	Y_E
1	+1 / 57,5	+1 / 3,6	3,00
2	+1 / 57,5	0 / 2,4	1,46
3	+1 / 57,5	-1 / 1,2	8,25
4	0 / 55,0	+1 / 3,6	1,50
5	0 / 55,0	0 / 2,4	2,40
6	0 / 55,0	-1 / 1,2	15,70
7	-1 / 52,5	+1 / 3,6	3,74
8	-1 / 52,5	0 / 2,4	10,50
9	-1 / 52,5	-1 / 1,2	35,00

From Equations (2), (3), (4) and (5) the coefficients reflecting the influence of individual factors, as well as the combination of them upon the alternating of the parameter ***Y***, can be derived. The concrete values are given in Table 3.

TABLE III

b_0	2,2589
b_1	-8,4517
b_2	-6,0883
b_{12}	6,5025
b_{11}	6,4117
b_{22}	3,7917

From the values of the coefficients it can be noted, that with increasing of the length ***l*** and the value of the capacitor ***C***, the value of SWR decreases. On the other hand there is a straight proportional impact of these factors in the second degree, as that of the value of capacitor ***C*** is twice less than that of the length ***l***. The straight proportional mixed influence, reflected by the coefficient b_{12} is also significant

Using Equation (1) the dependency of the parameter ***Y*** within the factorial space, can be built, which in this case is given by the expression:

$$Y = 2{,}2589 - 8{,}4517.x_1 - 6{,}0883.x_2 + \\ + 6{,}5025.x_1.x_2 + 6{,}4117.x_1^2 + 3{,}7917.x_2^2 . \quad (6)$$

Figure 5 depicts the dependency (6), obtained using the software MatLab [6].

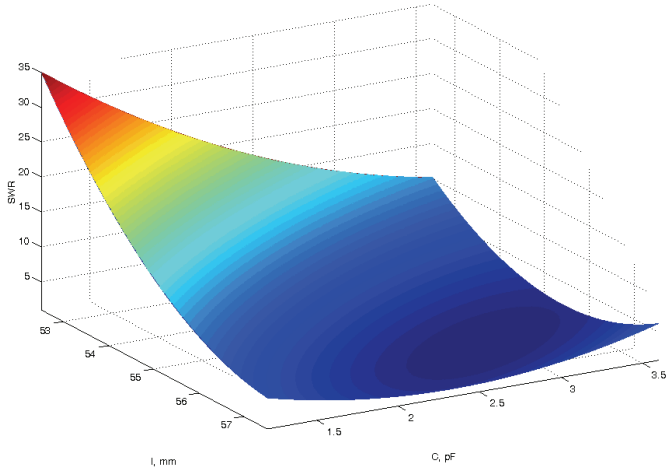


Fig. 5. Model Visualization

Several checks were made of the model (6) for determination the significance of the coefficients and the adequacy of the model. Using the criterion of W. G. Cochran it was proved that all coefficients in the model are significant. Therefore, to assess the adequacy of the model the t-criterion of Student (W. S. Gosset) is applied, with which the ***t*** value was calculated using the formula:

$$t = \frac{|Y_0 - b_0|\sqrt{m}}{\sqrt{\frac{1}{m}.\sum_{j=1}^{m} S_{Y_j}^2}} , \quad (7)$$

where:

Y_0 – the value of the parameter for null coded values of the factors (sample j=5);

m – number of samples;

S_{Yj} – dispersion of the values of the parameter within the bounds of individual sample.

The values of the parameter of empirical data Y_E and the corresponding values Y_M obtained by model (6) are given in Table IV. In the same table the values of dispersion S_{Yj} of the parameter within the individual sample are presented, which are calculated by the formula:

$$S_{Y_j} = Y_{E_j} - Y_{M_j} . \quad (8)$$

TABLE IV

j	Y_E	Y_M	S_{Yj}
1	3,00	6,27	-3,27
2	1,46	1,86	-0,46
3	8,25	10,12	-1,87
4	1,50	2,11	-0,61
5	2,40	4,13	-1,73
6	15,70	18,81	-3,11
7	3,74	5,44	-1,7
8	10,50	13,89	-3,39
9	35,00	35,00	0

The values of parameter Y_M are normalized, as it is assumed that its lower level is equal to the lowest possible value of SWR, thus avoiding the irregular values for SWR <1.

The calculated value for the criterion of Student is $t = 0,197$. The verification on inequality $t = 0,197 < t_{0,02;9} = 2,821$ proves that the model (6) is adequate. The value $t_{0,02;9} = 2,821$ is taken from table [5].

The result of performed experiment and the obtained model localize an extremum of the alternation of parameter Y. For values of the factors $x_1 = 0,420$ and $x_2 = 0,445$ that correspond to the length of the one arm of the radiating element $l = 56,05$ mm and capacity value of the capacitor $C = 2,935$ pF the minimum for the value of the SWR is obtained at frequency of 434 MHz. This means that at these values for the factors, the antenna will have the best emitting properties. Taking into consideration the operating frequency we can find the relative dimensions of the antenna as parts of the operating wavelength. For the length l, the value $l = 0,0811.\lambda$ is obtained. In that way the reactive component of the input impedance of the antenna can be determined. Considering that the capacitor is connected in series with the radiating element, then for the resonant frequency of the antenna radiation their reactive components compensate each other. Accordingly, the reactive component of the input impedance is equal to the impedance of the capacitor, but with reverse mark. From the obtained data, it follows that this component has inductive character and a value determined by the dependency:

$$X_a = +j\frac{1}{2\pi fC} = +j124,95\Omega\,. \qquad (9)$$

IV. CONCLUSION

As a result of the performed studies, based on the theory of the planned experiment, an optimization of the dimensions of the fractal antenna is achieved as well as of the capacity of the matching capacitor. A value $0,0811.\lambda$ was determined for the length of one arm of the fractal element. The value $-j124,95\Omega$ was also determined as the impedance for the matching reactive element. The achieved local extremum - minimum of the SWR parameter can be realized with technological feasible values of the factors.

REFERENCES

[1] Petkov P., "Fractal Antenna Arays", Trans. Sc. Conf. “Telecom'2007”, St. Constantine - Varna, Bulgaria, Oct. 11-12, 2007.

[2] Lemecki A.,P. Debicki, "Broadband properties of a minkowski fractal curve antenna", MIKON-2002. 14th International Conference on Volume 3, Issue , 2002 Page(s): 785 - 788 vol.3.

[3] Puente C.,J. Romeu, A. Cardama "The Koch monopole : A~Small Fractal Antenna", IEEE Transactions on Antennas and Propagation, Vol 48, No. 11, November 2000.

[4] Angelov K., P. Petkov, B. Bonev, Analysis of Fractal Designed Antennas, Sc. Conf. “Telecom'2008”, St. Constantine - Varna, Bulgaria, Oct. 9-10, 2008.

[5] Vuchkov I.N., Bojadjieva L.N., Quality Improvement with Design of Experiments. A response Surface Approach, “Kluwer Academic Publishers”, Dordrecht, Netherland, 2001.

[6] MATLAB IMAGE PROCESSING TOOLBOX. User's Guide, "The Math - Works Inc.", 2000. www.mathworks.com

Experimental Setup for BER Measuring of Free Space Optical System

Nikolai Kolev[1] Kalin Dimitrov[2] Yuliyan Velchev[3] Tsvetan Mitsev[4]

Abstract – **The paper describes measuring of BER during data transmission in FSO system. The used transmitting and receiving electronic blocks are connected in suitable manner with the designed by the authors BER testing system. The system is installed on the blocks 1 and 2 of the TU-Sofia. Preliminary results about system availability are presented.**

Keywords– **Bit Error Rate, Free Space Optics, Bit Error Rate Tester.**

I. INTRODUCTION

The bit error ratio (BER) is a measure of the percentage of bits that a system does not transmit or receive correctly. It is a dimensionless number ranging from 0 to 1. If the BER = 0, then all bits are transmitted correctly at the other extreme, if the BER = 1, every bit is received in error. Every transmission system (and every part of it) has an intrinsic bit error ratio, which can take on any real number between 0 and 1. The exact value may change, for example, with temperature or operating voltage, but it's a fundamental system property[1].

$$BER=Nerr/Nall \quad (1)$$

whereNerr - number of mistaken bits, Nall - number of all bits. It's very important constructed FSO system to be tested for BER in real atmospheric conditions [2, 3].

II. SCHEMATIC DESIGN

The device which is measuring the level of error BER is called BERT bit error rate tester. This article describes exactly such a scheme for measuring the BER [4].

Block diagram and real photos of the system is shown in Fig.1.The information is transmitted simultaneously on two channels one is atmospheric and the other is a conductor.

Once the information passing through both channels of their output signals are compared in the block BER measure. If there is a difference in the two signals the counter COUNTER shows an counts on the display. Since the delay of the signal varies in both channels they are connected delay elementsto thescheme. Delays are implemented in two ways.

The random number generator is realized by shift registers and logic elements exclusive OR (XOR) shown in Fig.2.The

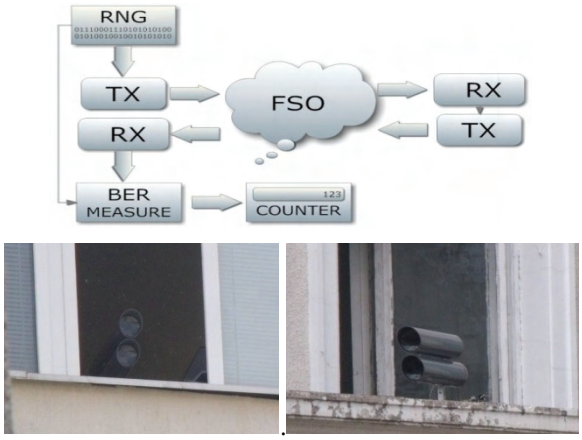


Fig. 1. Block scheme and real pictures of FSO

information is shifted on every tick of the clock generator from left to right, and last but one and last digit are connected to both inputs of the logical element XOR. After completion the logical operation result is returned to the first digit of the

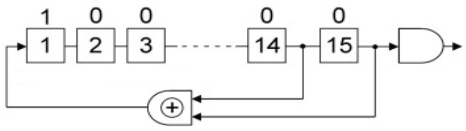


Fig. 2. Scheme of random number generator

shift registry, etc.

Fig.3 shows the principal realization of the random number generator.The shift registers are realized with integrated circuit 74HCT174 and a logic element exclusive OR with 74HC86. The circuit is supplied with voltage 5V. Clock signal is necessary for normal operation of the scheme. It is fed to the ninth pins of integrated circuits 74HC174.

Thecoarse delays of n number of ticks is implemented with D flip flops Fig.4. Depending on the required number of ticks of delay a different number of D flip flops are used [5].For shorter delays are used non-inverting buffers.

For comparison of two signals from the optical path and conductorare used logic element exclusive OR (XOR) Fig.5. To avoid error in the divergence of the two fronts of the signal after XOR logic element is added a logic element AND is added, comparing the two signals only when the clock signal is in high level.

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

[2]Kalin Dimitrov is with the Faculty of Telecommunications at Technical University of Sofia, E-mail: kld@tu-sofia.bg.

[3]Yuliyan Velchev is with the Faculty of Telecommunications at Technical University of Sofia.E-mail: julian_velchev@abv.bg.

[4]Tsvetan Mitsev is with the Faculty of Telecommunications at Technical University of Sofia.E-mail: mitzev@tu-sofia.bg.

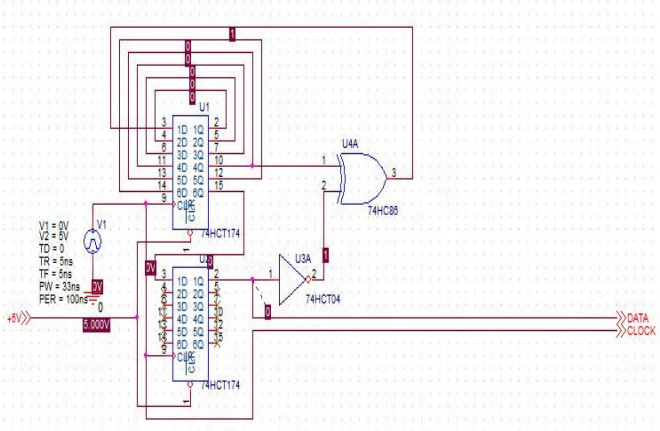


Fig. 3. Principal realization of random number generator

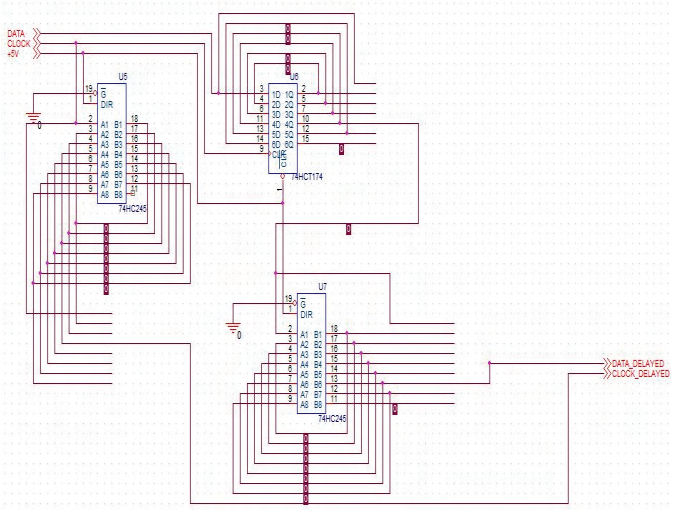


Fig. 4.Principal realization ofscheme for long delays

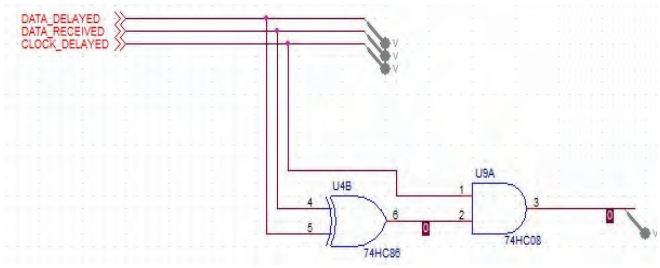


Fig. 5. Principal realization of comparing scheme

III. SIMULATION

Fig.6 and Fig.7 shows two computer simulations. In Fig.6 can be seen an error caused by misalignment of the two edges of the pulses: the first is the transmitted data and the second is the delayed original data. In Fig.7 the signals are aligned in time at which the error does not occur. Error signal at output of the scheme is low level.

The speed of FSO link is 10Mbps. The distance between Rx and Tx is 180 m. The receiver bandwidth is 16MHz. The spot size at the receiver is 900mm. The receiver aperture is 90mm.

We left the system to work for a week, but the whether conditions prevented the accumulation of any error.The

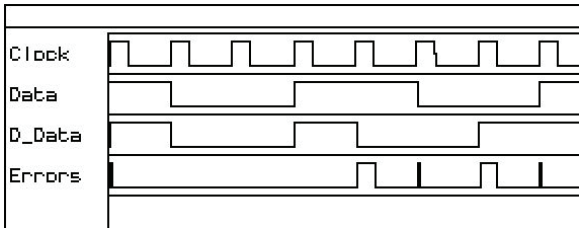


Fig. 6. Computer simulation of errors

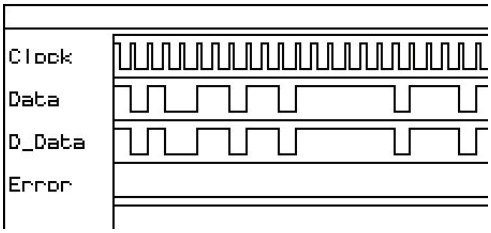


Fig. 7. Computer simulation without errors

nominal working distance of the system is 1km, but in the particular case we worked at 180 m. This also prevents accumulation of error.

IV. CONCLUSION

The described scheme can be appliedfor studding the dependence of the level of bit error of Free Space Optics systems and the dependency of destabilizing factors such as atmospheric conditions (fog, snow, rain, etc.). [6, 7]. In future we plan to work at nominal distance and with bigger time intervals witch will enable accumulation of error.

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

REFERENCES

[1] D. Derickson, M. Müller, Digital Communications Test and Measurement: High-Speed Physical Layer Characterization, Prentice Hall, 2007.

[2] E. Leitgeb, M. Löschnigg, U. Birnbacher, G. Schwarz, A. Merdonig, High Reliable Optical Wireless Links for the Last Mile Access, pp. 178 - 183, ICTON, IEEE, 2008.

[3] D. Pham Tien, et al, A study on transmission of RF signals over a turbulent free space optical link, MWP/APMP, pp. 173-176, IEEE, 2008.

[4] N. Kularatna, Digital and analogue instrumentation: testing and measurement, MPG Books Limited UK, pp.173 - 176,2008.

[5] M. Rafiquzzaman, Fundamentals of digital logic and microcomputer design , John Wiley & Sons, New Jersey, 2005

[6] H. Willebrand, B. Ghuman, Free Space Optics: Enabling Optical Connectivity in Today's Networks, Indianapolis, SAMS, 2001.

[7] R. Kvicala, V. Kvicera, M. Grabner, BER and Availability Measured on FSO, Radioengineering, vol. 16, no. 3, pp.7-12, 2007.

Impact of Plane Wave Excitation Parameters on Shielding Properties of Enclosure with Multiple Apertures

Tatjana Cvetković[1],Vesna Milutinović[1], Nebojša Dončov[2], Bratislav Milovanović[2]

Abstract- **In this paper the influence of parameters of plane wave excitation on shielding efficiency of enclosure with multiple apertures is analyzed. Two groups of two rectangular apertures placed on the adjacent enclosure walls, are considered. For analysis purposesTLM method implemented by means of commercial software is used. The numericalresults, proved to be valuable for the illustration of the effects of excitation parameters on the enclosureshielding performances.**

Keywords **–Electromagnetic shielding, multiple apertures, plane wave**

I. INTRODUCTION

The analysis of problems related to electromagnetic compatibility (EMC) is performed at the system design stage with the use of various numerical simulation techniques. The numerical method based onnetwork of transmission lines(Transmission Line Matrix – TLM) [1,2] and the finite difference method (Finite-Difference Time Domain – FDTD) [3] are widely applied to EMC analysis due to their characteristics. For analysis purpose, by using some of the previously mentioned methods, a detailed problem description in regards to the geometry and electromagnetic (EM) characteristics in the spatial domain is also required. In the time domain, their response to real excitation is simulated, while in the frequency domain the required parameters are calculated. With the use of numerical simulations we can determine the origin, nature and level of electromagnetic interference (EMI) that can affect the electronic system functioning. On the basis of the numerical analysis results, various procedures of electromagnetic protection can then be applied for reducing the coupling paths.

Shielded enclosures, surrounding the electronic system completely or partially, are commonly used as a solution to the problem of electronic system protection, i.e. reduction of the amount of EM radiation that reaches the system from the environment or is emitted into the environment by the system[4-5]. Materials with various EM characteristics and of various thicknesses are used for enclosures construction. In the practical application, enclosures usually have apertures on their walls in order to allow system access and control (connectors, cables, CD-DVD ROMs, etc), cooling and ventilation, or are due to imperfect technological realization. Therefore, electronic system performances in relation to EMC,besides the character of the excitation EM radiation source, the configuration of wire and dielectric structures within the system, also depend on the existence and nature of interconnection paths established through apertures enabling coupling between the EM source energy and sensitive electronic systems. The characterization of EM radiation penetrating into or out of the enclosure through apertures is shown in [6], for the purpose of estimating the impact of aperture existence on the enclosure shielding efficiency. Coupling through an aperture often has a dominant impact on the system operation in relation to EMC.

The level of EM radiation that penetrates into or out of the enclosure through the apertures, apart from the number of apertures, their geometry and mutual distance, can significantly depend on the excitation radiation parameters. Therefore, this paper discusses the effect of changing the polarization angle, azimuth and elevation of the electric field vector on the electric SE in the case of plane EM wave excitation. Multiple apertures, represented by two groups of two rectangular apertures, are assumed on the adjacent enclosure walls for system access and control purposes. Numerical TLM modelling method, implemented through a commercial software package, is used in order to estimate the efficiency of rectangular enclosure with multiple apertures over a frequency range up to 2 GHz. On the basis of the obtained numerical results, we have made appropriate conclusions, in relation to the effect of the analyzed excitation parameters on the enclosure shielding performances.

II. SHIELDING EFFECTIVENESS - SE

The parameter often used to estimate the enclosure performances is the shielding effectiveness (*SE*). The shielding effectiveness can be calculated with the use of a numerical simulation technique, in an analytical way and by measurements. It is expressed in dB and defined as the ratio between the incident field level in the appropriate point in the system without enclosure (E_i) and with enclosure (E_t) [1]:

$$SE = 20\log\left|E_i / E_t\right| \quad (1)$$

where E_i is the incident field level and $E_t$is the level of the field transmitted through the aperture.

*SE*is defined separately for electric (so-called electric *SE*) and magnetic field (so-called magnetic*SE*).Electric and magnetic *SE'*sdonot always have the same value. The minimum required *SE*value is 20dB. For most EMC problems, satisfactory effectiveness is in the range between 50 dB and 60 dB, but due to the increased number and complexity of electronic system it is desirable to achieve the highest possible effectiveness (about 100 dB).

[1]Tatjana Cvetkovic, Vesna Milutinovic - Republic Agency for Electronic Communications, Visnjiceva 8, 11000 Belgrade, Serbia,E-mail: tatjana.cvetkovic@ratel.rs,vesna.milutinovic@ratel.rs

[2]Nebojsa Doncov, Bratislav Milovanovic - The Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia,E-mail: nebojsa.doncov@elfak.ni.ac.rs,bata@elfak.ni.ac.rs

For the case of enclosure with perfectly conducting walls and with one rectangular aperture (i.e. slot because aperture length *l* is significantly larger than width *w*) placed at the center of the front wall (Fig.1), there is an analytical formulation of *SE*, based on the equivalent circuit model shown in Fig.2 [1,7].The model is derived for an excitation in the form of plane wave that propagates in a direction perpendicular to the front wall and whose electric field is polarized vertically in relation to the longer side of the aperture. In this case aperture has a considerable effect on the current flow induced on front wall leading to larger penetration of the incident field into enclosure.

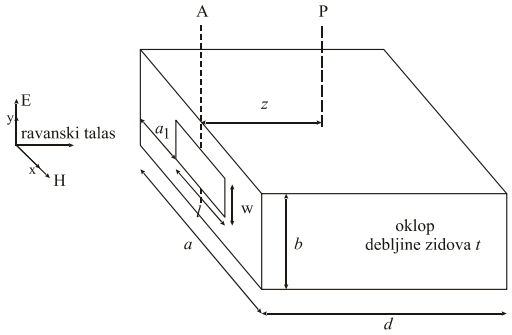


Fig.1. An enclosure with the rectangular cross section and single rectangular aperture

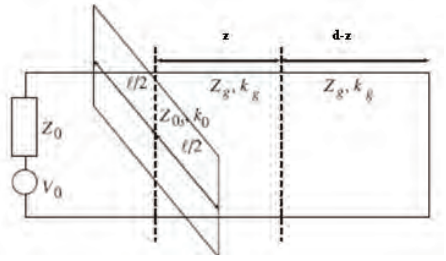


Fig.2. Equivalent circuit model of enclosure with a rectangular aperture, excited by a plane wave

The elements of the equivalent circuit model from Fig.2 represent the source of EM radiation, the aperture and enclosure itself. The incident plane wave is described with voltage generator, where V_0 represents the incident field amplitude and $Z_0$is the free space impedance ($Z_0 \approx 377\ \Omega$). The rectangular aperture is represented by a coplanar stripline that is short-circuited on both ends, where the total width of the coplanar stripline equals the height of the enclosure *b*, while the distance between strips equals the width of the aperture. The enclosure excited by the field generated at the aperture and short-circuited on the other end is describedas a short-circuited rectangular waveguide in which the TE_{10} mode propagates along the z-axis.

With the application of the Thevenin's Theorem, the calculation of the electric field within the enclosure at the distance *z*from the wall with rectangular aperture is reduced to calculation of voltage *V*(*z*) in the equivalent circuit. Since in the absence of enclosure the electric field at the same point is equal to a half of the voltage V_0, electric *SE* is:

$$SE = 20\log\left(\frac{V_0}{2V(z)}\right) \qquad (2)$$

The described model for calculation of the electric *SE* takes into consideration the existence of only one aperture situated in the center of a wall. If there are*n*apertures on the same wall placed at a distance where their mutual coupling can be neglected, equivalent circuit model should be modified to include *n* coplanar striplines, short-circuited on both ends and connected in series.

However, in reality enclosures can have multiple apertures on several wallsand incident plane wave polarization and propagation directioncan be different from the simplified case shown in Fig.1. In that case more complex techniques for the electric *SE*calculation should be used.It is necessary to perform decomposition of the electric field vector into *x*, *y* andz components.The calculation of voltage *V*(*z*) at the point at the distance *z* from the enclosure wall with apertures can be obtained through superposition of the voltage components as described in [8]. Such process can be very complicated, time-consuming and restricted by aperture dimensions and their mutual arrangement that can be taken into account. Therefore, in order to conduct an efficient analysis of plane wave propagation and polarization direction influence on the electric *SE*of the enclosure with multiple apertures represented by two groups of two rectangular apertures on the adjacent walls, numerical TLM method, implemented through a commercial software package, is applied in this paper and numerical simulation results are presented in the following section.

III. Numerical Analysis

For the calculation of the electric *SE*with the use of numerical TLM simulation, in this paper we used a metal enclosure with the rectangular cross-section whose dimensions are 300x120x300mm. An enclosure has two groups of two rectangular apertures the same size (30x20 mm) on adjacent walls as shown in Fig.3. The thickness of perfectly conductive walls of enclosure with rectangular apertures was t = 3 mm. The distance between the apertures was selected for the purpose of reducing their mutualcoupling effect and impact on *SE*. The frequency range of interest in this paper is from 0 to 2 GHz. We have already analyzed the effect of incident plane wave parameters on *SE*of an enclosure with two rectangular apertures of the same dimensions but situated only on one wall of the enclosure [9]. Fig. 3 shows the proposed enclosure model with plane wave parameters that we will change for the purpose of analyzing their impact on *SE*.

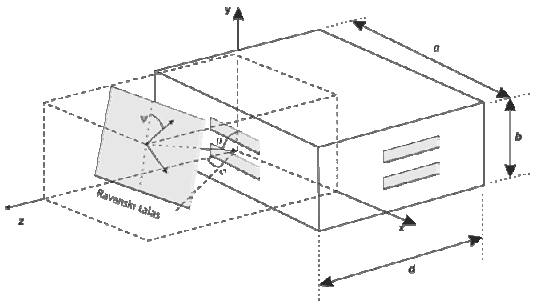
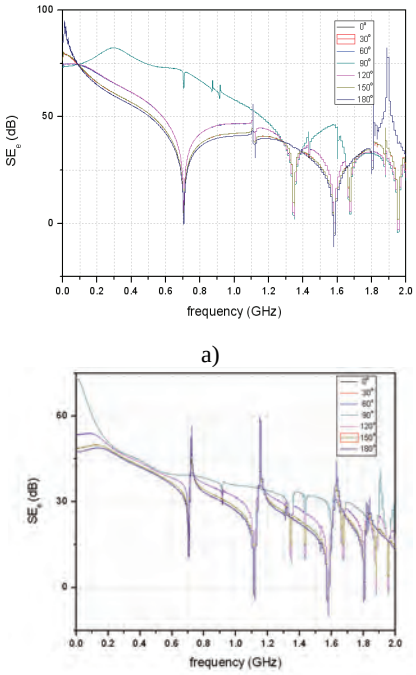


Fig.3. An enclosure with two groups of two rectangular apertures on adjacent walls and with plane wave parameters (angles ψ, θ i φ)

The polarization angle $\psi$of the plane wave electric field vector propagating in the direction that is vertical to one wall of the enclosure with apertures (θ=90° and φ=0°) is changed within the range from 0° to 180°, with steps of 30°.The change of SE_e is given in Fig.4, at the point in the center of the enclosure (Fig.4a), and at the point closer to the front wall with apertures (Fig.4b).



a)

b)

Fig.4. $SE_e$of the proposed enclosure model depending on the polarization angle of the electric field vector, observed in: a) the center of the enclosure, and b) a point closer to the front wall apertures

It can be noticed that a change of the polarization angle of the electric field vector from 0° to 90°, results in increase of SE_e.Highest value of SE_e, for the proposed enclosure model,is reached for the angle ψ=90°, (the electric field has the x component only), which is especially visible at the point in the center of the enclosure. Further change of the polarization angle of the electric field vector up to 180° results in reduction of the SE_e. We can also observe overlapping of curves for 0° and 180°, 30° and 150°, and 60° and 120°. The effect of polarization change is higher at the point closer to the apertures. Electric effectiveness at this point is 20dB lower than at the point in the center of the enclosure.

Fig.5 shows the change of SE_ewithin the observed frequency range for the polarization angle of the electric field vector ψ=90° for the proposed enclosure model at three points within the enclosure (point coordinates are given in the graph). With increasing the distance from the apertures the electric effectiveness increases. The SE_e will be highest at the point that is farthest from theapertures (point 150,60,30), while at points in the vicinity of the apertures it is considerably lower.

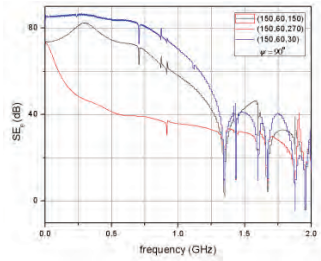


Fig.5 The change of SE_e for ψ=90° at various points within the enclosure

Fig.6 shows the change of SE_eat the first and second resonant frequency as function of the polarization angle at center point of enclosure, whose coordinates are given in the graph.

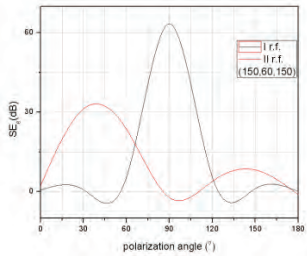
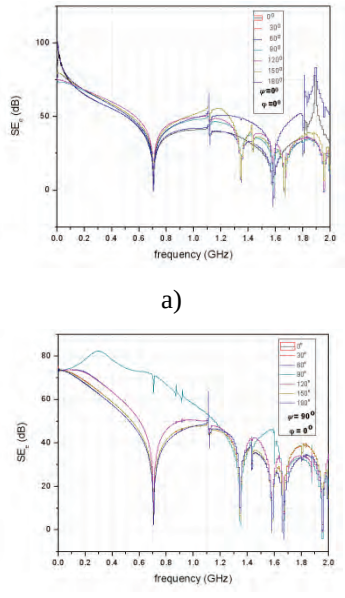


Fig.6. The change of SE_e with polarization angle at center point of enclosure at the first and second resonant frequency

The most critical case, from the aspect of electric effectiveness, is the point situated in the vicinity of apertures in *y-z* plane.The highest effectiveness is recorded at the point that is furthest from the apertures.

For constant polarization of the electric field vector (ψ= 0° and ψ=90°)and constant azimuth (φ=0°) we changed the elevation angle (the excitation position in the elevation plane around the enclosure wall with apertures) from 0° to 180°, with step of 30°. Fig. 7 illustrates the effect of the electric SE_echange at the point in the center of the enclosure for ψ= 0°, φ=0°and ψ=90°, φ=0°.



a)

b)

Fig.7. *SE* due to the change of the elevation angle of the electric field vector for a) ψ= 0°,φ=0° and b) ψ=90°, φ=0°

It can be concluded that a change of the elevation angle of the electric field vector for ψ=0° and φ=0° to the second resonant frequency has no considerable effect on SE_e (curves partially overlap). For ψ=90° and φ=0° the highest electric *SE* is for θ=90°.

Fig.8 shows the change of SE_ein the function of the elevation angle with ψ =90°,φ=0°and ψ =0°, φ=0°, at the first resonant frequency.

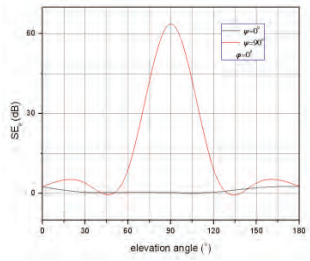


Fig.8. SE_e in the function of the elevation angle for ψ=90°, φ=0° and ψ =0°, φ=0°on the first resonant frequency

In the following case, as shown in Fig.9, we analyzed the effect of a change of azimuth of the electric field vector, from -90° to 90°, with steps of 30° on SE_e with constant polarization angles ψ =90° and the elevation θ=90°.

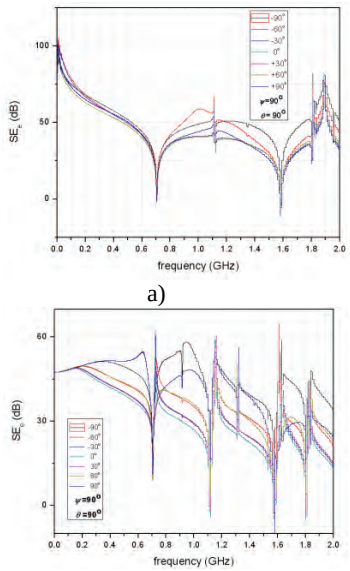


Fig.9. SE_e of the proposed enclosure model depending on angle φ
a) in the center of the enclosure, b) at a point closer to the front side with apertures

With a change of azimuth, for the proposed enclosure modelthe curves for electric effectiveness mostly overlap up to the first resonant frequency. Between the first and the second resonant frequency, the highest SE_e is for φ = -60°, and above the second resonant frequency for φ = -90°.

IV. Conclusion

Shielded enclosures, as one of the most frequent types of electromagnetic protection, have apertures on their walls. The purpose of the apertures is various, for access to the system for its control and power supply, the need for cooling and removal of surplus heat from the system, etc. Through the apertures the EM radiation penetrates into the enclosure, i.e. the space outside it, which degrades its protective function. Assuming that conductance of the enclosure walls is very high, the EM coupling through the apertures has dominant impact on functioning of the electronic systems regarding EMC. For evaluation of the shielding effectiveness, in this work we applied the numerical TLM method, implemented through an appropriate software package. The proposed enclosure model is rectangular and has two rectangular apertures at fixed mutual distance on two adjacent sides of the enclosure. We analysed the effect of excitation parameters of a plane EM wave, polarization, elevation and azimuth of the electric field vector on the electric *SE* of the enclosure. Based on the results obtained through numerical simulation, we can select an optimal position of the system in the enclosure, so that the influence of EM interference on the electronic system is minimal. For most practical EMC problems, excitation in the form of a plane wave represents approximation of real excitation. Under real conditions, when EM radiation reaches the enclosure at an approximate angle, the situation is more complex. In the upcoming period, the authors intend to carry out detailed investigation of the excitation parameter impact on the system shielding function.

References

[1] C. Christopoulos: The Transmission-Line Modelling (TLM) Method, IEEE Press in association with Oxford University Press, Piscataway, NJ,1995.

[2] Li, M., Nuebel, J., Drewniak, J.L., DuBroff, R.E., Hubing, T.H., Van Doren, T.P., "EMI from Cavity Modes of Shielding Enclosures – FDTD Modelling and Measurements", *IEEE Trans. Electromagn. Compat*, Vol.42, No.1, pp. 29-38, 2000.

[3] V. Trenkic, R. Scaramuzza: "Modelling of arbitrary slot structures using transmission line matrix (TLM) method", Proceedings of International Symposium on Electromagnetic Compatibility, Zurich, Switzerland, pp.393-396,2001.

[4] C. Christopoulos: Principles and Techniques of Electromagnetic Compatibility, CRC Press, 2000.

[5] Kaiser, K.L., *Electromagnetic Compatibility* Handbook, CRC Press, 2005.

[6] Mendez, H.A., "Shielding Theory of Enclosures with Apertures", *IEEE Trans. Electromagn. Compat*, Vol. 20, No. 2, pp. 296–305, 1978.

[7] M.P. Robinson, T.M. Benson, C. Christopoulos, J.F. Dawson, M.D. Ganley, A.C. Marvin, S.J. Porter, D.W.P. Thomas: „Analytical Formulation for the Shielding Efectiveness of Enclosure with Apertures", IEEE, Transactions on Electromagnetic Compatibility, Vol.40, No.3, 1998, pp.240-248.

[8] J. Shim, D.G. Kam, J.H. Kwon, J. Kim:"Circuital Modeling and Measurement of Shielding Effectiveness Against Oblique Incident Plane Wave on Apertures in Multiple Sides of Rectangular Enclosure" IEEE Transactions on electromagnetic compatibility, Vol.52, No.3, August 2010.

[9] T. Cvetković, V. Milutinović, N. Dončov, B. Milovanović: „Analysis of plane wave polarization and propagation direction influence on the efficiency of rectangular enclosure with apertures", INFOTEH 2011, Jahorina.

Analysis of the Shielding Effectiveness of Enclosure with Multiple Circular Apertures on Adjacent Walls

Vesna Milutinovic[1], Tatjana Cvetkovic[1], Nebojsa Doncov[2] and Bratislav Milovanovic[2]

Abstract **–In this paper an electromagnetic coupling through apertures and its effect on shielding performance of enclosuresareanalyzed. Multiple apertures, represented by two groups of two circular aperturesplaced on the adjacent rectangular enclosure walls,are considered. The electric effectiveness has been calculatedin the frequency range of EMC interest by applying a numerical TLM method implemented through a commercial software package.A plane wave which propagates in a direction perpendicular to oneof the wallscontaining two aperturesis used as an excitation while the distance between the apertures in each group varies in horizontal or vertical direction.**

*Keywords***–Enclosure, multiple aperture, coupling, shielding effectiveness.**

I. INTRODUCTION

In order to reducean emissions from electronic equipment and/or achievean equipment immunity, an electromagnetic (EM) enclosure is being used. Thisenclosure, made of conductive material of suitable thickness and with different EM properties, affects on the amount of EM radiation that reaches the electric circuit from the external environment, but also determines how much energy from the circuit is radiated in the environment. Many devices have enclosures with more apertures for outgoing or incoming cable penetration, control panels, heat dissipation, airing or other purpose. The characteristics of electronic circuit blocks and devices housed within the enclosure can be affected by EM threates through various coupling ports commonly used in electronic systems. Performance of enclosure are determined by the shielding effectiveness (SE), which is defined as the ratio of field strength in the presence and absence of enclosure. The presence of apertures degrades the shielding effectiveness, so it is very important that the coupling through the apertures is taken into account for accurate calculations of SE. There are several methods already developed for the calculation of SE of metal enclosures with apertures on their walls, such as analytical formulations [1], which relies on Fourier transformation and the model analogy. A more complex approach to this problem requires solving the sophisticated problem of scattering using the Mendez`s method [2]. A slots and apertures have a tendency to become a coupling route of electromagnetic interference (EMI) from the external environment inside the enclosure, which degrades the shielding effectiveness. Due to EMI coupling an electronic equipment may degrade the operation of other equipment in the same frequency range. It is very important to understand how factors such as aperture size, shape and mutual spacing, thickness of the enclosure and the position of the apertures affect on the SE and how essential they are in reduction of EMI and sensitivity.

Differential numerical techniques in the time domain, such as the Finite-Difference Time Domain (FDTD) [3] and the method of modeling using trasnmission lines network (Transmission Line Matrix - TLM) [4], owing to its characteristics, have found their application in solving many EMC (*electromagnetic compatibility)* problems in a wide frequency range. The application of these numerical methods for the analysis of practical EMC problems requires a detailed description of geometrical and EM characteristics of the problem and calculation of appropriate parameters in the frequency range of interest. For the vast majority of EMC problems, as the basic requirement it is imposed an adequate numerical modeling of interactions of excited EM fields with geometry of small but in a electrically sense important structures (thin wire structures, complex wired circuits, slots, apertures, etc), embedded within the physical large and shielded systems.

Apertures presence influence to the shielding performance of enclosure has been extensively studied [5]. Simple aperture pattern and orientation has been studied in [6] and with its formulation, the electric and magnetic shielding effectiveness of a rectangular enclosure with one or more apertures in a wall can quickly be predicted. The radiation leakage through the array of apertures is also significantand presented in [7], by the method of moments (MoM). An enclosure with different apperture orientation and dimension is analysed by using FDTD numerical simulations in [8]. A quantitative relationship between EMI and number of apertures and their dimensions is given in the paper [9]. EMI shows a strong depedence on various factors like aperture patterns, their dimensions and number, as well as their orientation according to their orientation on the wall of the enclosure or a plane wave[10-12].Shielding effectiveness was calculated as functions of frequency, aperture dimensions and position not only of an aperturewithin the enclosure, but an incident plane wave also in [13-14].

In this paper, the TLM method implemented through a commercial software package was used to study an effect of

[1]VesnaMilutinovic is with the Republic Agency for Electronic Communications, Visnjiceva8, 11000 Belgrade, Serbia, E-mail: vesna.milutinovic@ratel.rs.

[1]Tatjana Cvetkovic is with the Republic Agency for Electronic Communications, Visnjiceva 8, 11000 Belgrade, Serbia, E-mail: tatjana.cvetkovic@ratel.rs.

[2]Nebojsa Doncov is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nebojsa.doncov@elfak.ni.ac.rs.

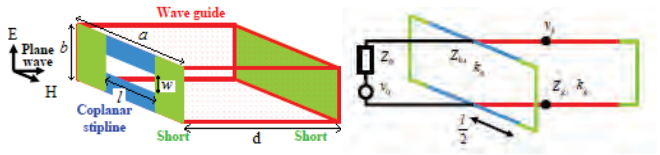
[2]Bratislav Milovanovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: bata@elfak.ni.ac.rs.

aperture spacing in vertical and horizontal direction on the electric SE ofrectangular enclosure over a frequency range up to 2 GHz. It is assumed that two groups of two circular apertures, placed on the adjacent enclosure walls are present due to outgoing or incoming cable penetration. On the basis of obtained numerical results, relevant conclusions are carried out from the point of effects of multiple aperture coupling to the shielding effectiveness of enclosure.

II. Modeling of Apertures

A. Circuit Model for Shielding Effectiveness Calculation

There are many techniques that deal with multiplenumbers of apertures in an enclosure. Approximate analytical methodsare accurate but applicable only to simple geometries. A simpleanalytical method has been introduced byRobinson et al. based on aTLM [7]. In this method, the rectangular enclosure(Fig. 1a) is modeled by a short-circuited rectangular waveguide, the apertureis represented by a coplanar strip transmission lineand the plane wave incident as excitation isThevenin equivalentwith parameters which represent the strength of the incident field, and Z_0 impedance of free space (~ 377 Ω) like in Fig.1b.The exact value of the voltage v_0 for the calculation of SE is not relevant because it is expressed as a ratio of two fields.By neglectingthe mutual coupling between multipleapertures, the perforated wallimpedance is the sum of the individual elements.



a) b)

Fig. 1. Rectangular enclosure with an aperture and its equivalent circuit

Characteristic impedance ofshort circuited coplanar strip transmission line section, representing an aperture, can be calculated as:

$$Z_{0cs} = 120\pi^2 \left[ln\left(2\frac{1+\sqrt[4]{1-(2w_e/b)^2}}{1-\sqrt[4]{1-(2w_e/b)^2}} \right) \right]^{-1} \quad (1)$$

where

$$w_e = w - \frac{5t}{4\pi}\left(1 + ln\left(\frac{4\pi w}{t}\right)\right) \quad (2)$$

Enclosure is presented as a shorted rectangular waveguide in which the propagation is along the z-axis. Assuming TE_{10} mode, the characteristic impedance and phase constant in the waveguide can be expressed, respectively, as:

$$Z_g = Z_0 / \sqrt{1-(\lambda/2a)^2} \quad (3)$$

$$k_g = k_0 \sqrt{1-(\lambda/2a)^2} \quad (4)$$

An equivalent circuit of complete problem is shown in Figure 2b, from which it can be seen that when a section of transmission line is short circuited at the load end, Z_{ap_half} and Z_{ap} may be written, respectively, as:

$$Z_{ap\_half} = jZ_{0s}\, tan\frac{k_0 l}{2} \quad (5)$$

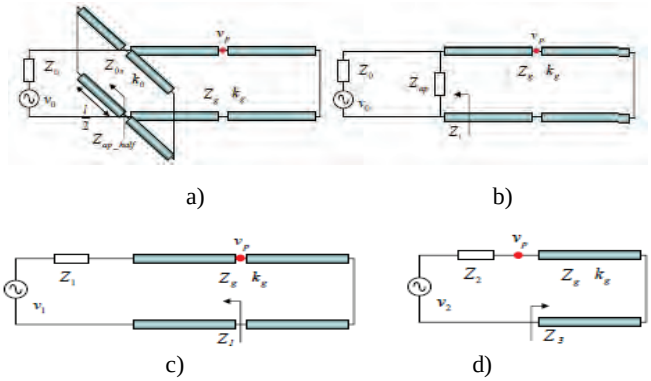
$$Z_{ap} = \frac{l}{a}(Z_{ap\_half} \| Z_{ap\_half}) = \frac{1}{2}\frac{l}{a} jZ_{0s} \tan\frac{k_0 l}{2} \quad (6)$$

Calculation of the electric field inside the enclosure at a distance p from the wall with anaperture is reduced to the calculation of voltage v_p in the equivalent circuit. Equivalent circuit can be further transformed into a Thevenin's equivalent circuit in Figure 2c, where $v_1$and $Z_1$are its components which represent incident field and the impedance of the aperture in Figure 2b. Finally, reduction of Thevenin's equivalent circuit, looking left and right from the coordinate z inside the enclosure gives the circuit in Figure 2d, whose elements are:

$$V_2 = \frac{V_1}{cos(Z_g p) + jZ_1 sin(k_g p)/Z_g}$$

$$Z_2 = \frac{Z_1 + jZ_g tan(k_g p)}{1 + jZ_s tan(k_g p)/Z_g}$$

$$Z_3 = jZ_g \tan[k_g(d-p)] \quad (7)$$



a) b)

c) d)

Fig. 2. Equivalent circuit of the complete problem

Voltage on the distance $p$can be then calculated as:

$$Vp = V_2 Z_3 /(Z_2 + Z_3) \quad (8)$$

Since in the absence of an enclosure electric field at the same point is equalto half of the voltage V_0, the final expression for SE is:

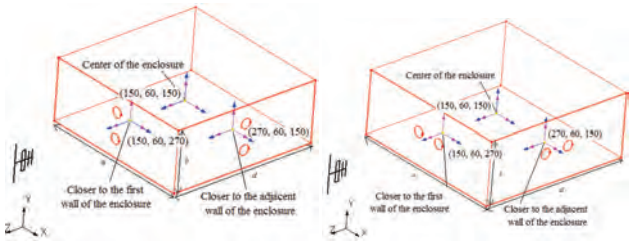
$$SE = 20\log\left(\frac{V_0}{2V(p)}\right) \quad (9)$$

It can be shown that the previously given equations can be also applied for the calculation SE of enclosure with a circular aperture of radius r placed in the center of the front wall, with the approximation $l = w = \sqrt{\pi} r$.

Limitations of previous circuit model are such that only aperture on single side for normal incident wave and single mode is supported.In cases where the aperture is not in the center of the wall, when there are two or more apertures in the same or different walls of the enclosure or when an incident wave encounters to the plane of the wall with apertures with different propagation and polarization angles,more complex techniques should be applied. Only then it is possible to get complete analysis of the influence of these factors on the SE of enclosure. In this paper,numerical TLM method, implemented through a commercial software package,is applied for the analysis of the impact of changing the horizontal or vertical spacing betweentwo circular apertures in each group of apertures positioned on the two adjacent walls, on the electrical efficiency of the enclosure, and results are presented in the next section.

B. Numerical analysis

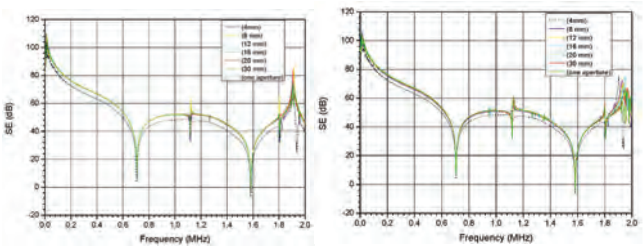
The purpose of this study is toexamine the role of various apertures on coupling behavior and assist the design of electronic systems.In this paper, the electrical SE of an enclosure of rectangular cross-section with following dimensions: a = 300 mm, b = 120 mm and d = 300 mm and with two round apertures on the front and adjacent walls (Fig.3) is calculated by the numerical simulation. Apertures are placed one above the other symmetrically according to the center of the wall (change of spacing between the apertures was performed in a vertical direction), or one nearanother (changing the spacing between the apertures was performed in a horizontal direction). The thickness of perfectly conducting metal walls with circular aperturesis t = 2 mm. An excitationin the form of a plane wave whose direction of propagation was normal to the plane of the front wall with apertures was used. Circular apertures have a diameter of 2r = 20 mm.



a) b)

Fig. 3. An enclosure with two round apertures

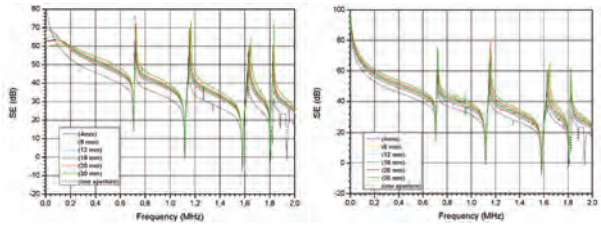
The effect of increasing the spacing between the apertures on the electrical SE is illustrated by the example of the circular apertures in the center point of the enclosure and in the points near apertures of adjacent walls for the vertical direction (Fig. 4a), and for the horizontal direction (Fig. 4b). The spacing between the circular apertures was beingincreased in steps of 4 mmup to 20 mm, but also the cases when the distance is greater than aperture diameter (30 mm) and when there is one larger circular aperture with the same size as the considered two small ones were considered. From Figures 4a and 4b it can be seen that, for various values of the spacing between the apertures, the electrical curves for SE, calculated at the center of the enclosure, partially overlap, where minimum SE is for one large aperture of the same area as two smaller apertures.Shape of the curve with the change spacing remains the same, indicating that the spacing considered only affects the field attenuation during the propagation through the inside of the aperture. A minimum value of electrical efficiency is obtained in the case of one larger aperture, which indicates that SE of enclosure is higher in case of having more small apertures than the aperture same size as one larger aperture.



a) b)

Fig. 4. SE of enclosureat the center point, with two apertures placed: a)one above another, b) one near another

In the point near apertures on the front wall (Fig. 5), aminimum value of SEis obtained for one larger aperture, and the maximum for the aperture spacing of 30 mm. Besides, SE curves progressively are decreasing with reducing vertical or horizontal aperture spacingwhere the higher values of SE are obtained for apertures placed one near another.



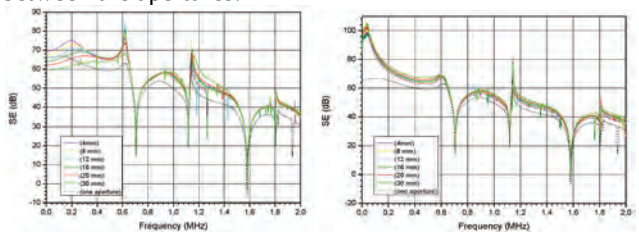
a) b)

Fig. 5. SE of enclosure at the point nearer to frontal wall, with two apertures placed: a) one above another, b) one near another

In the point near apertures on the adjacent wall (Fig. 6), aminimum value of SEis obtained for one larger aperture, and from 0 MHz to 0,6 GHz SE curves progressively are increasing with reducing vertical aperture spacingwherefrom 0,6 GHz to 2 GHzthe electrical curves for SE, partially overlap. At the other hand,SE curves for horizontal aperture spacingin the entire frequency range observed partially overlap. A minimum value of electrical efficiency is obtained in the case of one larger aperture.

Although the aperture spacing has an insignificant effect on the position of resonance points, it has a significant effect on SE of a shielding enclosure at resonance points, especially

at higher frequencies. From Fig. 7 it can be seen that the SE does not change the valuewith changing the vertical aperture spacing for the first and third resonance, but gets negative around the third resonance, so it is advisable to avoid it in the design process. For the other resonance frequencies of the shielding enclosure, the SE is changing with the distance between the apertures.



a) b)

Fig. 6. SE of enclosure at the point nearer second wall, with twoapertures placed: a) one above another, b) one near another

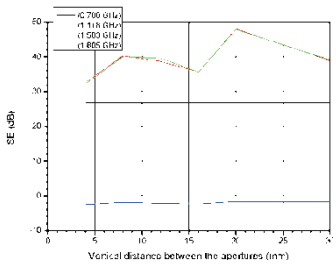


Fig. 7. SE for different resonant frequencies depending on the aperture spacing for the point in the center of enclosure

III. Conclusion

The computationin the early stage of design allows the identification of potential problems such as interference from the environment, the position of the mutual coupling aperture in the enclosure, the sensitivity of the equipment inside the enclosure etc. In this paper, the impact of the mutual spacing of multiple circular apertures, whose diameters are significantly smaller than the wavelength, placed on adjacent walls is considered. It can be concluded that changing the distance between two circular aperturesfrom4 mm to 30 mm in the vertical direction does not affect greatly the SE, provided that the above mentioned effect was more pronounced at the point near the apertures especially at lower frequencies to 200 MHz, and that the lower efficiency is with a larger aperture of equal surfaceas small ones by 5 dB. When changing the distance in the horizontal direction the greatest impact in changing SE can be seen at high frequency and above 1.8 GHz. The calculations show that the enclosure resonates at approximately four frequencies(Fig. 7). Below the resonant frequency, SE decreases with frequencies and increases with distance from the aperture.In all of the above mentioned analysis it is noticed that the smallest SE is for the case of one larger aperture.The studies of resonance behavior are important in the electromagnetic compatibility field, which can help designers to avoid resonance phenomenon in design process.

References

[1] H. H. Park and H. J. Eom, “Electromagnetic penetration into a rectangular cavity with multiple rectangular apertures in a conducting plane” IEEETrans. Electromagn. Compat., vol. 42, no. 3, pp. 303–307, Aug. 2000

[2] H. A. Mendez, “Shielding theory of enclosures with apertures” IEEE Trans. Electromagn. Compat., vol. EMC-20, no. 2, pp. 296–305, May1978.

[3] K.S. Kunz, R. J. Luebbers: The Finite Difference Time Domain Method for Electromagnetics, CRC Press, Boca Raton, FL, 1993.

[4] C. Christopoulos: The Transmission-Line Modelling (TLM) Method, IEEE Press in association with Oxford University Press, Piscataway, NJ, 1995.

[5] M. Lin, J. L. Drewniak, S. Radu, J. Nuebel, T. H. Hubing, R. E. DuBroff, and T. P. Van Doren, “An EMI estimate for shielding-enclosure evaluation”IEEE Trans. Electromagn. Compat., vol. 43, pp. 295–304, Aug.

[6] Robinson, M. P., T. M. Benson, C. Christopoulos, J. F. Dawson, M. D. Ganley, A. C. Marvin, S. J. Porter, and D. W. P. Thomas, “Analytical formulation for the shielding effectiveness of enclosures with apertures” IEEE Trans. Electromagn. Compat., Vol. 40, No. 3, pp. 240–248, Aug. 1998.

[7] Shabista Ali, Daniel S. Weile, Thomas Clupper, “Effect of Near Field Radiators on theRadiation Leakage Through Perforated Shields” IEEE Trans. Electromagn. Compat., Vol. 47, No. 2, pp. 367–373, May 2005.

[8] Min Li Joe Nuebel, James L. Drewniak, Richard E. DuBroff, Todd H. Hubing, Thomas P. Van Doren, “EMI from Cavity Modes of Shielding Enclosures – FDTD Modeling and Measurements” IEEE Trans. Electromagn. Compat., Vol. 42, No. 1, pp. 29–38, Feb. 2000.

[9] Min Li Joe Nuebel, James L. Drewniak, Richard E. DuBroff, Todd H. Hubing, Thomas P. Van Doren, “EMI from Airflow Aperture Arrays in Shielding Enclosures – Experiments, FDTD, and MoM Modeling Modeling” IEEE Trans. Electromagn. Compat., Vol. 42, No. 3, pp. 265–275, Aug. 2000.

[10] Jongjoo Shim, Dong Gun Kam, Jong Hwa Kwon, Joungho Kim,”Circuital Modeling and Measurement of Shielding Effectiveness against Oblique Incident Plane Wave on Apertures in Multiple Sides of Rectangular Enclosure” IEEE Trans. Electromagn. Compat., Vol. 52, No. 3, pp. 566–577, Aug. 2010.

[11] Bao-Lin Nie, Ping-An Du, Ya-Ting Yu, and Zheng Shi “Study of the Shielding Properties of Enclosures With Apertures at Higher Frequencies Using the Transmission-Line Modeling Method” IEEE Trans. Electromagn. Compat., Vol. 53, No. 1, pp. 73–81, Feb. 2011.

[12] Milovanović, B., Dončov, N., Milutinović V., Cvetković, T., “Numerical characterization of EM coupling through the apertures in the shielding enclosure from the viewpoint of electromagnetic compatibility”, Telecommunications - Scientific journal published by the Republic Agency for Telecommunications – RATEL, No.6, pp.73-82, 2010.

[13] Milutinović V., Cvetković, T., Dončov, N., Milovanović, B., “Shielding effectiveness of a rectangular enclosure with two apertures depending on distance between apertures”, (Yu Info 2011Proceedings).

[14] Cvetković, T., Milutinović V., Dončov, N., Milovanović, B., “Analysis of the influence of polarization and direction of propagation of a incident plane wave on the effectiveness of rectangular enclosures with apertures”, (Infoteh - Jahorina, March 2011).

Neural Network Based Software for Modeling Printed Pentagonal Dipole

Marija Milijić[1], Zoran Stanković[1], Ivan Milovanović[2], Aleksandar Nešić[3]

Abstract- **This paper presents software for modelling printed pentagonal dipole. The software is based on neural model that calculates dipole resonant frequency versus dipole dimension and substrate dielectric constant. The neural model training and test set consist of results get by WIPL-D software package. WIPL-D uses Method of Moments enabling high accuracy of results, but also causing a long simulation time. Unlike this software, artificial neural network can achieve both great simulation speed and the satisfactory accuracy.**

Keywords – **printed antenna, pentagonal dipole, neural network**

I. INTRODUCTION

Printed antennas, including microstrip and printed antennas, have several well known advantages over other antenna structures, including their low profile, light weight, low cost of production and compatibility with microwave monolithic integrated circuits (MMICs) and optoelectronic integrated circuits (OEICs) technologies [1]. Because of these merits, forms of the printed antenna have been utilized in many applications such as in mobile communication base stations, spaceborne satellite communication systems and even mobile communication handset terminals. Also, it was ease of fabrication and development.

For this reason, design techniques for printed antennas have attracted much attention from antennas researches [2]. The most known method used for modeling printed antennas is the electromagnetic simulation. Although it is very correct process, it has some disadvantages which can not satisfy requirements of communication systems designing under some circumstances. Its basic disadvantage is that electromagnetic simulation has high demands concerning the hardware resources necessary for its software implementation. The software implementation itself might be very complicated and faced with many difficulties. Also the time needed for numerical calculation by electromagnetic simulation could be unacceptably long. The method of moments (MoM) is a very popular algorithm of computer electromagnetic calculations. It is widely used for antenna simulations and electromagnetic wave scattering analysis due to its good accuracy. The MoM is based on transformation of field equations into a system of linear equations [3]. The mathematical basis for the MoM has been known for a long time however, the method was used for solving particular electromagnetic-field problems, such as the analysis of linear antennas in the early 1950s. The main disadvantages of the MoM is demand to solve a large number of time-consuming complex electromagnetic equations.

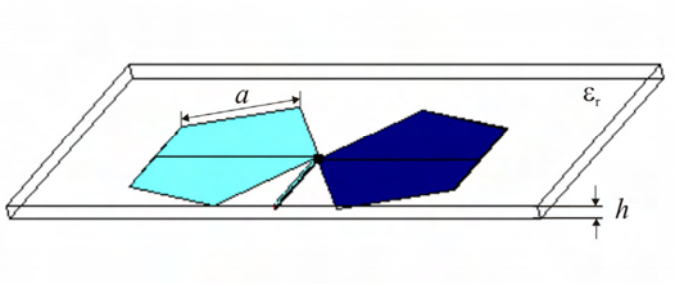


Fig.1. Printed pentagonal dipole

The previous researching, concerning the modeling of slotted patch antennas have showed that the neural network models can have satisfactory accuracy similar as MoM but also can have higher simulation speed then EM simulations [4-7]. In this researching, it is shown that neural network could be very successfully for slotted patch antenna modeling carried out in the field of signal and noise modeling of these devices. Recent research [8] point out that neural network can be good tool for modelling printed pentagonal dipole resonant frequency versus its dimension. This paper suggests neural model of pentagonal dipole that calculates resonant frequency versus dipole dimension and substrate dielectric constant. It is incorporate in the software enabling great simulation speed, same accuracy as MoM accuracy and user friendly work.

II. PRINTED PENTAGONAL DIPOLE

The printed pentagonal dipole is presented on Fig.1. It consists of two regular pentagons of dimension *a* that are situated on substrate with dielectric constant e_r and heigh *h*. One half of pentagon is in on side of dielectric and other half is on opposite side of dielectric. The dipole is fed by symmetrical microstrip lines of input impedance of 100 Ω. The corner reflector is consists of two metal plates which is situated at an angle of 45°. Proposed dipole is used as element of printed arrays enabling side (SLS) lobe suppression better than 34 dB in E-plane [9]. Such impressive SLSs are hardly achievable with conventional microstrip antenna arrays (with patches). In microstrip antenna arrays presented in literature, side lobe levels are suppressed 25 dB (related to main lobe) at best. For this reason, the tools for modelling printed pentagonal dipole are research in this paper.

[1]Marija Milijić and Zoran Stanković are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18 000 Niš, Serbia, e-mail:{marija.milijic, zoran.stankovic} @elfak.ni.ac.rs
[2]Ivan Milovanović Nešić is with University Singidunum, Serbia, e-mail: ivanmilovanovic@singidunum.ac.rs
[3]Aleksandar Nešić is with IMTEL Institute, Bulevar Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, e-mail: aca@insimtel.com

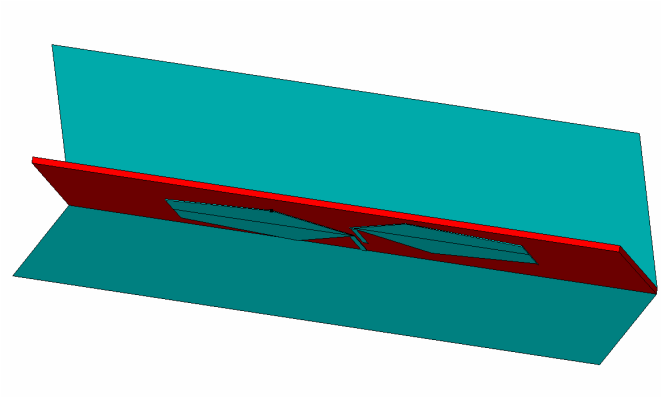


Fig. 2. Model of printed pentagonal dipole in WIPL-D software

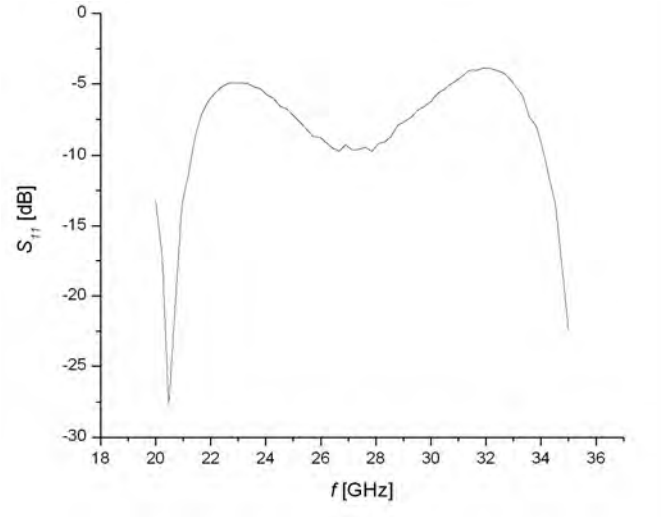


Fig. 3. Simulated S_{11} paremetrs for pentagonal dipole side a= 5mm obtained by software WIPL-D

III. Modeling Printed Pentagonal Dipole Using WIPL-D Software

WIPL-D is a commercial software for high frequency electromagnetic modeling and simulation. For over a decade, it has been a useful and reliable assistant to experts both in industry and academia. This software is specially tailored for research and design of antennas and microwave circuits as well as for analysis of scatterers and EMC problems [10]. It enables modeling of arbitrary metallic and dielectric 3D structures with wires, plates and parametric objects. Also, it uses higher order basis functions and the Method of Moments (MoM) to provide highly efficient analysis. For electrically large structures, special techniques such as multilevel fast multipole method and smart reduction of expansion order boost the performance even further.

The model of printed pentagonal dipole in WIPL-D software is presented in Fig. 2. WIPL-D software enables the simulation of Y-parameters, Z-parameters, S-parameters (Fig. 3.), radiation pattern, etc. The main disadvantage of WIPL-D software is long simulation time, especially when the process of modelling antennas must be done in some period of time.

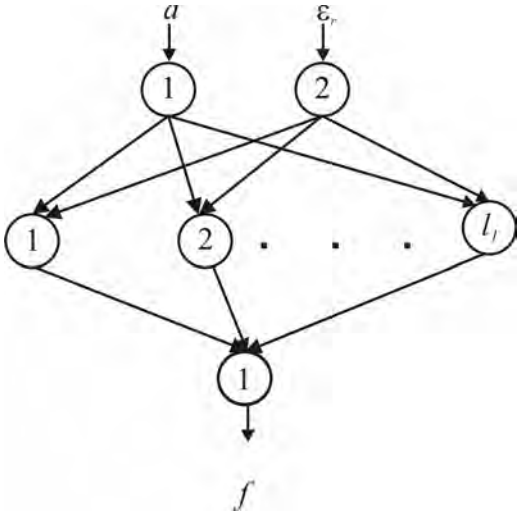


Fig. 4. Neural model of printed pentagonal dipole

IV. Modeling Printed Pentagonal Dipole Using Neural Networks

An Artificial Neural Network is very sophisticated modeling techniques capable of modeling extremely complex functions. Indeed, anywhere that there are problems of prediction, classification or control, neural networks can be introduced [11,12].

The neural network is inspired by the way biological nervous systems, such as the brain, process information. It is composed of a large number of highly interconnected processing elements - neurons working in unison to solve specific problems. Typically, a number of neurons are organized in layers. The input layer is not really neural at all: these neurons simply serve to introduce the values of the input variables. The latest layer of neural network is output layer and its neurons give output of neural network. Other neurons, which are not connected with input or output links, are hidden neurons. The hidden and output layer neurons are each connected to all of the neurons in the preceding layer. These neurons receive a number of inputs, either from original data or from the output of other neurons in the neural network. Each input comes via a connection that has a strength or weight. Each neuron also has a single threshold value. The weighted sum of the inputs is formed, and the threshold subtracted, to compose the activation of the neuron. The activation signal is passed through an activation function (transfer function) to produce the output of the neuron. In this way, signals flow from inputs, forwards through any hidden neurons, eventually reaching the output neurons and forming feedforward neural network.

Neural networks learn by example. The neural network user gathers representative data, and then invokes training algorithms to automatically learn the structure of the data. Neural networks are applicable in virtually every situation in which a relationship between the predictor variables-inputs and predicted variables - outputs exists, even when that relationship is very complex and not easy to articulate. The key feature of neural networks is that they learn the

input/output relationship through training. In supervised learning, the network user assembles a set of training data. The training data contains examples of inputs together with the corresponding outputs, and the network learns to infer the relationship between the two.

MLP neural network for modeling printed pentagonal dipole consists of input, one hidden and output layers (Fig.4.). The number of input layer neuron is equal to number of dipole parameters that determine modeling. In this application, there are two input parameters: dipole dimension a and substrate dielectric constant ε_r. The number of hidden layer neurons is variable during training process and output layer has one neuron that gives resonant frequency f_r. The MLP network models the function:

$$[f_r] = f(a, \varepsilon_r) \quad (1)$$

The activation functions of the hidden layers are sigmoid, while the neurons of the output layers have linear activation functions. The neural networks were trained using Levenberg-Marquardt method with 10^{-4} performance goal. The notation of MLP models is MLPn-l_1-l_2-…-l_{n-2} where n represents layer number and l_1-l_2-…-l_{n-2} are the numbers of neurons of its each hidden layer.

The values of resonant frequency fr necessary for the training and the test MLP neural networks, were obtained by WIPL-D software. This software uses method of moments to calculate S_{11} parameter for certain frequency f of pentagonal dipole with specific dimension. Training and test sets consist of only resonant frequency f_r defined by minimum value of S_{11} parameters for specific dipole.

Pentagonal dipole, modeled in this paper, has a substrate with height h=0.254 mm. The width of fed line w depends on substrate dielectric constant ε_r and its range is [0.23 mm, 0.198 mm][13]. The other dipole parameters are changeable. In the training set with 33 samples two input parameters have following range: 3 mm≤a≤7 mm and 2.1≤ε_r≤2.5.

V. Simulation Results

The test set contained 10 samples that have not been used in training process. Test results of successfully trained MLP networks are presented in the Table I together with the average test error (*ATE*), the worst case error (*WCE*) and the Pearson Product-Moment correlation coefficient (r^{PPM}). The minimum of average test error and the maximum value of r^{PPM} coefficient represent the basic criterion for selection the best MLP network. Selected neural model is MLP3-4.

At first, generalization level of MLP3-4 model should be checked. First, dependence of resonant frequency f_r on pentagonal dipole dimension a for different dielectric constant ε_r obtained by MLP3-4 compared with WIPL-D simulation is shown in Fig. 5. This figure shows the satisfying accuracy of neural model compared with MoM simulation results. The similar conclusion can be done in Fig. 6. that shows how resonant frequency f_r depends on dielectric constant ε_r for different values of pentagonal dipole dimension a.

TABLE I
TEST RESULTS

MLP model	*WCE* [%]	*ACE* [%]	r^{ppm}
MLP3-4	3.8106	1.3908	0.9992
MLP3-3	3.1403	1.6080	0.9990
MLP3-2	3.9077	1.4751	0.9990
MLP3-5	3.6858	1.5806	0.9988
MLP3-6	4.8766	1.6835	0.9983
MLP3-7	3.3281	1.6863	0.9983
MLP3-12	4.7610	1.9812	0.9977
MLP3-13	9.6557	4.6384	0.9891

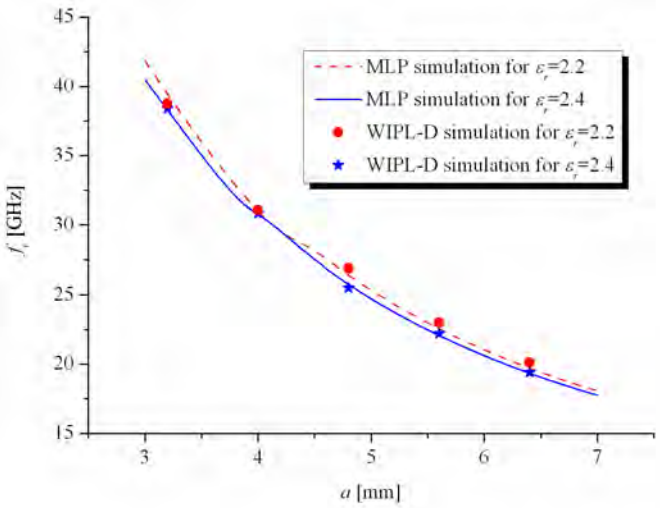


Fig. 5 Resonant frequency f_r versus pentagonal dipole dimension a for ε_r=2.2 and ε_r=2.4

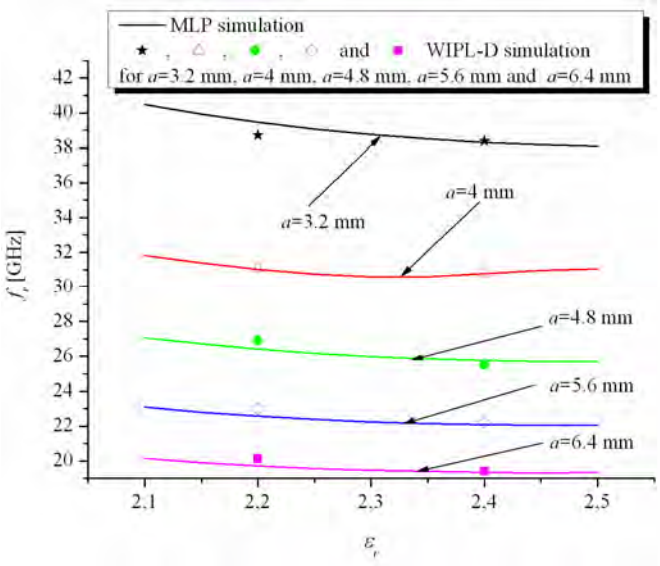


Fig. 6 Resonant frequency f_r versus dielectric constant ε_r for different pentagonal dipole dimension a

Further, proposed neural model improves pentagonal dipole modeling with great speed of work. Fig. 7. shows the dependence resonant frequency f_r on dipole dimension a and dielectric constant ε_r. This dependence is presented using 729

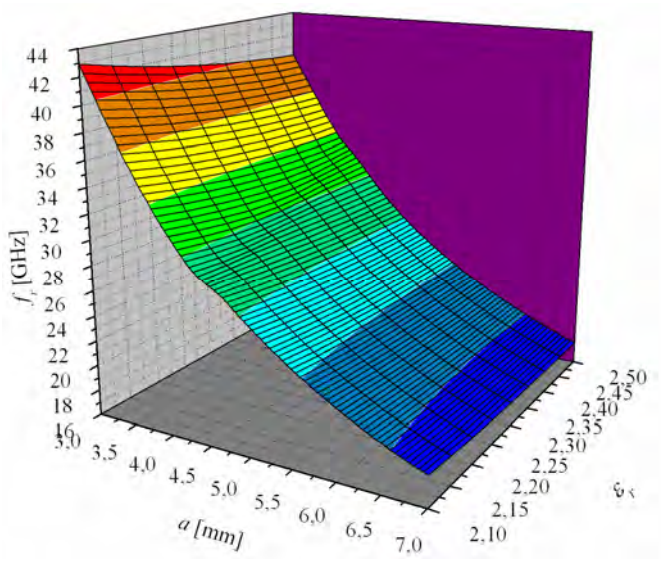


Fig.7 Resonant frequency f_r versus dielectric constant ε_r and pentagonal dipole dimension a

values of f_r obtained by MLP simulation for 2 seconds. If we use MoM simulation in WIPL-D software to obtain the same number of fr values, we will do it for a few days. For these reasons, MLP simulation is better alternative in applications where simulation has to be finished in certain period of time.

VI. Software "Pentagonal Dipole"

Software module "Pentagonal Dipole" (Fig. 8.), whose code is written in Visual C++ programming language, uses MLP3-4 model for calculate resonant frequency of printed pentagonal dipole. Range of input parameters is limited by the range of parameters from training process. When user inputs parameters, the values of resonant frequency f_r is calculated for the pentagonal dipole with these dimensions. Also, it offers the possibility of optimization. It means that user do not have to input both dipole parameters. The user can input one or no one parameters and this software module can calculate the dimension of dipole that has requested resonant frequency f_r.

VII. Conclusion

This paper presents the neural model of pentagonal dipole as an alternative to time-consuming detailed EM models. The proposed neural model is incorporated in software module "Pentagonal Dipole" that has a user friendly interface keeping similar accuracy as EM methods and surpassing EM methods with greater simulation speed.

Acknowledgements

This work was supported by the Ministry of Science and Technology Development of Serbia within the project No. TR-32052: 'Research and development of solutions for performance improvements of wireless communication systems in microwave and millimeter frequency bands'.

Fig.8 Software "Pentagonal Dipole"

References

[1] Lal Chand Godara, *Handbook of antennas in wireless communications*, CRC Press, 2002.
[2] Kin-Lu Wong, Planar antennas for wireless communications, John Wiley & Sons, 2003.
[3] B. Kolundzija, A. Djordjevic, "Electromagnetic Modeling of Composite Metallic and Dielectric Structures", Artech House, 2002.
[4] B. Milovanović, M. Milijić, A. Atanasković, Z. Stanković: "Modeling of Patch Antennas Using Neural Networks", 7th International Conference TELSIKS 2005, Niš, Serbia and Montenegro, September 28-30 2005, pp.385-388.
[5] B. Milovanović, Z. Stanković, M. Milijić, "Hybrid Empirical-Neural Model of Microvawe Slotted Patch Antennas", Eight Seminar NEUREL 2006, Belgrade, September 25-27 2006, pp.181-184.
[6] B. Milovanović, M. Milijić, Z. Stanković, „Neural Network Approach in Modeling Microwave Slotted Patch Antennas", XLIV International Scientific ICEST 2009, Veliko Tarnovo Bulgaria, June 25-27 2009, , Vol.1, pp. 37-40.
[7] M. Milijić, Z. Stanković, B. Milovanović, "Efficient Model for Slotted Patch Antenna Based on Neural Networks", 9th International Conference TELSIKS 2009, Niš, Serbia, October 7-9 2009, Vol.2, pp.384-387.
[8] M. Milijić, Z. Stanković, B. Milovanović, "Neuronski model pentogonalnog štampanog dipola", 55. Conference ETRAN, CD Paper proceeding, Banja Vrucica, Bosna&Herzegovina, 6-9. June 2011.
[9] A. Nesic,"Printed Antenna Arrays with Side Lobe Supression, Invited paper, 12-th Microcol Conference, Budapest, May, 2007.
[10] www.wipl-d.com
[11] S. Haykin, Neural Networks, New York, IEEE, 1994.
[12] Q. J. Zhang, K. C. Gupta, Neural Networks for RF and Microwave Design, Artech House,2000.
[13] Advanced Design System Documentation 2006A.

Transinformation of MPSKSC Diversity System in Weibull Fading

Milica I. Petković[1], Aleksandar T. Miljković[2], Bata V. Vasić[3] and Goran T. Đorđević[4]

Abstract – **In this paper, we analyze the diversity reception of phase-shift keying signals transmitted over Weibull fading channel. Selection combining (SC) of the signals is used on the reception. By means of computer simulation, transinformation and bit error rate are determined and constellation diagrams are drawn. The effects of the number of receiving antennas and fading depth on this system performance are observed.**

Keywords– **Phase-Shift Keying, Fading, Selection combining, Transinformation.**

I. INTRODUCTION

The development of wireless communication has exceeded the limits of the initial expectations becoming an indispensable part of modern life. There is a constant aspiration for greater date transfer and better services that are used in wireless communications.

In wireless communication, the variation of instantaneous value of the received signal, i.e. fading is one of the main causes of performance degradation. Diversity technique is certainly one of the most commonly used methods for minimizing fading effect and increasing the communication reliability without enlarging either transmitting power or bandwidth of the channel [1, 2]. In this paper diversity technique with selection combining (SC) is observed because its practical implementation simplicity.

The Weibull distribution is a flexible statistical model for describing multipath fading channels in both indoor and outdoor radio propagation environments. When the number of incoming radio paths is limited, the Rayleigh distribution may not be an appropriate fading model. Some evidence indicates that the signal amplitude can be well described by Weibull distribution in this situation. [3, 4]

In [5] the symbol error rate performance of dual-branch switched and stay combining (SSC) receivers in Weibull fading environment is studied.Paper [6] analyzes the performance of L-branch selection combining receiver over correlated Weibull fading channels in the presence of correlated Weibull-distributed cochannel interference.

While previously mentioned papers considered the error rate performance of different signal detection in Weibull fading, we evaluate transinformation in Weibull channel during Binary Phase-Shift Keying (BPSK) and Quadrature Phase-Shift Keying (QPSK) transmission while multibranch SC is applied.

[1]Milica I. Petković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: milicapetkovic86@gmail.com.

[2]Aleksandar T. Miljković is with the SingidunumUniversity, Danijelova 32, 11000 Belgrade, Serbia, E-mail: miljkovic6@yahoo.com.

[3]Bata V. Vasić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: bata.vasic@silicon-studio.com.

[4]Goran T. Đorđević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: goran@elfak.ni.ac.rs.

II. SYSTEM MODEL

Signal is transmitted from the transmitter to the receiver via channel with Weibull fading. The received signal envelope can be described by Weibull distribution given by

$$f_r(r) = \frac{m r^{m-1}}{\gamma} \exp\left[-\frac{r^m}{\gamma}\right], \quad m > 0, \quad r \geq 0, \quad (1)$$

where the index m is called the Weibull fading parameter and γ is a positive parameter related to the moments and the fading parameter. The more the value of parameter m, the less fading severity is.

Diversity techniques are applied to combine the multiple received signals of a diversity reception device into a single improved signal. Receiver with SC technique processes only one of the diversity branches, and because of that it is simpler for practical realization than maximum ratio combining (MRC) and equal gain combining (EGC) techniques. Assuming that noise power is equally distributed over branches, selection combining selects the branch with the highest signal-to-noise ratio (SNR), which is the branch with the strongest signal (see Fig. 1).

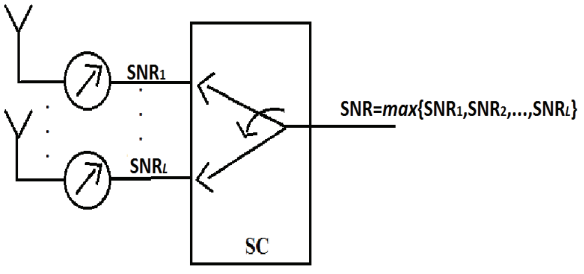


Fig. 1. Diversity receiver with selection combining scheme

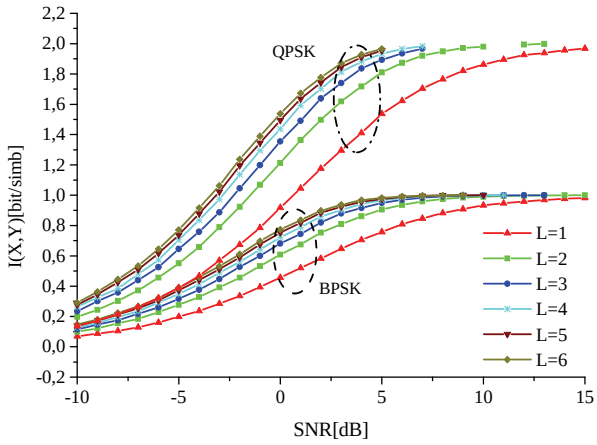


Fig. 2. Transinformation dependence on average SNR for different values of diversity order in Weibull channel with the Weibull fading parameter m=1.5

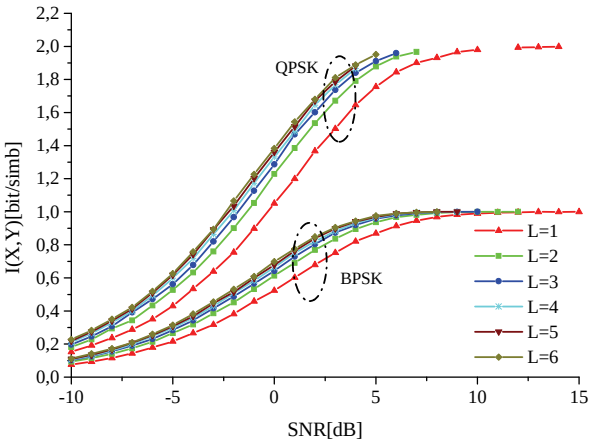


Fig. 3. Transinformation dependence on average SNR for different values of diversity order in Weibull channel with the Weibull fading parameter m=3

III. Performance Determination and Results

A. Transinformation

In this section it will be consider how transinformation depends on SNR [dB] in Weibull channel during BPSK and QPSK modulation. It will be also observed the effect of increasing the number of reception branches of SC combineron transinformation during BPSK transfer.

Transinformation can be calculated as difference between the average self-information of shipping list X, denoted by $H(X)$, and equivocation, denoted by $H(X/Y)$ and it is given by [7-9]:

$$I(X,Y) = H(X) - H(X/Y). \quad (2)$$

$H(X)$describes the measure of uncertainty about the random process X that the potential user has before the begining of transfer[7-9]:

$$H(X) = \sum_{i=1}^{m} P(x_i) ld \frac{1}{P(x_i)}. \quad (3)$$

The entropy of variable Y conditional on the variable X taking a certain value $x$represents average uncertainty about X list after receiving a symbol from Y list and it is given by [7-9]:

$$H(X/y_j) = \sum_{i=1}^{m} P(x_i/y_j) ld \frac{1}{P(x_i/y_j)}. \quad (4)$$

Equivocation illustrates the average user uncertainty about shipping list X when all symbols from Y list are received. It is given by [7-9]:

$$H(X/Y) = \sum_{j=1}^{r} P(y_j) H(X/y_j). \quad (5)$$

These formulas are used to calculate transinformation for different values of SNR.

Fig. 2 shows transinformation dependence on SNR in Weibull channel with m=1.5 during BPSK and QPSK signal transmission with selection combining on the reception during which the order of combiner is different (L=1-6). The same is given in Fig. 3, except for the parameter m=3. In Fig. 2, the transinformation values of different diversity order $L$for the same SNR can be read for BPSK. For SNR=5 dB, $I(X,Y)$=0.7573 bit/simb for L=1 and $I(X,Y)$=0.9824 bit/simb for L=6. It is noticed that when the diversity order is increased, the value of transinformation is higher. In Fig. 3, the same characteristics are read. For SNR=5 dB, $I(X,Y)$=0.8693bit/simb for L=1 and $I(X,Y)$=0.9751bit/simb for L=6. It is shown that transinformation is increased for 0.2251 bit/simb when the number of branches in SC reception is increased from 1 to 6, for SNR=5 dB when the Weibull fading parameter m=1.5 during BPSK transmission. When the Weibull fading parameter has the value m=3, the transinformation is increased for 0.1058 bit/simb. We can see that transinformation is more increased when the Weibull fading parameter m is lower, while the number L of SC branches is increasing for the same SNR. The conclusions for

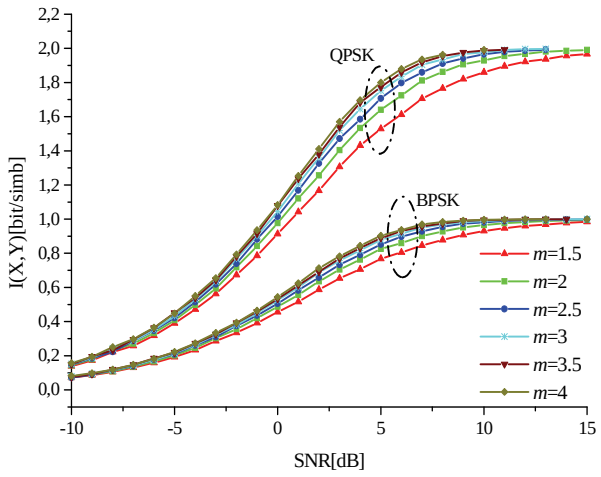


Fig. 4. Transinformation dependence on average SNR in Weibull channel withdifferent values of the Weibull fading parameter (L=1)

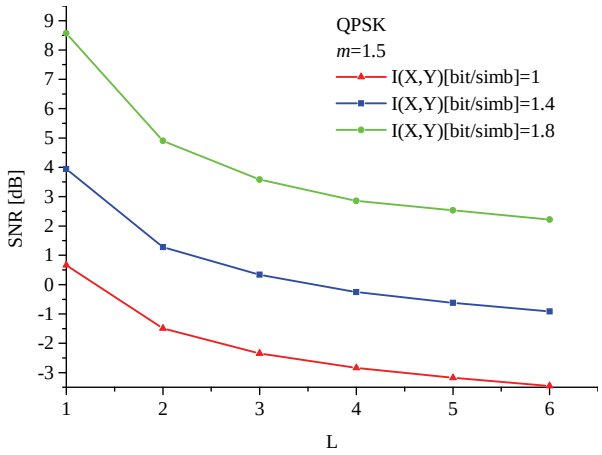


Fig. 5. Required SNR dependence on diversity order *L* in order to achieve given value of transinformation

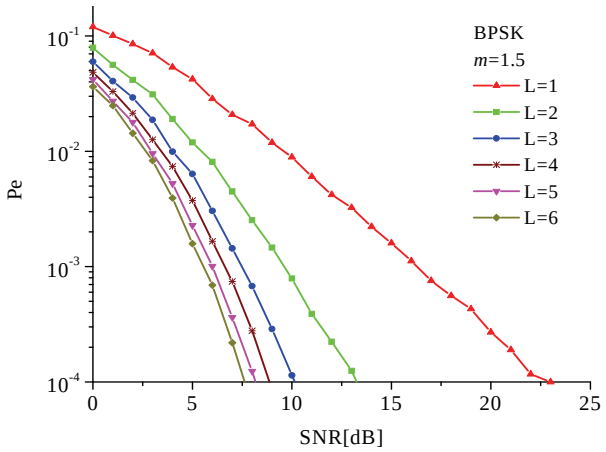


Fig. 7. Bit Error Rate dependence on average SNR for different values of diversity order in Weibull channel

QPSK are the same as for BPSK: the increase in the number of diversity order of SC combiner influences the increase in the value of transinformation and transinformation is more increased when the Weibull fading parameter *m* is lower, while the number *L* of SC branches is increasing for the same SNR.

In Fig 4. transinformation dependence on average SNR is shown during transmission of BPSK and QPSK signals over Weibull channel with different values of the Weibull fading parameter *m*. It is noticed that with the increase in Weibull fading parameter *m*, the transinformation is increased for the same values of SNR.

Fig. 5. shows the SNR dependence on diversity order for different values of transinformation in Weibull channel with *m*=1.5 for QPSK. In order to achieve transinformation of 1.8 bit/simb, if the number of the branches *L* is increased from 1 to 2, the required value of SNR is decreased from 8.5 to 5 dB. When the SC diversity order *L* is increased from 5 to 6, the required value of SNR is decreased from 3 to 2.5 dB. It means that if the SC diversity order is increased from 1 to 2, the value of the signal is lower for 3.5 dB (assuming that noise power is equally distributed over the branches), and when the SC diversity order is increased from 5 to 6, the value of the signal is lower for 0.5 dB. It is noticed that the gain in the signal strength is lower with higher SC diversity order.

The dependence shown in Fig. 6. shows how the Weibull fading parameter *m*influences on the SNR. We can notice that increasing the parameter *m* leads to lower signal strength if the value of transinformation is not changed. If we want a higher transinformation value, and the parameter *m* is same, we need a stronger signal (SNR, if the noise power is equal).

B. Bit Error Rate and Constellation Diagrams

In this section the influence of the number of SC receiving antennas *L* on it error rate (P_e) is observed. The results are presented in the graphs that show the probability of error dependence on SNR.

The bit error rate dependences on average SNR for different values of diversity order for BPSK in Weibull channel with the *m*=1.5 and *m*=3 are shown in Figs. 7 and 8.

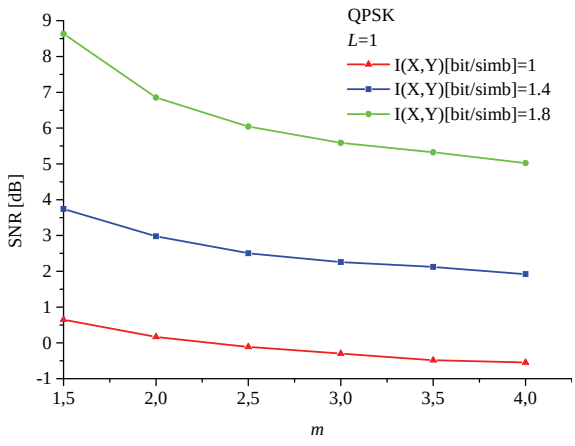


Fig. 6. Required SNR dependence on the Weibull fading parameter *m*in order to achieve given value of transinformation

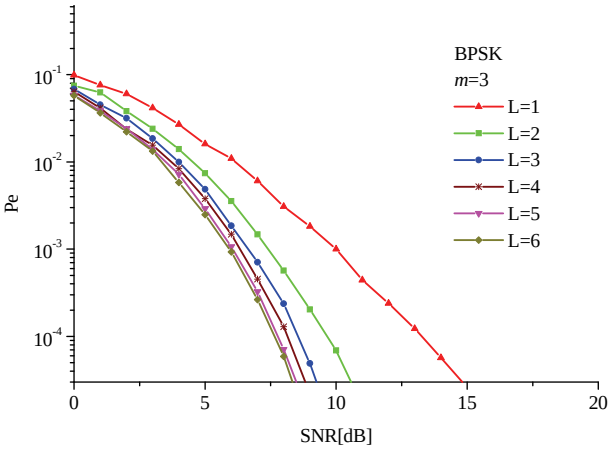


Fig. 8. Bit Error Rate dependence on average SNR for different values of diversity order in Weibull channel

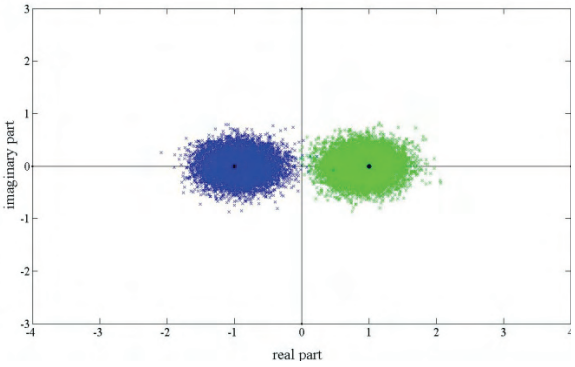


Fig. 9. Constellation diagrams for BPSK in Weibull channel with the Weibull fading parameter m=3 for L=1

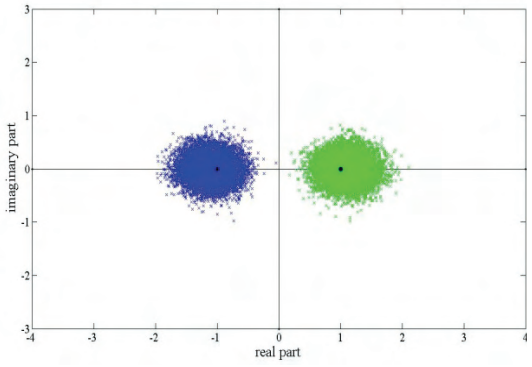


Fig. 10. Constellation diagrams for BPSK in Weibull channel with the Weibull fading parameter m=3 for L=4

By examining this results, we can notice that bit error rate is decreased when the diversity order is increased.

These results are confirmed by constellation diagrams. Constellation diagrams are suitable for graphical display of possible states, and for the influence of diversity order on decision-making. Constellation diagrams for BPSK in Weibull channel with the parameter m=3 for diversity orders L=1 and L=4 of selection combining are given in Figs. 9. and 10.

Black dots in these graphics show two possible states 0 and 1, which are different for π radians. Green and blue dots are 10.000 possible values on the reception, when 0 and 1 are sent. The black line represents the threshold of decision between the two states. It is noticed that abstraction of points around central value is lower when the order of SC diversity technique is higher (Abstraction of points is lower in the case L=4 then L=1). For higher number of branches in SC receiver, bit error rate is reduced, because fewer dots are crossing the limit of decision.

IV. Conclusion

In this paper we have analysed performance of SC receiver for BPSK and QPSK signals transmitted over Weibull fading channel. The effects of the number of receiving antennas and fading depth on this system performance are observed. It is noticed that the increase of the diversity order of SC receiver influences the increase in the values of transinformation. Also, it is noticed that the value of transinforamtion is more increased when the Weibull fading parameter m is lower. This is valid for both PSK transmitions.

Acknowledgement

This paper was supported in part by the Ministry of Science of Republic of Serbia under grant TR-32028.

References

[1] M. K. Simonm M. S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Perforamance Analysis*, New York, John Wiley & Sons, Inc., 2000.

[2] G. L. Stuber, *Principles of Mobile Communication*, Boston, Kluwer Academic Publishers, 2000.

[3] J. Cheng, C. Tellamura, N. C. Beaulieu, “Performance Analysis of Digital Modulations on Weibull Fading Channels”, Vehicular Technology Conference,vol. 1, 2003.

[4] J. Cheng, C. Tellamura, N. C. Beaulieu, “Performance of Digital Linear Modulations on Weibull Slow-Fading Channels”, IEEE Transactions on communications, vol. 52, no. 8, 2004.

[5] N. C. Sagias, G. K. Karagiannidis, G. S. Tombras, "Error-rate analysis of switches diversity receiver in Weibull fading", Electronics letters, vol. 40, no.11, 2004.

[6] M. Č. Stefanović, D. Lj. Drača, A. S. Panajotović, N. M. Sekulović, "Performance analysis of system with L-branch selection combining over correlated Weibull fading channels in the presence of cochannel interference", IEEE Transactions on communications, vol. 52, no. 11, 2004.

[7] G. Lukatela,*Statisticalof communication theory and information theory*, Beograd, Građevinska knjiga, 1981. (in Serbian).

[8] D. B. Drajić, P. N. Ivaniš, *Introduction to information theory and coding*, Beograd, Akademska misao, 2009. (in Sebian).

[9] J. G. Proakis, *Digital communications*, Fourth edition, New York, McGraw-Hill, Inc., 2001.

Session PO3:

PO3 – SIGNAL PROCESSING, METROLOGY AND REMOTE SENSING

Performance Comparison of Chaotic and Classical Spreading Sequences

Galina Cherneva[1], Elena Dimkina[2]

Abstract – **In this paper theauto-correlation and BER performances of the chaotic sequencefor spectrum spreading are investigated by numerical simulations and compared with that of the pseudorandomM-sequence.**

Keywords– **spread spectrum communication, chaotic sequence, pseudorandom sequences, correlation properties.**

I. INTRODUCTION

In conventional spread spectrum (SS) communication systems pseudorandom(pseudo-noise(PN))signals are used for broadening thespectrum by modulating the phase (in direct-sequence (DS)), orthe frequency (in frequency hopping (FH)) of the carrier signal. The most popular pseudorandom spreading sequences are the maximal length sequences shift register (M- sequences) [1].

A different type of spreading sequence, that is used for broadening the spectrum, is the chaotic sequence [2]. Use of chaotic signals as spreading codes in a SS systems has been shown to be a promising way of applying chaos digital communication purposes [2, 3].

In this paper, the auto-correlation and BER performances of the chaotic sequence, belonging to the family of piecewise-affine Markov (PWAM) maps,are investigated by numerical simulations and compared with that of the well known pseudorandom spreading M- sequence.

II. PROPERTIES OF CHAOTIC SEQUENCE

Use of chaos in a SS communicationsystems consist in replacing the standard pseudo-noise generator by a chaotic dynamical system.

A chaotic dynamical system is an unpredictable, deterministic and uncorrelated system, that exhibits noise-like behaviour trough its sensitive dependence on its initial conditions.

A nonlineardynamical system model based on its *n*previousvalues can be described as [3]:

$$\boldsymbol{x}[n+1] = \boldsymbol{f}(\boldsymbol{x}[n],\lambda), \tag{1}$$

where

$$\mathbf{x}[n] = (x_1[n], x_2[n], \ldots\ldots, x_m[n]) \tag{2}$$

is the state,

$\boldsymbol{f} = (f_1, f_1, \ldots\ldots. f_m)$ is the discrete functional transformation (maps) the state $\mathbf{x}[n]$ to the next state $\mathbf{x}[n+1]$,and $\lambda$is the bifurcating parameter.

Starting with an initial condition $\mathbf{x}[0]$, repeated applications of the map $\boldsymbol{f}$ give rise to the sequence of the points $\{\mathbf{x}[n]\}$.

A new sequence $\{\mathbf{y}[n]\}$ is obtained by mapping $\mathbf{x}[n]$ by a function $\boldsymbol{f_1}$, that is

$$\mathbf{y}[n] = \boldsymbol{f_1}(\boldsymbol{x}[n],\lambda). \tag{3}$$

So the chaotic sequences are non-converging and non-periodic sequences. A large number of these reproducible and random-like signals can be generated by changing initial value. Also the signals generated from chaotic dynamical system have broad power and flat spectrum in the frequency domain.

The correlation property of a spreading sequence plays an important role in the detection process

In this study, the family of (*l*, *t*)-tailed shifts with *l* even and $t < l/2$ [4] is employed to illustrate that the chaotic spreading sequence can achieve nearly optimal autocorrelation performance.The tailed shift map belongs to the family of piecewise-affine Markov (PWAM) maps [4].

(l, t)-tailed shifts areaffine in each of the intervals

$$x_j = \left[(j-1)/l \, , \, j/l \right]. \tag{4}$$

Theintervals from x_1 to x_{l-t} are mapped onto $x_{t+1}...x_l$ while the last *t* intervals aremapped onto $x_1...x_t$.

Consider a function $\boldsymbol{f}$: [0,1] →[0,1] iterated starting from an initialcondition $\mathbf{x}[0]$ uniformly distributed in [0,1] to produce the sequence (1).Thissequence is quantized by the bipolar threshold function

$$\boldsymbol{f_1}\text{: } [0,1] \rightarrow [1,-1] \tag{5}$$

centered at $1/2$ and the spreading sequences are taken to be $\mathbf{y}[n] = \boldsymbol{f_1}(\boldsymbol{x}[n])$ for $n = 0,\ldots., N-1$.

The auto-correlation function of the tailed shift map can be written as [4]:

$$R_n = \frac{1}{l}\sum_{i=1}^{l}\sum_{j=1}^{l} f_1(x_i) f_1(x_j) \boldsymbol{K^n_{ij}}, \tag{6}$$

where $\boldsymbol{f_1}(\boldsymbol{x}_j)$ indicates the value of $\boldsymbol{f_1}$ for all thepoints in $\boldsymbol{x}_j$;

$$\boldsymbol{K^n}_{ij} = \frac{1}{l}\begin{bmatrix} 1-h^{n-1} & 1-h^{n-2} \\ 1-h^{n} & 1-h^{n-1} \end{bmatrix}; \tag{7}$$

[1]Galina Cherneva, Higher School of Transport, Department of Electrical Engineering, Bulgaria, 1574 Sofia, 158 Geo Milev Str., E-mail: cherneva@vtu.bg

[2]Elena Dimkina, Higher School of Transport, Department of Electrical Engineering, Bulgaria, 1574 Sofia, 158 Geo Milev Str., E-mail: elena.dimkina@abv.bg

$$h = -\frac{t}{l-t}. \quad (8)$$

Figure 1shows the auto-correlation of a chaotic sequencefor l=4 , t=1 ((4,1)- tailed shift).

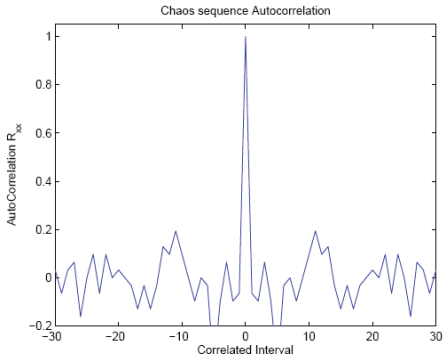


Fig.1. The auto-correlation of the chaotic sequence

III. Performances Comparison of Chaotic and M- sequences

There are several properties of chaotic sequences that make them superior to conventional PN sequences.

Firstly PN sequences are generated by shift registers which are periodic in nature. Chaotic sequences on the other band are considered very secure because of its aperiodicity und unique property of high sensitivity to initial conditions.

Secondly, for a given M -stage linear feedback shift register, there is a limit to the maximum number of sequences that can be generated. So PN sequences are less in number and this limits the security. However, with the large number of chaotic maps available and the fact that changing the initial conditions, generates a completely new sequence, there are theoretically an infinite number of chaotic sequences that can be generated. Therefore the use of chaotic spreading sequence is more secure.

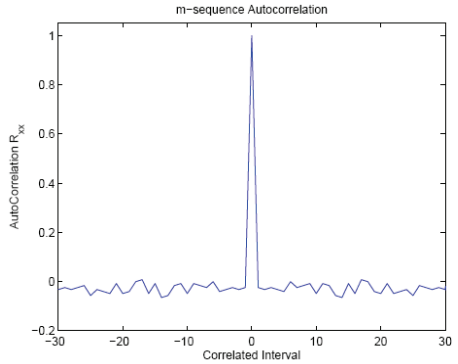


Fig.2. The auto-correlation of the M- sequence

We compare the auto-correlation properties of the chaotic sequence with that of the 31-bit M- sequence.Figure 2 shows the auto-correlation of a 31-bit M-sequence. The auto-correlation is given as N when the time lag is 0 and given as 1/N at all other times.In simulation it is found that the correlation bounds of the non-zero lags of the auto-correlation are about 0.21 and -0.24 for chaotic sequence. For M-sequence they are about 0 and (-0.24).

The bit error rate (BER) is one of the best performance measures in comparingdifferent communication systems. Following is a comparison of the BERs of conventional31-bitM-sequence to chaotic spreading sequence.Figure 3shows the performance of these sequences in AWGN Channel.

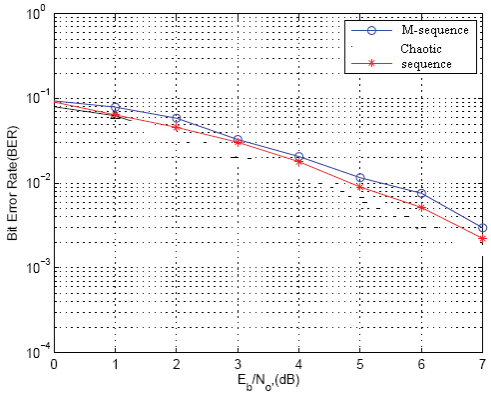


Fig.3. The BER results of using different spreading sequences

The chaotic sequenceseems toperform the best, followed by M-sequence.There seems tobe only about 0.5dB difference between these sequences.

IV. Conclusion

According to the analysis of the auto-correlation performance, it is found that the chaotic spreading sequences generated by PWAM mapshave a better auto-correlation performance than classical spreading M-sequences.

Besides this, the use of chaotic spreading sequences has some otheradvantages in the application of SS communicationsystems to classical sequences.One advantage is the availability of an enormous numberof different sequences of a given length as compared to the maximal length sequences. Generation and regeneration of chaotic sequences is very simple and involvesthe storage of only a few parameters and functions even for very long sequences. The inherent aperiodic and sensitive initial conditions features in chaotic sequences are definitelyproperties that can be used to make a system more secure.

References

[1] M.K.Simon, J.K. Omura, R.A.Sholtz, Spread Spectrum Communications. Handbook, Hardcover . 2001

[2] F.C.Lau, C.K.Tse. Chaos-Based Digital Communication Systems. Springer-Verlag. 2003.

[3] S.Mandal,S.Banerjee. AChaos BasedSpread Spectrum CommunicationSystem. Conf. Nonlinear Sys. Dynamics.Kharagpur.28-30.12.2003

[4] G.Mazzini.R.Rovatti,G.Setti. Interference minimization by autocorrelation shaping in asynchronous DS-CDMA systems:Chaos-Based spreading is nearly optimal. Electron.Lett.Vol.34.1998.

Fast Querying in Database with Images by Using Multiresolution

Mitko Kostov, Mile Petkovski, Ilija Jolevski[1]

Abstract – **In this paper we present an algorithm for fast querying in database with images. It uses multiresolution technique, all the images are decomposed in few levels and the most important wavelet coefficients are used for calculation of pseudohash. When searching for an image-query in the database, pseudohash is calculated from the image-query and used in a simple sql select statement.**

Keywords – **Wavelets, multiresolution, database, images, query, pseudohash.**

I. INTRODUCTION

Wavelet transforms have received significant attention recently from mathematicians, signal analysts and engineers as a new tool for feature extraction, signal and image compression, edge detection and denoising. Unlike the traditional Fourier techniques, wavelets are localized both in time and frequency domain. This feature makes them suitable for the analysis of nonstationary signals.

This paper considers a practical implementation of the wavelet transform for a fast searching in a database with images. Images are decomposed in a few levels. Our algorithm calculates pseudohash information from images wavelet coefficients at a low-resolution level and stores such information in a database. When searching for an image-query in the database, pseudohash is calculated from the image-query and used in a simple sql statement to select images-candidates from the database that match some defined criteria.

The paper is organized as follows. The wavelet theory is summarized in Section 2. Section 3 presents the algorithm for fast querying in a database with images. The experimental results are presented in Section 4. Section 5 concludes the paper.

II. WAVELET THEORY

The Discrete Wavelet Transform (DWT) decomposes a signal into a set of orthogonal components describing the signal variation across the scale [1]. The orthogonal components are generated by dilations and translations of a prototype function ψ, called mother wavelet.

In analogy with other function expansions, a function f is presented for each discrete coordinate t as a sum of a wavelet expansion up to certain scale J plus a residual term, that is:

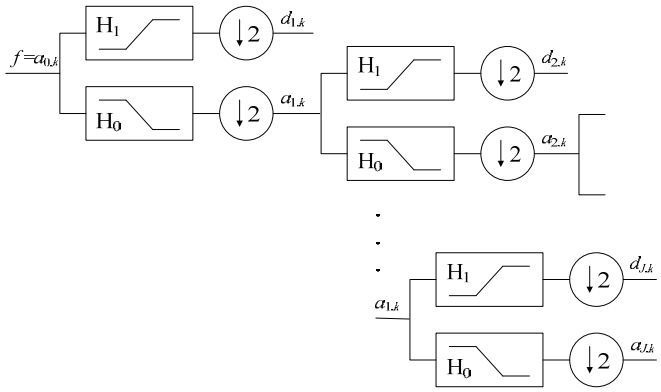


Fig. 1. Discrete wavelet transform tree.

$$f(t) = \sum_{j=1}^{J} \sum_{k=1}^{2^{-j}M} d_{jk} \psi_{jk}(t) + \sum_{k=1}^{2^{-J}M} a_{Jk} \phi_{Jk}(t) \qquad (1)$$

where ψ_{jk} and ϕ_{jk} denote wavelet and scaling function, respectively, the indexes j and k are for dilatation and translation, and a_{Jk} and d_{jk} are approximation and detail coefficients.

Wavelet decompositions and multiresolution concepts are closely related to filter bank theory. For this reason, it is helpful to view the scaling and wavelet function as a low pass and high pass filters, $\mathbf{H_0}$ and $\mathbf{H_1}$, respectively. The wavelet transform is applied to low pass results (approximations) as it is illustrated in Fig. 1.

The most popular form of conventional wavelet-based signal filtering [1], can be expressed by:

$$\begin{aligned} &\{\mathbf{A}^{(k)}, \mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \cdots, \mathbf{D}^{(k)}\} = \mathrm{DWT}(\mathbf{s}), \\ &\mathbf{s}^{*} = \mathrm{IDWT}\left(f\left(\mathbf{A}^{(k)}, \mathbf{h}^{(1)}.\times \mathbf{D}^{(1)}, \mathbf{h}^{(2)}.\times \mathbf{D}^{(2)}, \cdots, \mathbf{h}^{(k)}.\times \mathbf{D}^{(k)}\right)\right) \end{aligned} \qquad (2)$$

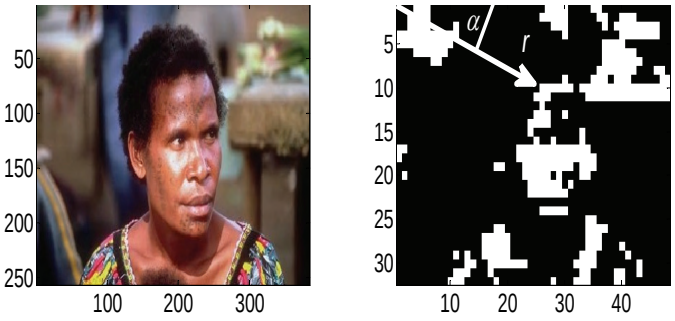
where $\mathbf{s}$ is input signal, $\mathbf{s}^{*}$ is filtered signal, $\mathbf{A}^{(k)}$ and $\mathbf{D}^{(k)}$ are approximation and detail coefficients at level k, respectively, f is a function of the modified detail and approximation coefficients, .× is element-by-element multiplying and

$$\mathbf{h}^{(k)} = \left[h_1^{(k)}, h_2^{(k)}, \cdots, h_j^{(k)}\right]^T \qquad (3)$$

are weighting coefficients of the corresponding detail coefficients at level k.

In case of conventional hard threshold filtering the weighting coefficients are

[1]Mitko Kostov, Mile Petkovski and Ilija Jolevski are with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mails: mitko.kostov@uklo.edu.mk, mile.petkovski@uklo.edu.mk, ilija.jolevski@uklo.edu.mk.



(a) (b)

Fig. 2. (a) Image with resolution 384x256; (b) The most important wavelet approximation coefficients at level 3, resolution 48x32 (calculated with haar wavelet.

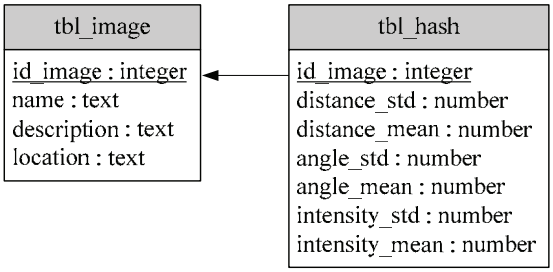


Fig. 3. Relations schema for the database with images.

$$h_j^{(k)}(hard) = \begin{cases} 1, & \text{if } \left|D_j^{(k)}\right| > \tau^{(k)} \\ 0, & \text{otherwise} \end{cases}, \qquad (4)$$

while for the soft threshold filtering they are

$$h_j^{(k)}(soft) = \begin{cases} 1 - \dfrac{\tau^{(k)}\,\mathrm{sgn}\left(D_j^{(k)}\right)}{D_j^{(k)}}, & \text{if } \left|D_j^{(k)}\right| > \tau^{(k)} \\ 0, & \text{otherwise} \end{cases}, \qquad (5)$$

where $\tau^{(k)}$ is user specified threshold for the *k*-th level details.

III. The Algorithm

A. Basic Algorithm

The main idea is to search for a particular image-query in a large database with images and to select a few images as candidates. The images-candidates would be considered visually if some of them match the image-query.

Our database keeps pseudohash information for a large number of images. The pseudohash is calculated from the most important wavelet coefficients from a low-resolution level. Namely, the wavelet transform tends to concentrate the energy of a signal into a small number of coefficients, while a large number of coefficients have small energy. By applying a threshold given with (4) the most important wavelet coefficients are selected.

For the purpose of calculating the pseudohash, RGB images are converted to YCbCr colour space, where Y is the luminance (intensity) component and Cb (blue chrominance) and Cr (red chrominance) are the blue-difference and red-difference chroma components, respectively. The Y components are taken into consideration and wavelet transform is applied. After filtering the wavelet approximation coefficients obtained at certain low-resolution level, the most important coefficients are selected. For these coefficients, three variables are defined: distance, angle and intensity, as it is illustrated in Fig. 2. An image with resolution 48x32 is shown in Fig. 2a and the most important wavelet approximation coefficients calculated with haar wavelet at level 3, are shown in Fig. 2b. For the three variables (distance, angle and intensity) both mean value and standard deviation are calculated and they compose the pseudohash information. Hence, an image pseudohash is consisted from six values: mean value and standard deviation for distance, angle and intensity of the non-zero pixels.

A database that keeps pseudohash information for the images contains a few relations with their schemas given in Fig. 3. In the relation tbl_image, the attribute id_image is the primary key. This relation contains description of the images: name, description and location (if the images are picture files in the file system). The relation tbl_hash contains pseudohash information for each image. The primary key, id_image, at the same time is foreign key that takes its values from the primary key of the relation tbl_image.

B. Extension

The algorithm is extended to search for circularly shifted images by involving Fourier transform, which is time/space invariant. Instead of calculating the wavelet coefficients from the Y component of the original images, the Fourier transform is applied to the Y components of all the images which eliminates the space information, then magnitudes of the Fourier coefficients are calculated, and at the end the inverse Fourier transform is applied. These modified images are used to calculate wavelet coefficients and pseudohash information as it is described in the basic algorithm.

When a database with schema shown in Fig. 3 is created and contains images pseudohash information, the process of querying an image can be summarized with the block-diagram shown in Fig. 4.

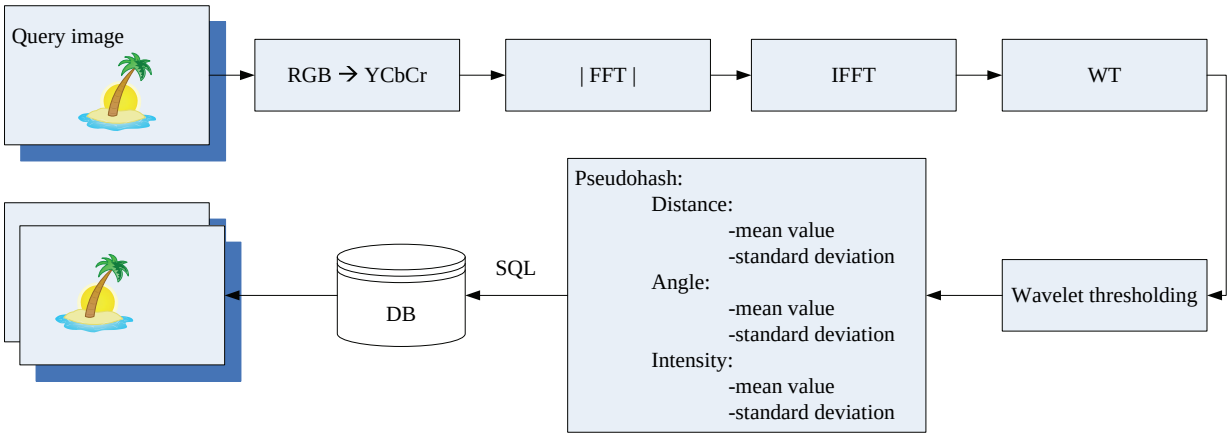


Fig. 4. Block diagram of querying with the proposed algorithm.

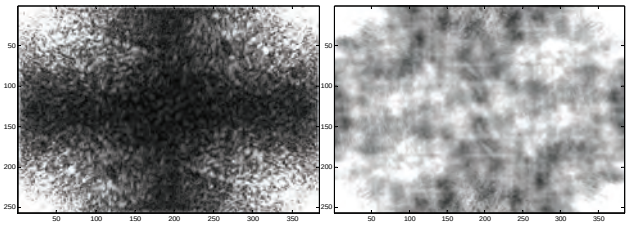
Fig. 5. Part of the database with 1000 images used for experiments.

IV. Experimental Results

In this Section, our experimental results are explained. The experiments for fast searching are made with 1000 images which pseudohash information are stored in a Microsoft Access 2007 database with schema given in Fig. 3. Some of these images are shown in Fig. 5. The database contains a lot of similar images with people, animals, landscapes, objects, etc. The database does not contain the images themselves; the images are picture files in the file system.

All these images are converted in YCbCr colour space, and the Fourier transform is applied over their Y components. Next, inverse Fourier transform is applied to the magnitudes of the Fourier coefficients. The obtained images from these successive operations of Fourier transform and inverse Fourier transform over the image from Fig. 2a are shown in Fig. 6a and Fig. 6b.

The haar wavelet transform in three levels is applied over



(a) (b)

Fig. 6. (a) Magnitude of the Fourier transform over the image from Fig. 2a; (b) Inverse Fourier transform from Fig. 6a.

(a) (b)

Fig. 7. (a) Circularly shifted version of the image from Fig. 2a; (b) Image with missing pixels.

the images obtained with inverse Fourier transform. The most important 20% pixels from the wavelet approximation coefficients at the third level are kept (Fig. 2b). It can be noticed that resolution of the wavelet coefficients at this level is 48x32, which means that only a few coefficients are taken into consideration for the calculation the pseudohash. For the image in Fig. 2, the number of non-zero pixels after applying wavelet threshold is only 307. These coefficients are normalized so the maximum intensity is 255.

For the selected wavelet coefficients, mean value and standard deviation are calculated for the variables defined with distance, angle and intensity and they are stored in the relation tbl_hash (Fig. 3).

Table I shows pseudohash data for a part of the images from our database. The all attributes domains are integer in order to save space. The value of the attribute id_image for the image from Fig. 2 is 1.

In the querying process, the same algorithm is applied over an image-query and its pseudohash data is calculated according to the block-diagram from Fig. 4. A simple SQL SELECT statement is used to select the image from the database which pseudohash values corresponds to pseudohash data calculated from the image-query.

In addition, if the image-query is a circularly shifted version of an image in the database (Fig. 7a), it has the same pseudohash data as the the original image due to the using of the Fourier transform. Moreover, if the image-query misses some pixels or parts of the image-query miss as it is shown in Fig. 7b, its pseudohash data differs slightly from the pseudohash of the original. The original image from the database still can be selected by loosening the criterion in the

TABLE I
PSEUDOHASH DATA FOR A PART OF THE IMAGES IN THE DATABASE

id_ image	dist_ std	dist_ mean	ang_ std	ang_ mean	int_ std	int_ mean
0	16	32	27	50	24	11
1	14	33	28	39	22	10
2	15	33	28	37	21	9
3	16	32	24	57	26	12
4	14	32	25	33	25	11
5	17	33	28	41	22	10
6	15	33	27	39	19	8
7	17	32	27	36	22	9
8	18	33	30	44	24	11
9	14	32	26	39	20	8
10	15	31	24	35	22	10
11	15	33	29	53	24	11
12	17	33	27	39	21	9
13	15	32	28	48	19	8
14	17	32	28	38	28	13
15	18	33	30	44	23	10

~ ~ ~

sql statement and involving the operator ‘between’ in the ‘where’ clause to select a range of data between two values. The calculated pseudohash data for the image from Fig. 7b is (- 13 33 28 39 22 10), what is very similar to the pseudohash for the image with id_image=1 (Fig. 2). The missing pixels in the image in Fig. 7b are not estimated.

V. CONCLUSION

In this paper we propose an algorithm for fast quering in database with images. The wavelet transform is exploited in order to select the most important wavelet coefficients. The most important wavelet coefficients are used to calculate pseudohash information. Pseudohash for a big number of images is stored in a database.

The algorithm is applied to an image-query and on the basis of calculated pseudohash, images-candidates from the database are selected.

Our future work will be focused on extending the algorithm for searching when the image-query is blurred or rotated, or it is manually drawn picture, a low-resolution image from a scanner or video camera.

REFERENCES

[1] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambrige Press, 1996.

[2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice – Hall, 1992.

[3] D. L. Donoho, "Wavelet Thresholding and W.V.D.: A 10-minute Tour", *Int. Conf. on Wavelets and Applications*, Toulouse, France, June 1992.

[4] C. E. Jacobs, A. Finkelstein, D. H. Salesin, "Fast multiresolution image querying", SIGGRAPH '95 Proceedings of the 22nd annual conference on Computer graphics and interactive techniques.

Algorithm for Object Recognition and Tracking on FPGA

Rosen Spirov[1], Dimiter Kovachev[2]

Abstract – **The paper presents the FPGA implementation for detecting the position of a moving object in a video sequence. By optical flow algorithms is possible to determine the approximate relative motion of the object. The frames are collected, subtracted from the background, filtered, enhanced and then the points are computed.**

Keywords– **Image processing, Adaptive filtering, Kalman filtering, FPGA, VHDL.**

I. INTRODUCTION

Optical Flow is a useful method in the object tracking branch and it can calculate the motion of each pixel between two frames, and thus it provides a possible way to get the trajectory of objects [1]. When the camera moves, a global motion will be added tothe local motion, which complicates the issue. Among the several tracking methods, point tracking is easy as shown in Fig.1.

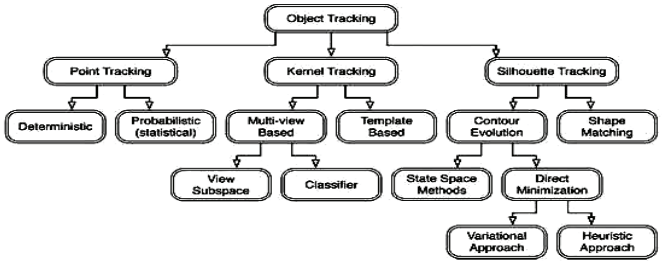


Fig.1The tracking methods

Objects detected in consecutive frames are represented by points and the association of the points is based on the previous object state which can include object position and motion. We first use image correlation to determine the global motion, and subtract the global motion for each pixel in the image so only local motion remains. We ran the optical flow on the modified frames and calculated the motion. This motion data is stored for tracking, requiring that we store all the Optical Flow data of all the pixels in each frame [2]. Having the data for all the pixels in each frame ensures that we can check each pixel's motion in each frame.

In this thesis we use a combination of optical flow and image correlation to deal with this problem, and have good experimental results.For trajectory estimation, we incorporate a Kalman Filter with the optical flow. Not only have to smooth the motion history, but have to estimate the motion into the next frame. The addition of a spatial-temporal filter improves the results in our later process.

[1]Rosen Spirov is with the Faculty of Electronic Engineering, TechnicalUniversity - Varna, 1, Studentska Str., 9010 Varna, Bulgaria,E-mail: rosexel@abv.bg.

[2]DimiterKovachev is with the Faculty of Electronic Engineering, TechnicalUniversity - Varna, 1, Studentska Str., 9010 Varna, Bulgaria, E-mail: dmk@abv.bg.

The basic hardware architecture, using extended Kalman filter-based method for calculating a trajectory by tracking features at an unknown location on Earth's surface, provided the topography is known, is given in [3]. The proposed model is implemented using VHDL and simulated and synthesized into an FPGA. The hardware design was implemented on an Altera DE2 board and Quartus II tools.

Traditionally Kalman filtering has proved to be satisfactorily in resolving many problems involved in predicting the position of moving targets [4], and is even useful for complex motion prediction. The capability of Kalman filtering to predict position allows us to overcome the artefact produced by this inherent processing latency, thus increasing the system'sreliable detection distance.

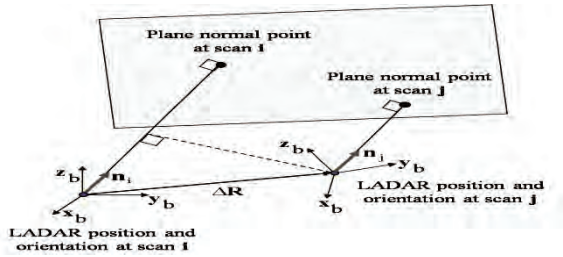


Fig.2. Navigation-changes in perceived location of the normal point between scan i and scan j are applied to estimate position changes

In Fig.2ΔR is the delta position vector, displacement vector between scans *i*, at time *ti*, and scan *j*, at time *tj*, in this case); n_i is the plane normal vector whose components are resolved in the Ladar body frame at scan epoch *ti; nj* is the plane normal vector whose components are. Note that in the navigation frame, the planar surface normal vectors at epoch's *ti* and *tj* are equal since resolved in the Ladar body frame at scan epoch *tj*; and, ρi and ρj are the shortest distances from the Ladar to the plane at epochs *ti* and *tj*, respectively stationary planar surfaces are assumed. However, expressed in the Ladar body frame both normal vectors are likely to be unequal due to the body frame rotation between epoch's *ti* and *tj*. From the geometry presented in Fig.1, a relationship can be derived between the projection of the displacement vector (between epoch's *ti* and *tj*) onto the planar surface normal vector and the change in the normal point range between scans *i* and *j* is shown in Eq.1:

$$\Delta \mathbf{R} \cdot \mathbf{n}_i = \rho_i - \rho_j \qquad (1)$$

Given M associated planar surfaces, a set of linear equations like (7) can be set up in matrix form is given in Eq.2:

$$\mathbf{H} \cdot \Delta \mathbf{R} = \Delta \boldsymbol{\rho} \quad (2)$$

Were:

$$\mathbf{H} = \begin{bmatrix} \mathbf{n}_{i,1}^T \\ \mathrm{M} \\ \mathbf{n}_{i,M}^T \end{bmatrix}, \quad \Delta \mathbf{r} = \begin{bmatrix} \rho_{i,1} - \rho_{j,1} \\ \mathrm{M} \\ \rho_{i,M} - \rho_{j,M} \end{bmatrix} \quad (3)$$

Note that a minimum of three non-collinear planar surfaces is required for the observation matrix, **H,** to be non-singular and thus allowing for a unique solution of Eq.2. The dynamic-state INS (Integrated Navigation Systems) calibration uses a Kalman filter to periodically estimate inertial error states. The estimation process is based on a complementary Kalman filter methodology [4] which employs differences between INS and laser scanner observables as filter measurements. Correspondingly, laser scanner observables of the Kalman filter are formulated as follows for the scan at time epoch t_m in Eq.4:

$$\Delta\rho_{LS}(t_m) = \begin{bmatrix} \rho_1(t_{m-1}) - \rho_1(t_m) \\ \cdots \\ \rho_N(t_{m-1}) - \rho_N(t_m) \end{bmatrix} \quad (4)$$

where N is the number of features for which is match is found time epoch t_m and t_{m-1}. Equivalent observables can be synthesized from INS measurements by transformation of the INS displacement vector into the range domain as follows in Eq.5 and Eq.6:

$$\Delta\rho_{INS}(t_m) = \mathbf{H}(t_{m-1})\left(\Delta\mathbf{R}_{INS}(t_m) + \Delta\mathbf{C}_b^n(t_m)\mathbf{l}_b\right) \quad (5)$$

$$\Delta\mathbf{C}_b^n(t_m) = \mathbf{C}_b^n(t_m) - \mathbf{C}_b^n(t_{m-1}) \quad (6)$$

As mentioned previously, filter measurements are defined as differences between inertial and laser scanner observables Eq.7:

$$\mathbf{y}_{Kalman}(t_m) = \Delta\rho_{INS}(t_m) - \Delta\rho_{LS}(t_m) \quad (7)$$

The filter operates with dynamic states only. Particular filter states include: errors in position changes between consecutive scans, velocity errors, attitude errors, gyro biases, and accelerometer biases in Eq.8:

$$\delta\mathbf{x} = \left[\delta\Delta\mathbf{R}_n^T \quad \delta\mathbf{v}_n^T \quad \boldsymbol{\psi}^T \quad \mathbf{a}_b^T \quad \mathbf{b}_b^T\right]^T \quad (8)$$

For this state vector, the observation matrix $\mathbf{H}_{Kalman}$ can be derived directly by augmenting the geometry matrix of Eq.3 with zero elements is shown in Eq.9:

$$\mathbf{H}_{Kalman}(t_m) = \begin{bmatrix} \mathbf{n}_1^T(t_{m-1}) & 0 & \mathrm{L} & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ \mathbf{n}_M^T(t_{m-1}) & 0 & \mathrm{L} & 0 \end{bmatrix} \quad (9)$$

The measurement noise matrix $\mathbf{R}_{Kalman}$ is derived from the line and planar surface estimation processes performing a comprehensive covariance analysis of the feature extraction method. As a result, the current position error contributes to the position error for the next scan where the new line is used for navigation. A position drift is thus created. Statistics of image sequences and noise can be estimated if these signals are really stationary. Generaladaptive3-D spatial-temporal filters are very complex and prohibitive for real-time implementation[5]. One advantages of KalmanFilter is that it does not need all the motion history to estimate the next stage motion, itjust needs the nearest one from the current motion.

The video sequences captured from the aircraft are quite noisy in nature. If we can get rid of some of the noise, the Kalman Filter will provide us with a better result. Based on the relationship of some neighbor pixels and the same pixel in several adjacent frames, we want to use the spatial or temporal filters before running Kalman Filter to delete the noisy data. The spatial filter is described as Eq.10:

Pt_{new}=1/2Pt +1/16(Nt_1+ Nt_2+ ……+ Nt_8)(10)

where **Pt**is the central pixel of a 3 x3 block, **N**t_1**N**t_2….**N**$t_8$are the eight neighbors of the central pixel **Pt**, **Pt***new*is the new value of **P**. The second filter is a temporal filter. In this approach the motion of a given pixel is correlated with motion of the same pixel in neighboring frames in time in Eq.11.

Pt_{new}=1/2Pt +1/4(Pt-*1*+ Pt+*1*)(11)

where Pt is the middle frame of three adjacent frames in a video sequence, Pt-**1** and Pt+**1**are the previous frame and the next frame, Ptnew is the new value of Pt. In our experiment, we run the spatial filter first, save the filtered data, and then run the temporal filter on the saved data. This combination is a spatial-temporal filter. In our work, the spatial-temporal filter works well. The areas marked with circles are the local motions, shown in fig.3.

Fig.3The frame from aerial video and its global motion removed version

The global motion removal step, the image correlation can just get integral magnitude, so the global motion detected can not reflect the real sub-pixel displacement, there is error existing. In order to delete the disturbing noisy points, we set a threshold in the data pool. In the following figures, we know all the targets are moving upwards, so we delete all the other directions' information. After that, a clean result will be shown in fig. 4.

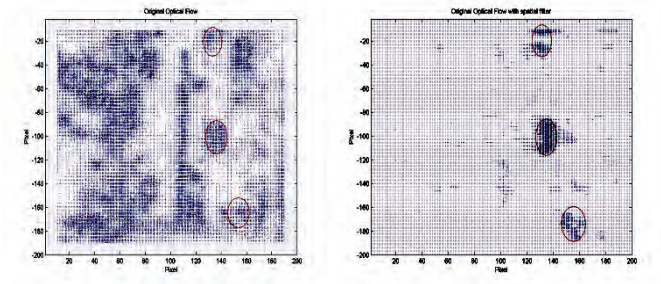


Fig.4 The local motion detection and with threshold and spatial filter

After setting the threshold, we run the spatial filter first, we find that it could increase the density of the object's vector cluster, since the filter constrains the relationship among the pixels come from the same region so it can help pixel modify its optical flow result according to the neighbors' data. Then, we run the temporal filter after the spatial filter. The spatial-temporal filter increases the density of the targets further. It is good for the estimation process.

II.ALGORITHM FOR OBJECTRECOGNITION

Each frame is fed into the program which subjects the frame to process of object recognition to achieve a noise free enhanced image containing only the object. At the rate of 1 frame per second, the enhanced image is fed to the tracking program. This analyzes each frame and computes the first white pixel that represents the object. This is under the assumption that object is made up of uniform material. This point is then plotted on a new image as the first position of the object. Subsequent frames are collected, subtracted from the background, filtered, enhanced and then the points are computed. The program of the combination of Threshold, Spatial-Temporal Filter, and Kalman Filter acquires the frames and plots the individual points. The code is:

```
%% read from the .txt file
filename1='ofx%d.txt';
filename2='ofy%d.txt';
filename3='spatialofx%d.txt';
filename4='spatialofy%d.txt';
filename5='temporalofx%d.txt';
filename6='temporalofy%d.txt';
filename9='zx.txt';
filename10='zy.txt';
imgname='original%d.jpg';
imgname1='original_shreshhold%d.jpg';
imgname2='original_spatial_filter%d.jpg';
imgname3='original_temporal_filter%d.jpg';
for index=2:30
a=sprintf(filename1,index);
b=sprintf(filename2,index);
fid=fopen(a);
[flowx,countx]=fscanf(fid,'%f');
fclose(fid);
fid=fopen(b);
[flowy,county]=fscanf(fid,'%f');
fclose(fid);
%%change the data to the image format……
%% spatial filter……
%%change the data to the image format…….
%%temporal filter……
%% kalman filter data prepare part
row=168;
col=145;
fid=fopen('icx.txt','r');
[icx,countx]=fscanf(fid,'%f');
fclose(fid);
fid=fopen('icy.txt','r');
[icy,county]=fscanf(fid,'%f');
fclose(fid);
ofx_kal(1:20)=0;
ofy_kal(1:20)=0;
%position update based on of
for imgindex=3:22
display(imgindex);
a2=sprintf(filename5,imgindex);
b2=sprintf(filename6,imgindex);
fid=fopen(a2,'r');
[ofx,count]=fscanf(fid,'%f',[200,200]);
ofx=ofx';
fclose(fid);
fid=fopen(b2,'r');
[ofy,count]=fscanf(fid,'%f',[200,200]);
ofy=ofy';
fclose(fid);
ofx_update(imgindex-2)=ofx(row,col);
ofy_update(imgindex-2)=ofy(row,col);
row=row+ofy_update(imgindex-2)+icy(imgindex);
col=col+icx(imgindex);
row=round(row);
col=round(col);
ofx_kal(imgindex-2)=ofx_update(imgindex-2);
ofy_kal(imgindex-2)=ofy_update(imgindex-2);
end
%% kalman filter data prepare part finished
%% kalman filter main part
%ZX,ZY are the testing data
%SysX,SysY are the optimized state
%PX,PY are the optimized state covariance
%eSysX,eSysY are the estimated state
%ePX,ePY are estimated state covariance
%KgX,KgY are the kalman gain
filename7='ofx_kal.txt';
filename8='ofy_kal.txt';
ZX=ofx_kal;
ZY=ofy_kal;
SysX_final(1:20)=0;
SysY_final(1:20)=0;
PX(1)=1;
PY(1)=1;
SysX_final(1)=-1;
SysY_final(1)=-1;
%for index_outside=2:11
index_outside=2;
SysX(1:20)=-1;
SysY(1:20)=-1;
for index=index_outside:index_outside+8
eSysX(index)=SysX(index-1);
eSysY(index)=SysY(index-1);
ePX(index)=PX(index-1)+1;
ePY(index)=PY(index-1)+1;
KgX(index)=ePX(index)/(ePX(index)+0.05);
KgY(index)=ePY(index)/(ePY(index)+0.05);
PX(index)=(1-KgX(index))*ePX(index);
PY(index)=(1-KgY(index))*ePY(index);
SysX(index)=eSysX(index)+KgX(index)*(ZX(index)-eSysX(index));
SysY(index)=eSysY(index)+KgY(index)*(ZY(index)-eSysY(index));
SysX_final(index)=SysX(index);
SysY_final(index)=SysY(index);
%end
fid=fopen(filename7);
dlmwrite(filename7,SysX_final,'delimiter',' ','newline','PC');
%fclose(fid);
fid=fopen(filename8);
dlmwrite(filename8,SysY_final,'delimiter',' ','newline','PC');
%fclose(fid);
fid=fopen(filename9);
dlmwrite(filename9,ZX,'delimiter',' ','newline','PC');
fid=fopen(filename10);
dlmwrite(filename10,ZY,'delimiter',' ','newline','PC');
fclose('all');
```

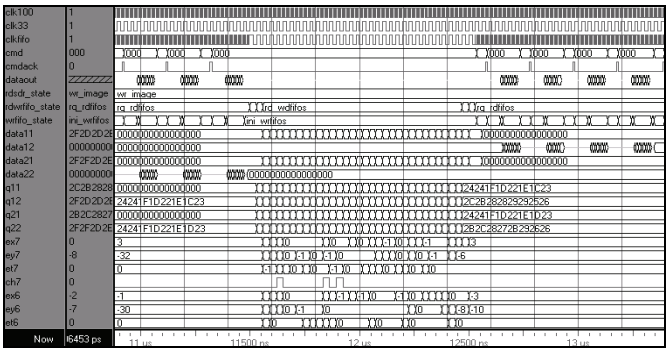


Fig.5 System simulation on ModelSim-Altera.

The advantage of parallel processing in FPGA leads to a substantial increase in performance and accuracy in processing, extraction of information than in the simulation in Matlab as shown in fig.6 and simulation in fig.7.

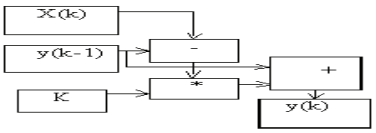


Fig.6 Pixel Kalman Filter Calculation

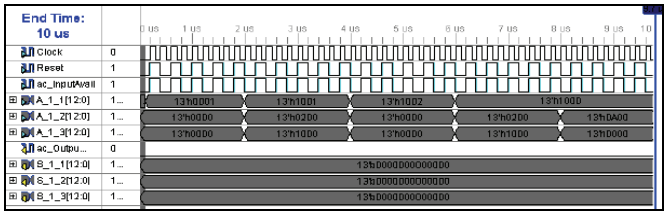


Fig.7 Behavioral simulation of the Kalman Filter

III. RESULTS

The above hardware design was implemented on an Altera Quartus II board and ModelSim, shown in fig.8and was able to operation time is about 60 clock cycle, which about 0.6us at 100MHz clock pulse, so the operation speed can be up to 1.5MHz. The whole design requires 4168 ALUTs and 241 registers- occupancy of resources is about 49% as in Tab.1.

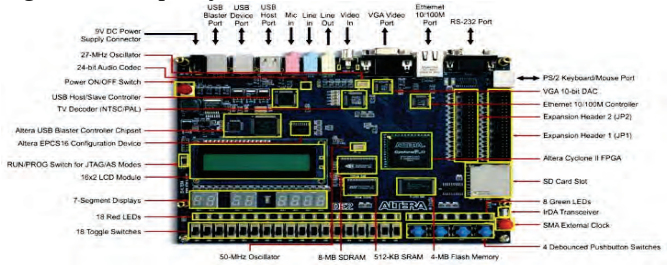


Fig.8 The Altera DE2 board

TABLE I

Sr. No.	Information	Count	%use
1	Noofslice	2145of32640	7%
2	Slice LUTs	3626of32640	14%
3	SliceLUTs Used as logic	3626of32640	14%
4	LUTFlip-Fl.pair used	4168	
5	LUT Flip-Fl. pairs with	2023of4168	49%
6	LUT Flip-Flop pair	542of4168	17%
7	Fully used FFpairs	1603of4168	41%
8	BondedIOBs	82 of 480	21%
9	DSP48Es	16 of 288	7%

The output and parameters are alignet such that one memory controller can handle reads and writes to input buffers. Hardware resources for the parameter calculation approximately 1664 Logic Cells. When we compare the outputs obtained from Matlab and FPGA, we find the outputs obtained using the AlteraDE2CycloneIIFPGAkitare computationally efficient. The pixel values are scaled and the outputs are comparable to the ones obtained using Matlab. Comparison between Optical flow [6] and Kalman Filter estimation shows the tracking result in Tab2.

TABLE II

Frame Number	Optical Flow	Kalman Filter
Frame1	-0.49883	-1
Frame2	-0.48995	-0.50219
Frame3	-0.48615	-0.54989
Frame4	-0.55982	-0.55775
Frame5	-0.65988	-0.65408
Frame6	-0.71012	-0.70244
Frame7	-0.42995	-0.43128
Frame8	-0.55981	-0.50974
Frame9	-0.45776	-0.46228
Frame10	-0.64996	-0.64356

One is from Optical Flow data and another one comes from the Kalman Filter estimation. In the first two frames, the Kalman Filter needs a period to converge. After the convergence, the Kalman Filter can estimate the motion well.

IV. CONCLUSION

Optical flow shows good results to detect and track the local motion, but suffers from some problems. One problem is when there exists both local and global motion. If these two kinds of motion exist at the same time, we should use some method to remove the global motion first and then run optical flow to detect the remaining motion. We used the image correlation first to delete the global motion and then ran the optical flow to get the local motions. We also added the Kalman Filter to the project to smooth the motion history and estimate the future trajectory of objects. The threshold and spatial-temporal filter helped the Kalman Filter deleting most of the noise efficiently. From the experimental results we see that this idea could improve tracking results for aerial video. Errors exist in the tracking process. Some are caused by the quality of the video, and others may caused by the algorithms limitations.

REFERENCES

[1] Maybeck, Peter S. Stochastic Models Estimation and Control, Vol I. Academic Press, Inc., Orlando, 32887.

[2] Hagen, E. Navigation by Optical Flow, In Proc. of IAPR Intern.Conf.-Patt. Recgn. Vol.1, 1992, pp. 700–703.

[3] Lazarov, A. D. Spatial correlation algorithm for ISAR image Reconstruction- 2000 IEEE - Rad.Conf. Virginia, USA, 7-12 P.

[4] Brown and P. Y. C. Hwang, Introduction to Random Signals and Applied Kalman Filtering, 3 rd Ed.2006.

[5] Pellerin D.and S. Thibault. Practical FPGA programming in C.Prentice Hall PTR, ISBN: 0-13-154318-0.ANN

[6] J. Diaz, E. Ros,S. Mota,Fpgabased real-time optical flow system, IEEE Transactions on, vol. 16, pp.274-279, Feb. 2006.

Image Stitching – Basic Problems and Approaches for Their Solutions

Yulka Petkova[1], Todorka Georgieva[2]

Abstract –**Combining images in a greater image is widely used in computer vision. The paper is a brief survey of published methods and techniques for image stitching. Basic steps of image stitching procedure are described. Different approaches for their solutions are presented and analysed.**

Keywords –**Image Stitching, Template Matching, Image Processing, Representative Points, Control Points, Panoramas**

I. INTRODUCTION

The goal of image stitching is to assemble together different smaller images into a single high-resolution seamless image. Image stitching has become very important nowadays. It is due to the limited abilities of conventional capturing devices to produce great images. A wide variety of applications need this process. Some of them are:

- Combining satellite images, producing a greater image of an area [20];
- Combining aviation and astronomical images[11];
- Producing large medical images for the purposes of diagnostic [7, 10];
- Combining microscope images for so called Virtual Microscopy [17];
- Videoconferences [1, 5,8, 9, 19];
- Architectural walk-through [4];
- Making panorama high-resolution photo-images, using a set of pictures, taking even by handheld cameras or mobile devices [3, 20], etc.

Image stitching process proceeds after images acquisition and preprocessing. After the images have been acquired, some processing have to be applied basically to remove undesired noises inserted during the first step. The image stitching it self consists of two basic steps – Image registration and alignment and Image assembling, as it is shown in fig. 1.

Image registration is focused on establishing correspondence between objects from one image with the objects from the other image. During the step of image alignment a proper mathematical model, which connects pixel coordinates from one image with the pixel coordinates from another image has to be formulated. After that the right alignment between images must be established. During the second step aligned images are blended, removing the seam between them.

The common problem of all the steps in image stitching process is the great computational complexity. This is the reason a lot of investigations and researches to be made in order to accelerate computations, saving the quality of input images.

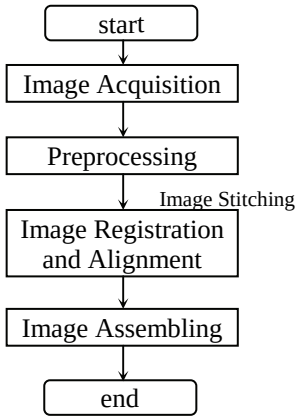


Fig. 1.Image stitching process

II. METHODS FOR IMAGE REGISTRATION

The key challenge in image stitching is the displacement of the objects in two different views of the same scene (parallax); moreover the displacement is different for objects at different depth levels (for cameras that do not have the same optical center). Thus, the objective of image registration stage is to find all matching (i. e. overlapping) images. Connected sets of image matches will be stitched later in a greater image. The problem is quadratic in the number of images, since each image could potentially match every other.

A method for image registration has to deal with a lot of problems, due to the methods of image acquisition. These problems can be:

- Differences between the intensities of the stitched images. They can be a result of changes in lightening, varying of angles between camera and lightening source, changing of the contrast between images, etc.;
- Presence of super illuminated areas, due to reflective objects in the scene;
- Presence of noise, due to the blurred lenses, dust, etc.
- Object moving during the process of images acquisition.

A method for image registration usually consists of four components. They are: a set of parameters which have to be compared; similarity measure, searching area and searching strategy.

Set of parameters for comparisons

Methods for automatic image registration can be divided in two major groups: direct or pixel-to-pixel comparison and feature-based comparison.

[1]Yulka Petkova is with the Department of Computer Sciences and Engineering at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: jppet@abv.bg

[2]Todorka Georgieva is with the Communication Department at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: todorka.georgieva@mail.bg

The advantage of direct methods [9, 17, 21], is that they use all possible information about the images, because they measure the contribution of every pixel in the image. Thus the final result is a precise registration. The basic disadvantage is that they are computationally very expensive and computational complexity is strongly dependent on the resolution, especially in the case of high-definition images. This is the reason they are not very suitable for real-time applications and they are not so widely used in practice. But direct methods can be used to refine the results, obtained by feature-based methods.

Feature-based methods use limited set of features, which are involved in comparisons. These features can be contours, edges, texture, colors, etc. [12, 13, 14, 17, 22, 24]. Set of features has to be properly chosen for any application. The basic advantage of these methods is the reduced computational complexity. As a rule these methods give precise enough results for the most of applications they are used in. It determines the great variety of algorithms, based on these methods [2].

For matching sequential frames in a video, the direct approach always works. For matching partially overlapped images for the purposes of photo-panoramas making it is not so useful.

Sometimes combinations of both feature-based and direct methods are used. These methods compare intensities of the selected control points (corners, edges, contours, etc.) [2, 18]. In other applications a feature-based methods are used for coarse registration and after that a more accurate direct comparison is used to refine the results [4].

Zoghlami et al. [24] use line segments together with control points to estimate homographies between compared images. Brown and Lowe [2] propose an approach, which basic advantage is that using invariant SIFT (Scale Invariant Feature Transform)features allows set of input images to be compared despite of rotations, scaling and different illuminations in them. Kumar et al. [10] propose to match the histograms of the component images in parts and to find correct correspondence between them (correct pixel coordinates of the relating pixels).

To summarize it can be noted that direct methods are more accurate but slower than feature-based. A lot of approaches use advantages of both direct and feature-based methods.

Similarity measures

Similarity measures are functions which return a value, corresponding to the similarity between comparing features. Similarities can be evaluated between orientation, size, color, intensities, etc. Values, obtained by computing the similarity measures are used to determine transformations, necessary for image alignment.

A suitable similarity measure should be chosen depending on type of comparison. The most frequently used similarity measures are:

- Sum of Squared Differences – SSD:

$$E_{SSD}(u)=\sum_i [I_1(x_i+u)-I_0(x_i)]^2=\sum_i e_i^2 \quad (1)$$

where $I_0(x_i)$ is the template image, sampled in discrete pixel location x_i, $I_1(x_i+u)$ is the corresponding part of a searched image, where the similarity measure is computed, u is the displacement; e_i is the residual error.

- Robust error metrics – SRD (Sum of Robust Differences):

$$E_{SRD}(u)=\sum_i \rho(I_1(x_i+u)-I_0(x_i))=\sum_i \rho(e_i) \quad (2)$$

$\rho(e_i)$ is a robust function that grows less quickly than the quadratic penalty associated with least squares.

- Sum of Absolute Differences – SAD:

$$E_{SAD}(u)=\sum_i |I_1(x_i+u)-I_0(x_i)|=\sum_i |e_i| \quad (3)$$

This function is preferred in motion estimation for video coding because of its speed. It is not proper for gradient descent approaches, because it is not differentiable at the origin.

- German-McClure function:

$$\rho_{GM}(x)=\frac{x^2}{1+x^2/a^2} \quad (4)$$

This is an example of a smoothly varying function that is quadratic for small values but grows more slowly away from the origin. Here a is a constant that can be thought of as an outlier threshold, x is the intensity.

- Normalized Cross-Correlation:

$$E_{NCC}(u)=\frac{\sum_i [I_0(x_i)-\bar{I}_0][I_1(x_i+u)-\bar{I}_1]}{\sqrt{\sum_i [I_0(x_i)-\bar{I}_0]^2 \cdot [I_1(x_i+u)-\bar{I}_1]^2}} \quad (5)$$

where $\bar{I}_o=\frac{1}{N}\sum_i I_0(x_i)$ and $\bar{I}_1=\frac{1}{N}\sum_i I_1(x_i+u)$ are the mean images (usually intensities) of the corresponding parts of compared images; N is the number of pixels in the part. E_{NCC} is in the range [-1, 1], it guaranties high reliability of the results and this makes it suitable for some higher-level applications.

Normalized Cross-Correlation has its interpretation in frequency domain, where the convolution in the spatial domain corresponds to multiplication in Fourier domain. Applying Fast Fourier Transform algorithm significantly decreases computational complexity.

Phase Correlation is also used in some applications [2] where motion estimation has to be computed.

Kumar et al. [4] propose mutual information to be used for establishing the best matching.

$$MI(I_0,I_1)=\frac{H(I_0)+H(I_1)}{H(I_0,I_1)} \quad (6)$$

where $H(I_0), H(I_1)$ are entropies of images; $H(I_0,I_1)$ is the joint entropy of two images; $MI(I_0,I_1)$ is the mutual information.

Patrik Nyman [16] proposes using of SURF (Speeded Up Robust Features) for image registration and alignment. SURF are compared according to the Euclidean distance and the minimum distance between them is found.

Brawn et al. [2]use SIFT features which are located at scale space maxima/minima of a difference of Gaussian function. Scale and orientation establishment at each feature location

gives a similarity invariant frame in which to make measurements. SIFT features are invariant under rotation and scale. This is the reason their method can handle images with arbitrary orientations and zoom.

To summarize it can be noted that the most precise similarity measure is NCC, but it is computationally very expensive. A trade-off between precision and computational complexity is usually made depending on the requirements to the particular application.

III. Searching Strategy

Searching strategy is an algorithm, which decides how to choose next transformations from the searching set. Searching methods usually use pattern matching for image alignment [18].

The simplest searching algorithm is an exhaustive comparison with the template which calculates the similarity measure for each position and transformation in the searched set (so called brute-force algorithm). In this way the optimal similarity measure is guaranteed to be the globally optimal measure. But this algorithm has got a great computational complexity, thus it works slowly which makes it unworkable for real-time applications.

Acceleration of algorithms, keeping high reliability and precision, is another provocation to researchers. One of the possible decisions is to restrict the searched positions, for example – to search around the most probable position which is known in advance [2].Other techniques use different feature sets. In this method binary images, generated from the selected features(usually corner or edge points) are compared.

Another technique is so called coarse-to-fine search. It is an iteration process which uses a coarse determination of the most probable position. After it has been determined a fine search around it is performed in order to refine the result.

Registration with step search strategy is proposed by Tzi in [23]. In this strategy only five positions are evaluated for their respective similarity measures in each iteration of the search. The five positions include a center point and four points respectively in the north, east, south and west of the center point. The distances between the center point and the four other points are the step sizes. Initially, the vertical step is half the distance between the center point and the top of border, and the horizontal step size is half the distance between the center point and the left border. During the processing step sizes are corrected. When both step sizes reach 1, the similarity measures of eight positions around the center point are evaluated to determine the position with the best similarity measure. Since the step size is reduced by half in each iteration, the search algorithm converges a solution very quickly.

Some algorithms use combinations of methods. One such method is registration with binary edge image and restricted search set. With this combination, the chance of misalignment can be reduced by limiting the search to a defined neighborhood, within the optimal overlapping position is guaranteed to occur.

Another example of combined algorithms is a combination of restricted search set and step search. Chia-Yen Chen in [4]proposes the feature set to be the averaged intensity or the binary edge image. The similarity measure is either the sum of absolute differences or the standard deviation of the intensity differences. A restricted search set is used to decrease the chance of the algorithm converging towards the local optimum away from the position with the globally optimal similarity measure.

Other techniques use image pyramids to find the optimal match and transformation between two images at successively higher resolutions. Once the pyramids have been built, the registration is fast, but the process of building images is very time consuming.

To summarize it can be said that brute-force searching strategy is the most reliable but computationally most expensive. The other strategies are faster, but there is a real chance to skip the right correspondence between images.

IV. Methods for Image Assembling

After the steps of registration and alignment images have to be assembled or blended in a common image. Image assembling is a process of adjusting the values of pixels in the registered and aligned images, such that when the images are joined, the transition from one image to the other is invisible. With other words - the basic problem of this process is how to merge the images so that the seam between them to be visually undetectable. A seam is the artificial edge generated by the intensity differences of pixels immediately next to where the images are joined.

Methods for image blending can be separated in two categories: transition smoothing methods and finding the optimal seam.

Transition smoothing methods try to minimize the seam between images smoothing the edges of the image. The basic disadvantage of these methods is that blurry areas are created. Recently some methods, using multi-resolution blending, wavelets and gradient-domain blending are published [21].Gradient blending calculates a smooth weighted blend from one side of the overlapping parts to the other. The effect reduces issues like varying background intensities and provides the smooth edge transition between adjacent images. These methods need finding a least square solution of a Poison equation which is computationally very expensive [6].

Methods with optimal seam [21] try to put the seam where differences between images are as small as possible. Patrik Nyman [16] proposes so called watershed segmentation to be used for seam position establishment. In watershed segmentation one regards the image as a topological map, where watershed barriers define the different segments. The source and sink of the problem are defined as the non-overlapping regions from the first and second image respectively. The segments from the watershed algorithm are set as the nodes in the graph. The total sum in the difference image of the boundary pixels between two segments is set as the weight between the segments. Using a max-flow algorithm the minimum cut is found. It is then the optimal seam between both images.

One approach to remove the seam is to perform the intensity adjustment locally, within a defined neighborhood of

the seam [15]. Another approach is to perform a global intensity adjustment on the images to be merged, so that apart from the intensity values outside the overlapping regions may also need to be adjusted [15].

Chia-Yen Chen [4] investigates four different intensity adjustments: linear distribution of intensity differences; linear distribution of median intensity differences; intensity adjustment with respect to median filtered regions and intensity adjustment with respect to corresponding pixels in overlapping region. The second method used by him gave the best results. The reason is that the adjustment to the original intensity levels is kept to a necessary minimum. The amount of adjustment is also proportional to the intensity differences between the joined images. Therefore, large intensity differences between the images indicate that the merged images may not retain the quality of the original images as well as when the intensity differences are small.

For obtaining a seamless stitching Kumar et al. [10] propose triangulation averaging to be applied on the overlapping area of the assembled images. The overlapped area of the left image is multiplied with averaging image whose intensity starts with 0 and changes to 1. The overlapped area of the right image is multiplied with averaging image whose intensity starts with 1 and changes to 0. The algorithm does not change the original quality of the images except in the overlapped area, but it produces a seamless image.

No quantitative evaluations are used for the quality of image blending. Visual evaluation is always used.

To summarize it can be noted that better methods for seamless image blending are based on finding an optimal seam between assembled images.

V. CONCLUSIONS

In this paper different methods and approaches for image stitching are briefly considered in order to be helpful for better understanding of different stages involved in generation of panoramic images. According to the published results a comparison between them is made.

Set of parameters, taking part in comparisons, similarity measures and different searching strategies are described. Methods for seamless image assembling (or blending) are also briefly presented.

ACKNOWLEDGEMENTS

The work presented in this paper was partially supported within the project BG 051PO001-3.3.04/13 of the HR Development OP of the European Social Fund 2007-2013.

REFERENCES

[1] Adam M., C. Jung, S. Roth, G. Brunnett, Real-time Stereo-image Stitching using GPU-based Belief Propagation, VMV 2009, pp. 215-224

[2] Brown M., D. G. Lowe, Automatic Panoramic Image Stitching using Invariant Features, International Journal of Computer Vision, Volume 74 Issue 1, August 2007

[3] Brunton A., C. Shu, Belief Propagation for Panorama Generation, 3D Data Processing Visualization and Transmission, International Symposium on, pp. 885-892, 2006

[4] Chia-Yen Chen, Image Stitching – Comparisons and New Techniques, CITR-TR-30, Computer Science Department of the University of Auckland, (http://www.tcs.auckland.ac.nz)

[5] Coleshill E., Alexander Ferworn, Panoramic Spherical Video – The Space Ball, ICCSA, 2003

[6] Eden A., M. Uyttendaele, R. Szeliski, Seamless Image Stitching of Scenes with Large Motions and Exposure Differences, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2006), pp. 2498-2505

[7] Goossen A., T. Pralow, R. Grigat, Automatic Stitching of Digital Radiographies using Image Interpretation, 2009

[8] Heinrich G., M. Leitner, C. Jung, F. Logemann V. Hahn, A Platform for Audiovisual Telepresence using Model- and Data-based Wave-field Synthesys, 125th AES Convention San Francisco, 2008

[9] Irani M., P. Anandan, About Direct Methods, Vision Algorithms: Theory and Practice, No 1883 in LNCs, pp. 267-277, Springer-Verlag, 1999

[10] Kumar A., R. S. Bandaru, at al., Automatic Image Alignment and Stitching of Medical Images with Seam Blending, World Academy of Science, Engineering and Technology, 65, 2010, pp. 110 – 115

[11] Majumdar J., S. Vinay, S. Selvi, Registration and Mosaicing for Images Obtained from UAV, International Conference on Image Processing and Communications, pp. 198-203, 2004

[12] McLauchlan P., A. Jaenicke, Image Mosaicing Using Sequential Bundle Adjustment, Image and Vision Computing, 20 (9-10):751-759, 2002

[13] Mikolajczyk K., C. Schmid, Scale& Affine Invariant Interest Point Detectors, International Journal on Computer Vision 60 (1), 63-86, 2004, Kluwer Academic Publishers

[14] Mikolajczyk K., T. Tuytelaars, C. Schmid, A Comparison of Affine Region Detectors, International Journal on Computer Vision Springer Science + Business Media, Inc., DOI: 10.1007/sl 1263-005-3848-x

[15] Milgram D., Adaptive Techniques for Photomosaicing, IEEE Transactions on Computers, vol. C-26, pp. 1175-1180, 1977

[16] Nyman P., Image Stitching using Watersheds and Graph Cuts, Centre of Math. Sciences, Lund University, Sweden, 2007

[17] Romero E., F. Gomez, M. Iregui, Virtual Microscopy in Medical Images: a Survey, *Modern Research and Educational Topics in Microscopy,* Formatex 2007, pp. 996-1006

[18] Ritter G., J. Wilson, Handbook of Computer Vision Algorithms in Image Algebra, CRC Press, 1996, pp. 225-239

[19] Steedly D., et al., Efficiently registering video into panoramic mosaics, In Tenth International Conference on Computer Vision, Beijing, China, 2005, pp. 1300 - 1307

[20] Szeliski R., H. Shum, Creating full Vew Panoramic Image Mosaics and Environment Maps, SIGGRPH'97, Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 251-258, 1997

[21] Szeliski R., S. Kang, Direct Methods for Visual Scene Reconstruction, In IEEE Workshop on Representation of Visual Scenes, pp. 26-33, Cambridge, 1995

[22] Tuytelaars T., L. V. Gool, Matching Widely Separated Views Based on Affine Invariant Regions, International Journal on Computer Vision 59 (1), 61-85, 2004

[23] Tziritas G., C. Labit, Advances in Image Communication, vol. 4, Motion Analysis for Image Sequence Coding, Elcevier, 1994

[24] Zoghlami I., O. Faugeras, R. Deriche, Using Geometric Corners to build a 2D Mosaic from a Set of Images, In Proceedings of the International Conference on Computer Vision and Pattern Recognition, IEEE, June 1997

Effectiveness of Statistical Methods for Encoding Images

Todorka Georgieva[1]

Abstract- **The paper presents effectiveness of statistical methods for encoding images based on four coding methods: block method, coding set length with cipher, one dimensional coding method with modified Huffman cod (MH) and with READ code.Coding of static images could be optimized with scanning the filmed object and determine if this is a text, picture or both types. Each coding method has a need from processor power for satisfaction the speed needs, on the other hand the size of the coded chunk should be mentioned, however.**

Keywords **– effectiveness, coding, methods, images**

I. Introduction

Noise protection analysis is based on four coding methods:

1.Block method;

2.Coding set length with cipher;

3.One dimensional coding method with modified Huffman code (MH);

4.Method with READcode.

All considered methods accomplish lossless coding,if these isn't any noise in connection channel.[1]

The analyze results from comparison include presumptionthat mistakes in binary digit appear independently of each other. We use the following quantitative parameters for comparison, determining distinctions between original and reproduced figure:[2]

1.Coding data bits - the number of coding bits after compression;

2.Coding bits - the number of coding bits after compression, including all subsidiary bits;

3.Compression coefficient **(CC_1)**–ratio between coding data bits to the number of figure elements;

4.Compression coefficient (**CC_2**) - all coding to the number of figure elements ratio;

5.Bit-error probability (**BEP**)-ratio between errors in accepted figure and quantity of transmitted elements ;

6.Transition probability from white to black element (**TPB**)- ratio between quantity of elements, transformed from white to black, to all white elements;

7.Transition probability from black to white element(**TPW**) - ratio between quantity of elements, transformed from black to white, to all blackelements;

8.Symbol-error probability (**SEP**)– ratio between the amount of transitions from white to black and black to white, and quantity of elements in the figure.[3]

All coding methods have different noise protection and influence of channel noises is different for each method.

II. Analytical Results

The comparison analyze is with the defined parameters:

-the figure/image has 1024 * 256 elements.

-the figure/image format is 100*50 mm.

-the size of raster surface is 100 * 100 μm.

-probability of bit-error in the communication channel is in the rate 4.10^{-6} - 10^{-4}.

-number of sections causing errors in the channel (**NES)**

-number errors in image restoration (**NEI**)

The results are presented as follows:[4]

TABLE I
BLOCK METHOD

NES	NEI	BEC	SEP	TPW	TPB	BEP
1	167	0,4	1,924	1,801	9,93	1,67
3	169	1,2	1,928	1,82	9,94	0,563
5	173	2	1,94	1,89	9,97	0,346
7	183	2,7	1,98	2,11	10,1	0,261
9	9161	3,5	17,8	98,6	54,37	10,179
11	9163	4,3	17,82	98,6	54,42	8,33
13	83687	5	163	981	459,5	64,375
15	92568	6	180	1090	507	61,712

[1]Todorka Georgieva is with the Faculty of Electronic, TU Varna, Telecommunication Dep., Studentska 1, 9010 Varna, Bulgaria, E-mail: tedi_ng@mail.bg

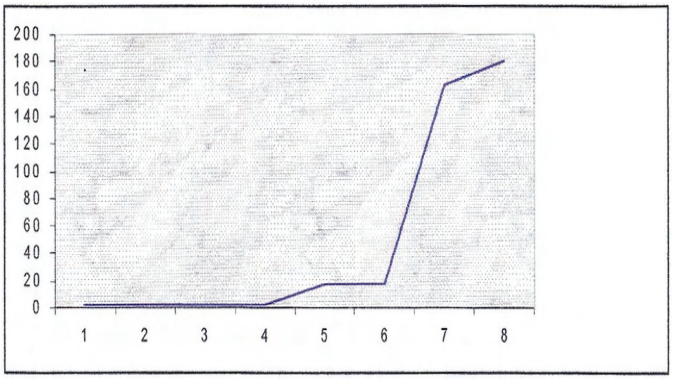


Fig. 1.Probability of error SEP depending on the increase of ***NES***. Block method

TABLE II
METHOD, CODING SET LENGTH WITH CIPHER

NES	NEI	BEC	SEP	TPW	TPB	BEP
1	764	0,6	1,56	1,29	2,25	7,64
3	1232	1,2	2,35	1,67	5,54	4,107
5	1656	3,1	3,16	2,4	6,74	3,312
7	2237	4,3	4,27		9,74	3,196
9	2624	5,5	5,02	3,55	11,88	2,916
11	2930	6,8	5,59	4,32	11,58	2,664
13	3070	8	5,87	4,79	10,9	2,362
15	3352	9,2	6,41	4,84	13,7	2,235
17	3473	10	6,64	4,98	14,38	2,043

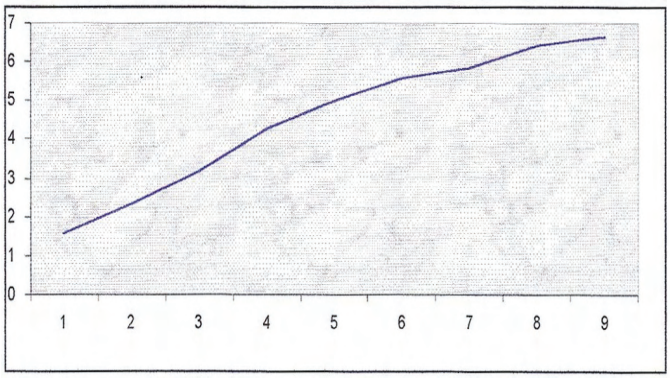


Fig. 2. Probability of error SEP depending on the increase of ***NES***. Method, coding set length with cipher

TABLE III
ONE DIMENSIONAL CODING METHOD OF SET LENGTH WITH MODIFIED HUFFMAN CODE (MH)

NES	NEI	BEC	SEP	TPW	TPB	BEP
1	89	0,7	0,17	1,02	4,89	0,89
3	402	2,1	0,768	4,5	22,6	1,34
5	571	3,5	1,09	6,42	31,9	1,142
7	668	4,9	1,27	7,6	36,9	0,954
9	1003	6,2	1,92	11,5	55,1	1,114
11	1772	7,6	3,39	18,5	106	1,611
13	2174	9	4,15	23,2	126	1,672
15	2415	10	4,62	24,4	148	1,61
17	3263	12	6,23	34,2	194	1,919

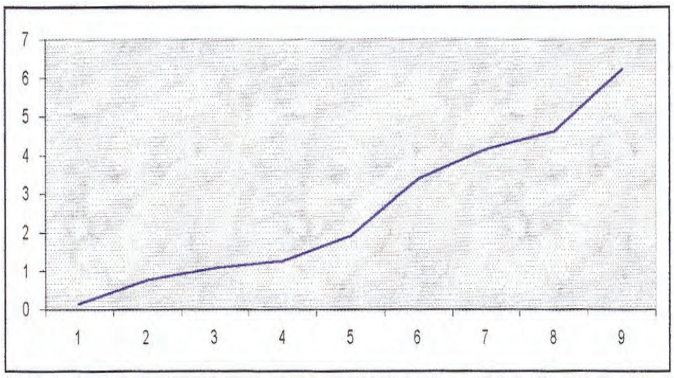


Fig. 3. Probability of error SEP depending on the increase of ***NES***.Huffman code (MH)

TABLE IV
CODING METHOD WITH READ CODE (κ = 2)

NES	NEI	BEC	SEP	TPW	TPB	BEP
1	668	0,86	1,315	7,652	3,89	6,88
3	900	2,6	1,72	10,157	5,02	3
5	1212	4,3	2,316	10,78	8,12	2,242
7	23805	6	45,49	272,38	131,04	34,007
9	24107	7,7	46,07	275,9	132,6	26,786
11	25133	9,5	48,031	287,66	138,31	22,848
13	25280	11,2	48,312	287,45	140	19,446
15	25586	13,1	48,897	290,83	141,7	17,057
17	26447	15	50,54	299,2	147,19	15,557

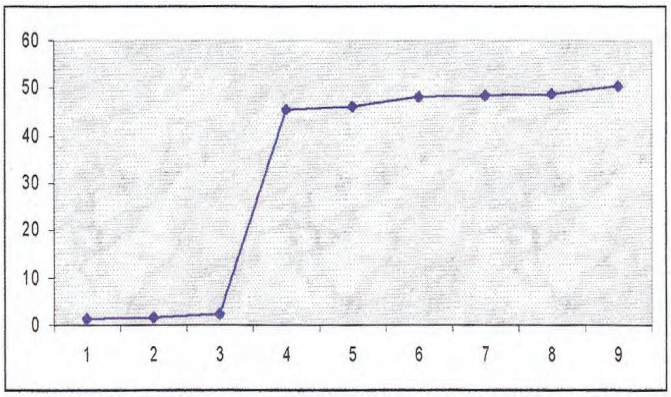


Fig. 4. Probability of error SEP depending on the increase of ***NES***. Coding method with READ code (κ = 2)

TABLE V
CODING METHOD WITH READ CODE (K = 4)

NES	NEI	BEC	SEP	TPW	TPB	BEP
1	2	0,997	0,0038		0,02	0,02
3	490	0,936	0,936	6,331	2,358	1,633
5	13917	26,59	26,59	156,8	77,75	27,834
7	14624	27,59	27,95	164,7	81,74	20,891
9	14948	28,57	28,57	169,5	83,02	16,609
11	9605	18,38	18,35	100,6	57,23	8,732
13	9617	18,39	18,38	100,87	57,23	7,398
15	11105	21,22	21,22	107,8	70,155	7,403
17	12429	23,75	23,75	124,4	76,75	7,311

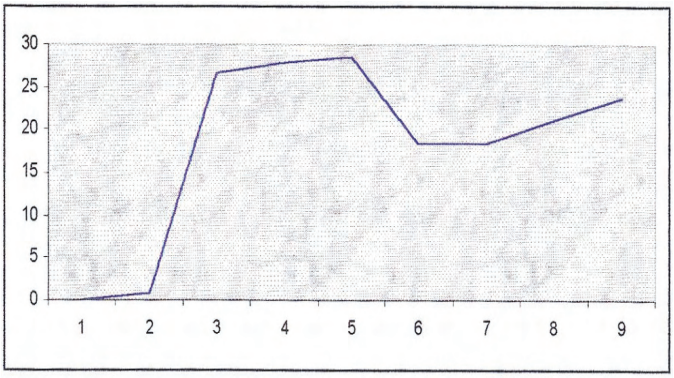


Fig. 5. Probability of error SEP depending on the increase of ***NES***. Coding method with READ code (κ =4)

We analyze results from comparison of the five coding methods with program modeling and parameters of transmitted message.
The results are present in Table VI.

TABLE VI
CODING METHOD. ANALYZE RESULTS.[6]

Coding method	Code sequence	Full Code sequence	CC_1	CC_2
Block method	253 824	253 824	0,484	0,484
With cipher	76 648	162 016	0,146	0,309
Modified Huffman code (MH)	137 776	143 920	0,263	0,274
READ code k = 2	109 744	115 888	0,209	0,210
READ code k = 4	36 176	102 320	0,183	0,195

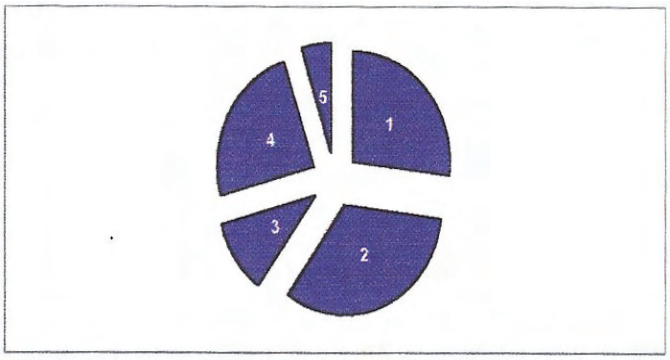


Fig.6 . Ratio between the studied methods of **SEP** in 3 sections, causing an error.

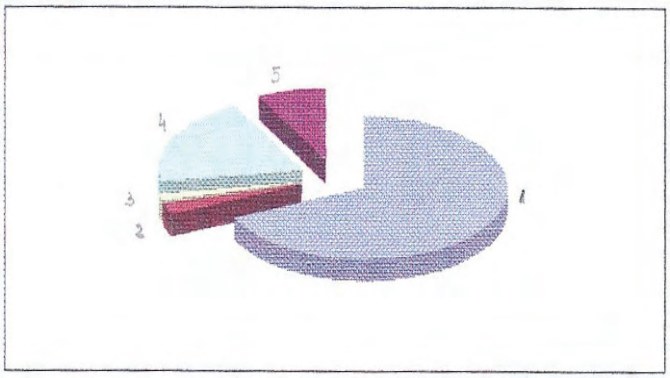


Fig.7. Ratio between the studied methods of **SEP** in 13 sections, causing an error.[5]

Bi-dimensional methods as block method lead to errors in some letters or words anywhere in the text, but one-dimensional have determinate effect in continuation of one working line.

Increasing of error probability increase general number errors in figure restoration and increase number of transformations from black to white element.

III. CONCLUSION

In block method errors in code sequence may balance each other in reproduction of transmitted message.

READ-code is better in noise protection than block method and reduce digital flow five times (when k=4, CC2 = 0.195 bit). For standard resolution parameter k=2, and for higher resolution k=4.

Coding set length with cipher is suited for increasing noise protection fore given period of time, but he has big excess.

Modified Huffman code (MH), used for fax machines group 3 has high stability, high compression levels, but in comparison with READ-code has more transformations white-black and black-white, that decreases it noise protection.

Probability of bit-error in the communication channel (**BEC**) *(10^{-5});

Obviously, all coding methods have different noise protection.

Increasing of error probability increase general number errors in figure restoration.

That also increase number of transformations from black to white element.

The effectiveness of coding methods is judged by probability of errors as a function of factors "worsening" the transfer of code sequences in communication channels.

REFERENCES

[1] Analysis of statistical noise protection coding methods, Georgieva T., Dimova R., RADIOELEKTRONIKA, Brno, Czech Republic, 2003.

[2] White book of bulgarian telecommunications, 2006

[3] The Signal Processing and Multimedia Group – The University of British Columbia ,2000г

[4] Coding of Still Pictures , JBIG Committee, Columbia, 2002г.

[5] ITU-T Recommendation SG8, 2002

[6] Evaluation of the effectiveness of information in telecommunication systems for data exchange,Georgieva T.,Union of scientists, Varna, 2010.

Comparison Between Steepest Ascent and Genetic Algorithm Optimization Methods in Series Based Software Direct Digital Synthesis of Signals

Mariana Shotova[1]

Abstract – **In this paper an optimization of the software method for Direct Digital Synthesis of signals, based on series approximation of sine wave is discussed. A 5th order polynomial is investigated and an optimization of the polynomial approximation is proposed and discussed. Two methods for spectral optimization, aimed at the reduction of the spurs' level, are compared – steepest ascent and genetic algorithm optimization. An increase of the dynamic range is achieved.**

Keywords – **DDS, series, polynomial, optimization, GAO, steepest ascent.**

I. INTRODUCTION

The Direct Digital Synthesis (DDS) is a technique for generating a high quality sine wave through a digitally defined frequency. The software implementation of DDS (SDDS) based on digital signal processor has two main versions – using ROM table of the sine wave, and series approximation of the sine wave. Measure of quality is the spurious free dynamic range D of the spectrum of the synthesized signal[1].

The advantages of the SDDS exploiting series approximation are: elimination of the ROM table, and better dynamic range. A drawback is the bigger number of the required mathematical operations, which results in lower sampling frequency.

There are several basic polynomial approximations which can be used in DDS [2,3,4].

The SDDS with 5th order polynomial is investigated here. By taking advantage of the sinus' symmetry – using approximation in the range [-π/2, π/2], the even-order components are eliminated. The ***polyfit*** approximation, normally used, is based on minimization of the root-mean-square deviation, while in DDS case it is important to minimize the spectral spurs. Therefore two methods for spectral optimization, aimed at reduction of the spur' levels, are compared.

The first one is based on the steepest ascent method [5] – it searches for a global maximum over a 3 dimensional area by alternating two coefficient of a fifth order polynomial. The application of steepest ascent to SDDS results in several sets of coefficient of the suggested polynomial, which increase the dynamic range.

A comparison between the steepest ascent and the genetic algorithm optimization(GAO) methods[6,7,8] is made. While the steepest ascent method provides quick search over the parameter space, GAO deals with the individuals in a population over several generations, and thus the time for searching the fittest coefficients increases. GAO is expected to find the global maximum even in the case of several extremes, at the cost of the bigger number of iterations and calculations.

[1]Mariana Iv. Shotova is with the Faculty of Electronics, Technical University of Varna, 9010, Varna, Bulgaria, E-mail: marianashotova@gmail.com

II. POLYNOMIAL APPROXIMATION OF 5TH ORDER

*A. Polynomial approximation with **polyfit***

A polynomial approximation of 5th order may be represented by the following equation:

$$\sin(\alpha) = a_1 - a_3\alpha^3 + a_5\alpha^5 , \quad (1)$$

where a_i, i=1,3,5, are the coefficients of the sinus approximation. The range of the argument is [-π/2, π/2] and thus the even-order components of the polynomial are eliminated. The computation of the polynomial coefficients is implemented by MATLAB's *polyfit* function, which minimizes the root-mean-square error. Thus for the polynomial (1) the coefficients are : a_1= 0.99977007, a_3= = -0.16582379 and a_5= 0.00757279.

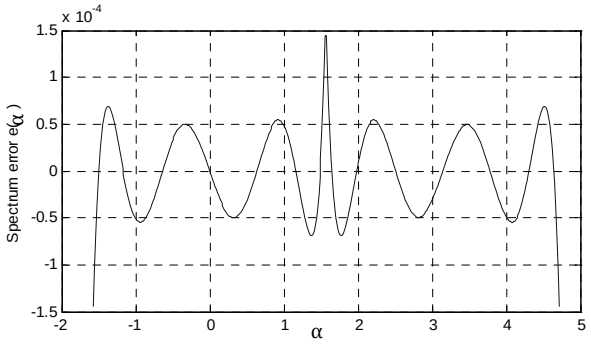


Fig.1 Sine wave error with 5th order polynomial

A plot of calculation error $e_{(\alpha)}$ over the range [-π/2, 3π/2] is shown in Fig.1.

The rearranged form of the series, closer to DSP is:

$$\sin(\alpha) = \alpha(a1 + \alpha^2 (a3 + \alpha^2 a5)) . \quad (2)$$

The argument α^2 is calculated in advance. The number of mathematical operations here is 3 multiplications and 2 addition/subtractions.

In the case of DDS the error spectrum is of interest (Fig.2). An error signal, which contains 16 periods of error "wave" $e_{(\alpha)}$ is composed, and FFT is applied. Since the amplitude of the synthesized signal is A = 1, (0 dB), the SFDR is defined by the level of the fifth harmonic at k = 81, $D = -L_5$ = 88.6dB.

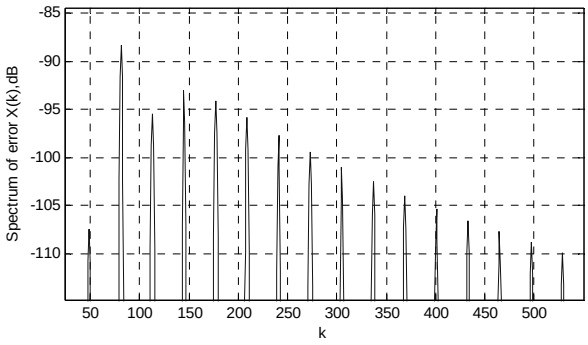


Fig.2. Error spectrum with 5th order polynomial.

B. Reducing the number of mathematical equations

As the first coefficient a_1 from (1) is very close to 1, it could be rounded. The other two coefficients a_3 and a_5 remain the same. Thus the number of mathematical operations is reduced to 2 multiplications and 2 addition/subtractions.

$$\sin(\alpha) = 1 - a_3\alpha^3 + a_5\alpha^5 \ . \qquad (3)$$

The calculation error $e_{(\alpha)}$ for polynomial (3) over the range $[-\pi/2, 3\pi/2]$ is shown in Fig.3.

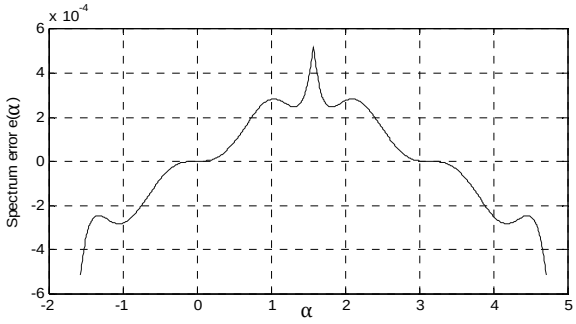


Fig.3 Sine wave error with 5th order polynomial

The spectrum of the obtained error $e_{(\alpha)}$ is depicted in Fig.4. The SFDR is defined by the level of the third harmonic at k = =49, $D = -L_3 = 88.7$dB.

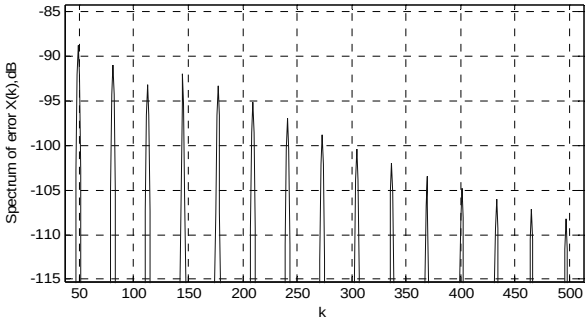


Fig.4. Error spectrum with 5th order polynomial.

Rounding the first coefficient a_1 doesn't affect the SFDR of the synthesized signal. Furthermore the number of mathematical operations is reduced without increasing the amplitude of the spectral spurs.

III. AN OPTIMIZATION ALGORITHM BASED ON STEEPEST ASCENT METHOD

A drawback of the ***polyfit*** function is that it is based on minimization of the root-mean-square error. In DDS it is important to minimize the spectral spurs. Therefore an algorithm aiming at the minimization of the spectral spurs based on steepest ascent/descent method is presented.

It searches for a global maximum over a two dimensional area by alternating two coefficient of a fifth order polynomial (3) with a step size alpha adjusted so that the function value is maximized along the direction by line search technique.

The goal is to find the best coefficients a3 and a5 at which the highest SFDR is achieved. The search area is shown in Fig.5. It is obtained by combining the two coefficients a3 and a5 and the resulting SFDR is given. The area is a ridge (Fig.6) with one global maximum which is the goal. To find the coefficient at which the maximum SFDR is achieved, an algorithm based on steepest ascent/descent method is presented.

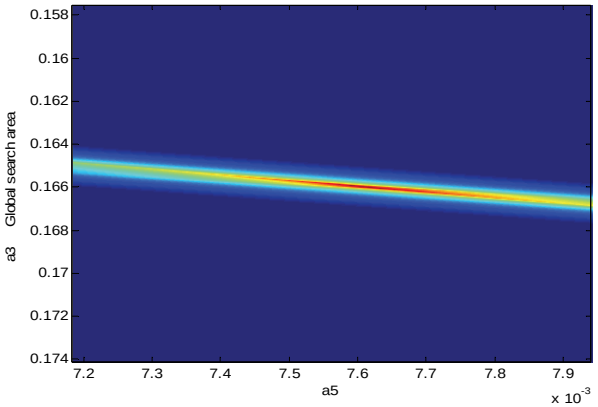


Fig.5. Search area of the steepest ascent method

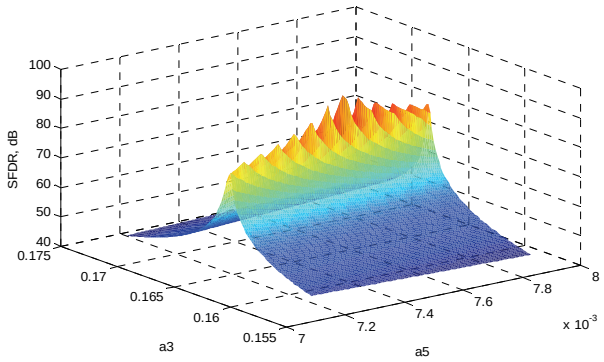


Fig.6. The search area of the steepest ascent method in 3D.

The algorithm is presented in Fig.7, where at the start two random coefficient a3, a5 and step sizes *alpha3* and *alpha5* respectively are defined and then the SFDR is calculated. For finding the ascending direction the two coefficient a3 and a5 are changed with the defined step sizes *alhpa3* and *alpha5* (addition). First the SFDR of the old coefficient a_3 and the new coefficient a_5 is calculated, then the SFDR of the new coefficient a_3 and the old coefficient a_5. After that the 3 obtained SFDRs are compared and the difference d3 and d5 is

calculated. The direction of ascending is defined by comparing the SFDR of the current cycle and the obtained SFDR of the cycle before – if the difference between the newly calculated SFRD and the old one (calculated in the cycle before) DD, is lower than the defined threshold, the direction for searching isn't changed. Else the two step sizes are changed and the algorithm returns to the beginning. The algorithms ends when the difference DD is bigger than the defined value of the threshold. Then it is assumed that the maximum SFRD is achieved.

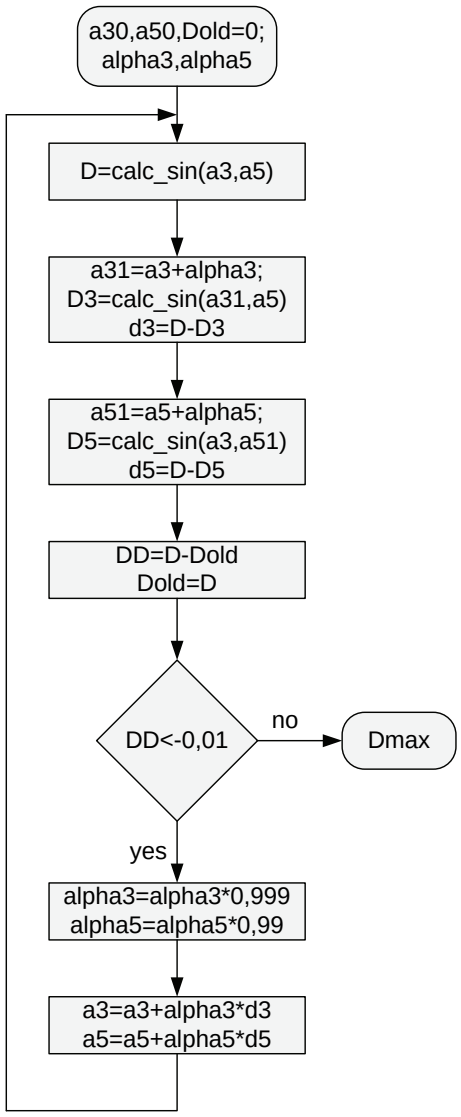


Fig.7 Block schema of the steepest ascent algorithm.

The path of steepest ascent algorithm is shown in Fig.7. The initial coefficients are a3=0.16564668, a5=0.00765671 and the calculated SFDR is D=68.1dB. The line search over the two dimentional area can be seen on fig.8. It can be seen that at the beginning of the search the start point in in the area with lower SFDRs. At the end of the algorithm, the last point defining the maximum SFDR is in the area with the highest values.

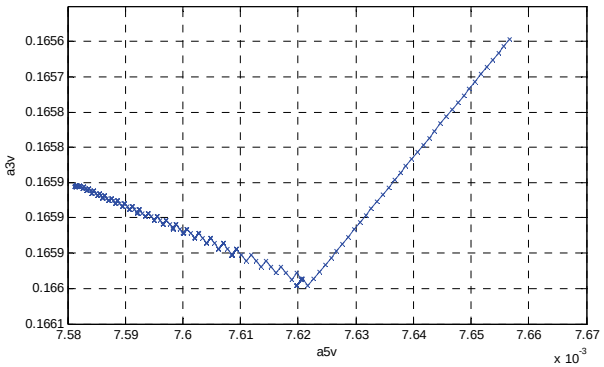


Fig.8. Path of steepest ascent algorithm.

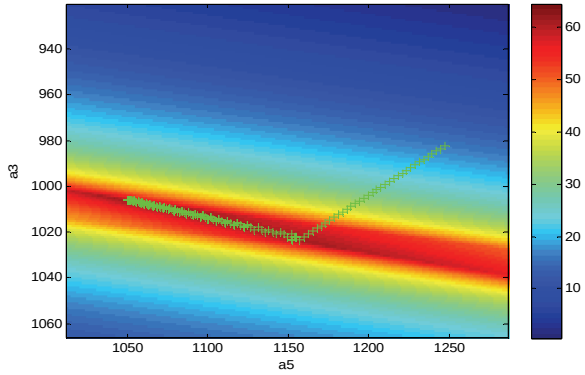


Fig.9. Path of steepest ascent over the search area.

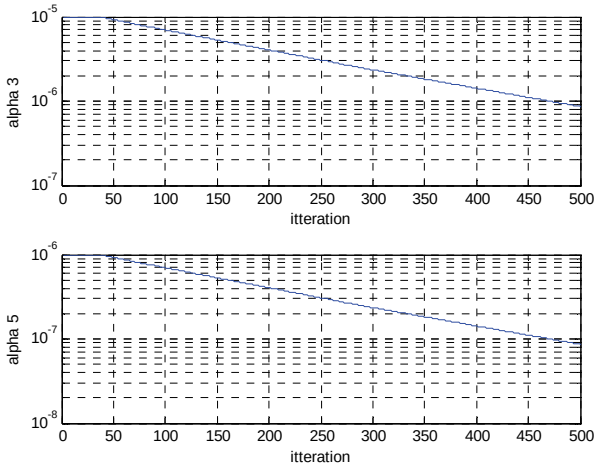


Fig.10 Change in the step sizes *alpha3* and *alpha5*.

Fig.10 represents the change in the step sized *alpha3* and *alpha5*. Because of the significants difference of the values of the two coefficient a3 and two different initial step sizes chould be assigned. In its initial state the steepest ascent technique is characterized by bigger step size and a constant direction. After approaching the area where the maximum is located, the step size begins to decrease; conversely the direction alternation is increased until the maximum is found.

The results using the inicial coefficient mentioned above over 500 iterations are – the new coefficients are

a_3=0.16585470, a_5= 0.00758128 and the resulting SFDR is D=91.3dB. Thus the SFDR of the polynomial ot 5th order (2) is increased with approximately 3dB.

IV. Comparison between Steepest Ascent Method and GAO

A comparison between the two algorithms is made.

The steepest ascent algorithms searches for the minimum of an N-dimensional function in the direction of the positive increment with a step size $alpha3_k$ and $alpha5_k$ at iteration k adjusted so that function value is maximized along the direction by a line search technique. At the current simulations the search is made over a complex plane of $4*10^6$ points for 500 iterations, generating 3 new sets of polynomial coefficient at each iterations.

GAO is a directed random search technique modeled on the natural evolution/selection process toward the survival of the fittest. The genetic operators deal with the individuals in a population over several generations to improve their fitness. The individuals are compared to chromosomes and are represented by a string of binary numbers. The step size remains the same. At the current simulations the search is made over a plane of 2^{48} points for 100 iterations with 90 individuals. One individual corresponds to a set of two polynomial coefficients a_3 and a_5.

Three sets of coefficients for each algorithm ensuring SFDR of 91dB are presented in Table I.

TABLE I
COMPARISON BETWEEN STEEPEST ASCENT AND GAO

Comparison	a3	a5	D[dB]
Steepest	0.16585281	0.00758103	91.2
Ascent	0.16585478	0.00758132	91.3
	0.16586387	0.00758484	91.4
GAO	0.16585968	0.00758353	91.4
	0.16585782	0.00758302	91.4
	0.16585784	0.00758302	91.4

The points are shown in Fig.11, where the SFDR obtained by the ***polyfit*** function is also presented with 'o'. The SFDR of the coefficients of steepest ascent algorithm is presented by '+' and the SFDR of the GAO coefficients is presented by '*'.

The marked points of the two algorithms are approximately close to one another. The difference in the SFDR is about 0.1dB. An exception is the SFDR obtained by the ***polyfit*** function where the SFDR differs with around 3 dB.

It can be seen that the steepest ascent is faster and also so accurate as the GAO. But there is a possibility if the algorithm comes across a local maximum to stop at it and not continue to search for the global one.

On the other hand the slower GAO can more accurately find the global maximum in a complex area with a lot of local extremes.

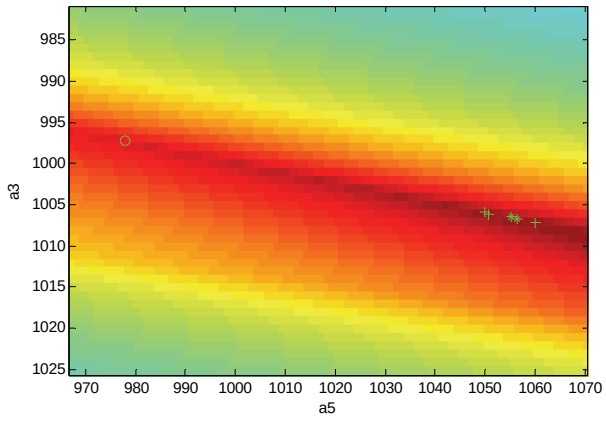


Fig.11. Positions of the resulted points on the search area

V. Conclusion

The optimization of a 5th order polynomial approximation of the sine wave was considered. Taking into account the sinus odd symmetry and rounding the first coefficient of the polynomial the number of mathematical operations necessary for the computation of the sine wave is reduced to 2 multiplications and 2 addition subtractions.

A comparison between two optimization algorithms is made. The advantage of the steepest descent algorithm is the lesser search time. GAO is slower but more accurate than the first one. Thus the results for the both algorithms are increasing the SFDR of the synthesized signal with about 3dB.

References

[1] B. Boyanov, "*Single table direct digital synthesis of signals, based on dsPIC digital signal controllers*", II International Congress on MEEMI 2005, Varna, Bulgaria, 2005.

[2] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs and mathematical tables. Applied Math. Series* 55. National Bureau of Standards, Washington, D.C., 1964.

[3] J.-A. Pineiro, S. Oberman, J.-M. Muller, and J. Bruguera, "*High-Speed Function Approximation Using a Minimax Quadratic Interpolator*", IEEE Transactions on Computers, vol. 54, No. 3, March 2005, pp. 304-318.

[4] N. Brisebarre, J.-M. Muller, and A. Tisserand. "*Computing machine-efficient polynomial approximations*" ACM Trans. Mathematical Software, 32(2): 236–256, June 2006.

[5] S. W. Smith, *"The scientist and engineering guide to Digital Signal processing"*,California Technical Publishing, 2002

[6] *Foundations of Genetic Algorithms 2*, Edited by L. Darrel Whittley, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[7] K. S. Tang, K. F. Man, S. Kwong and Q. He, "*Genetic algorithms and their Applications*", IEEE Signal Processing Magazine, vol. 13, No. 6, pp. 22-37, Nov. 1996.

[8] B. G. Boyanov, M. Iv. Shotova, "*Optimization of Polynomial Approximation in Series Based Software Direct Digital Synthesis of Signals*", ICEST 2010, Ohrid, Macedonia, pp191-194, June 2010.

Green's Function and Acoustic Standing Waves in Rectangular Loudspeaker Enclosures

Ekaterinoslav Sirakov [1], Hristo Zhivomirov[2], Boris Nikolov[3]

Abstract – **In this work are presented a theoretical analysis of acoustic standing wave and Green's function inside rectangular enclosures with rigid walls. The theory of room acoustics can be used for the analysis of sound field inside box of loudspeaker. Mathematical relations are presented for the calculation and researching of modal frequencies, standing sound waves and acoustics Green's function in a rectangular box. The results from the calculating and measuring modal frequencies and the box response are shown graphically and in a table.**

Keywords – **acoustic standing waves, Green's function, rectangular box of loudspeaker.**

I. Green's Function and Standing Wave

The Green's function for the sound field in a box with rigid walls can be defined with solutions to inhomogeneous Helmholtz differential equation [1, 2]:

$$\nabla^2 G(r,r_0) + k^2 \cdot G(r,r_0) = -\delta(r - r_0) \qquad (1)$$

where: G is Green's function [m^{-1}],

$k = \frac{\varpi}{c_0}$ - wave number [m^{-1}], c_0 - the speed of sound [m/s],

r -receiver position [m],

r_0 -source position [m],

δ - Dirac delta function,

∇^2 = div grad - Laplacian.

Using the method of images, the walls of the box can be replaced by the image sources obtained by reflecting the source point (and all its images) towards the walls.

The Green's function is the sound pressure at one point, $r(x,y,z)$, generated by a (normalized) point source at another point, $r_0(x_0,y_0,z_0)$ [3, 4]:

$$G(k,r,r_0) = \frac{-1}{V}\sum_N \frac{\Lambda_N \cdot \Psi_N(r) \cdot \Psi_N(r_0)}{\left(k^2 - k_N^2\right)} \qquad (2)$$

where: Ψ_N - mode function [dimensionless].

The Ψ_N - functions are the corresponding functions:

$$\Psi_N(x,y,z) = \cos(k_x \cdot x) \cdot \cos(k_y \cdot y) \cdot \cos(k_z \cdot z) \qquad (3)$$

N - integer, represents the three integers n_x, n_y and n_z (all values are integers between 0 and ∞).

$$\sum_N = \sum_{n_x=0}^{\infty} \sum_{n_y=0}^{\infty} \sum_{n_z=0}^{\infty}$$

V - volume [m^3],

$V = l_x \cdot l_y \cdot l_z$, l_x, l_y, l_z - dimensions of rectangular box [m],

x, y, z - Cartesian coordinates [m],

The modal amplitude, Λ_N (normalizing factor), depending on the modal numbers, given by [1, 4]:

for oblique waves (three dimensional modes):

$\Lambda_N = 8$ for $n_x > 0$, $n_y > 0$ and $n_z > 0$,

for tangential waves (two dimensional modes):

$\Lambda_N = 4$ when one of n_x, n_y or n_z is zero,

for axial waves (one dimensional modes):

$\Lambda_N = 2$ when two of n_x, n_y and/or n_z are zeros.

Fig. 1 and Fig. 2 shows the magnitude of the Green's function in a box as a function of the frequency and the receiver's position.

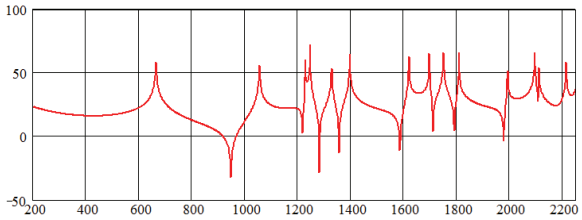


Fig. 1. Magnitude of the Green's function in a box as a function of the frequency.

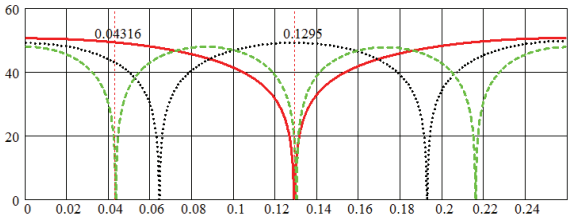


Fig. 2. Magnitude of the Green's function in a box as a function of the receiver position in x axis for natural frequency f_1=664 *Hz* - solid line, f_5=2 .f_1= *1328 Hz* - dot line and f_{11}=3.f_1= *1992 Hz* – dash line (l_x = 0.259, m).

[1]Ekaterinoslav Sirakov is with the Department of Radio Engineering, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria
E-mail: katiosirakov@abv.bg

[2]Hristo Zhivomirov is a student with the Department of Radio Engineering, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria
E-mail: hristo_car@abv.bg

[3]Boris Nikolov is with the Department of Radio Engineering, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria
E-mail: boris84@abv.bg

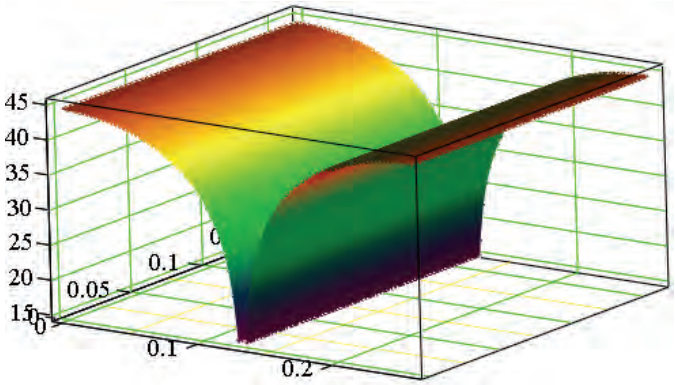


Fig. 3. The distribution of magnitude of sound pressure in a rectangular box in *x* axial sound wave, f=664 Hz, of mode (1, 0, 0).

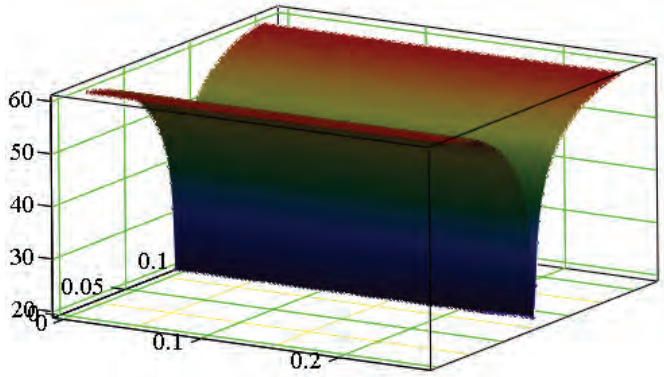


Fig. 4. The distribution of magnitude of sound pressure in a rectangular box in *y* axial sound wave, *f=1055 Hz*, of mode (0, 1, 0).

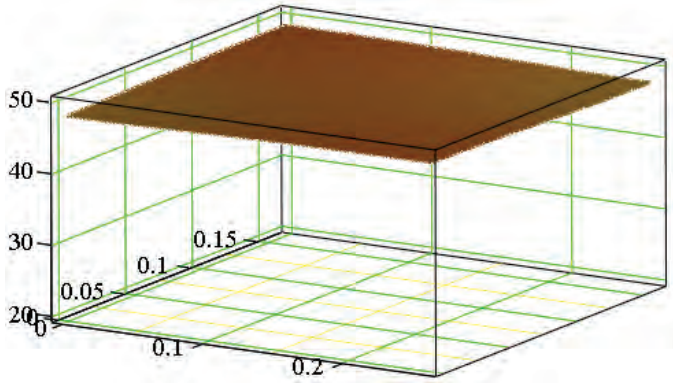


Fig. 5. The distribution of magnitude of sound pressure in a rectangular box in z axial sound wave, *f*=1229 *Hz*, of mode (0, 0, 1).

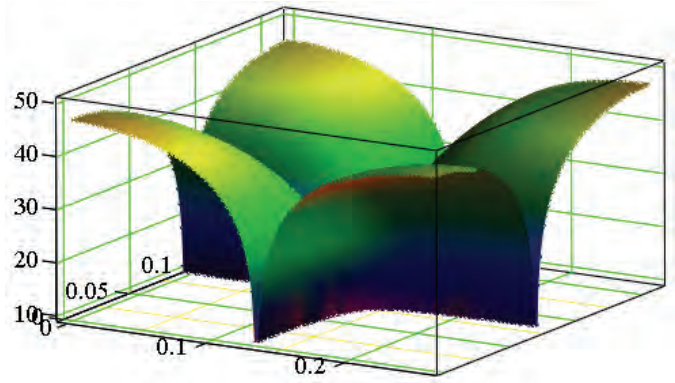


Fig. 6. The distribution of magnitude of sound pressure in a rectangular box in x, y tangential sound wave, f=1247 Hz, of mode (1, 1, 0).

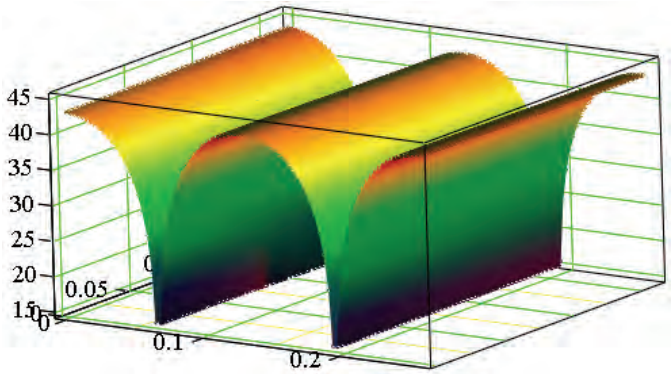


Fig. 7. The distribution of magnitude of sound pressure in a rectangular box in *x* axial sound wave, f=1328 Hz, of mode (2, 0, 0).

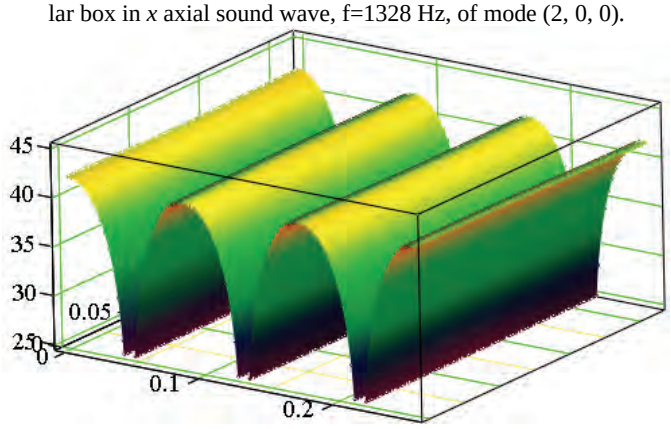


Fig. 8. The distribution of magnitude of sound pressure in a rectangular box in *x* axial sound wave, f=1992 Hz, of mode (3, 0, 0).

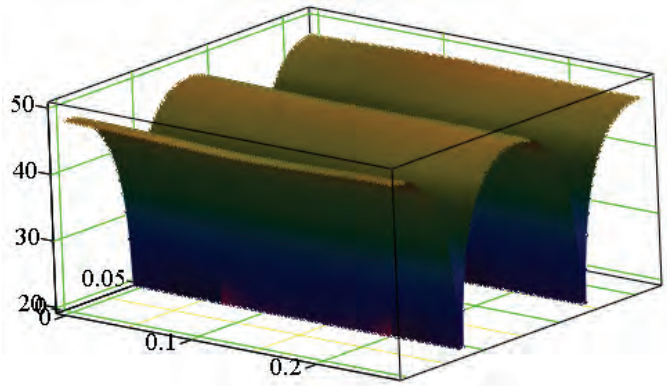


Fig. 9. The distribution of magnitude of sound pressure in a rectangular box in *y* axial sound wave, *f=2110 Hz*, of mode (0, 2, 0).

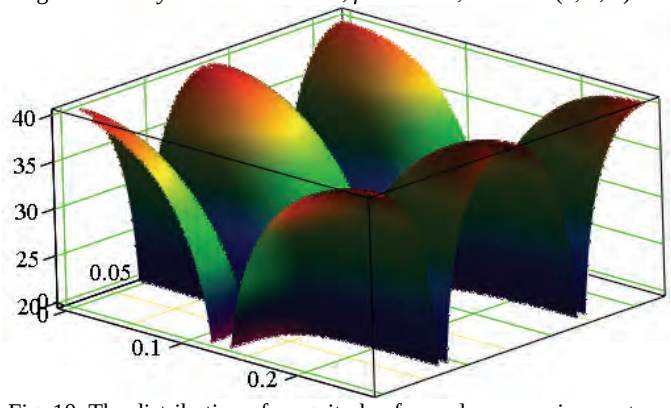


Fig. 10. The distribution of magnitude of sound pressure in a rectangular box in x,y tangential sound wave, *f=2212 Hz*, of mode (1, 2, 0)

The characteristic frequencies f_N of these standing waves are given by:

$$k_N^2 = \left(\frac{2 \cdot \pi \cdot f_N}{c_0} \right)^2 = k_x^{\,2} + k_y^{\,2} + k_z^{\,2} \quad (4)$$

The constant k_N can be represented with its up on x, y and z

$$k_x = \frac{n_x \cdot \pi}{l_x}, \qquad k_y = \frac{n_y \cdot \pi}{l_y}, \qquad k_z = \frac{n_z \cdot \pi}{l_z}$$

The modal frequency of a rectangular box is given by Eq. (4) which can be rewritten as:

$$f_N = \sqrt{f_x^2 + f_y^2 + f_z^2}$$

$$f_N = \frac{c_0}{2} \cdot \sqrt{\left(\frac{n_x}{l_x} \right)^2 + \left(\frac{n_y}{l_y} \right)^2 + \left(\frac{n_z}{l_z} \right)^2} \quad (5)$$

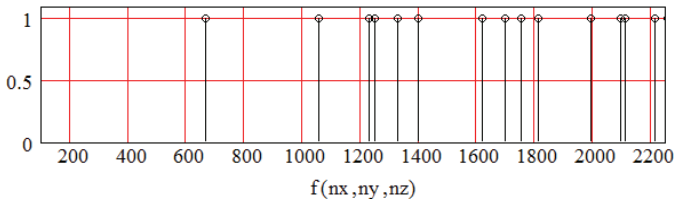


Fig. 11. Plots of mode distribution [5].

TABLE I
THE FOURTEEN LOWEST NORMAL MODES AND THEIR NATURAL FREQUENCIES FOR A BOX WITH RIGID WALLS AT THE SPEED OF SOUND 344 m/s [5].

№	mode	n_x, n_y, n_z	Frequency, Hz
1	x axial	1, 0, 0	664
2	y axial	0, 1, 0	1055
3	z axial	0, 0, 1	1229
4	x, y tangential	1, 1, 0	1247
5	x axial	2, 0, 0	1328
6	x, z tangential	1, 0, 1	1397
7	y, z tangential	0, 1, 1	1620
8	x, y tangential	2, 1, 0	1696
9	x, y, z oblique	1, 1, 1	1750
10	x, z tangential	2, 0, 1	1809
11	x axial	3, 0, 0	1992
12	x, y, z oblique	2, 1, 1	2095
13	y axial	0, 2, 0	2110
14	x, y tangential	1, 2, 0	2212

A graphical representation of the theoretical sound pressure distribution of the *x* axial mode 1, 0, 0 in a model rectangular box was show in Fig. 3.

The sound pressure is zero in the vertical plane at the center of the box and maximum at its ends. The distribution of the sound pressure in the *xy* plane at *y* and *z* axial modes, respectively, is shown in Fig. 4 and Fig. 5.

Three-dimensional representations of the sound pressure distribution in a rectangular box for a tangential mode are represented in Fig. 6 (1, 1, 0) and Fig. 10 (1, 2, 0).

From Figs. 3 to 10 can be noted that sound pressure is maximum at the corners for all modes.

II. ENCLOSURE RESPONSE MEASUREMENTS IN MODEL BOX

The characteristics of the sound pressure in a rectangular loudspeaker enclosure, measured with Realtime Analyzer [6] application software, are presented in graphical form in Fig.12.

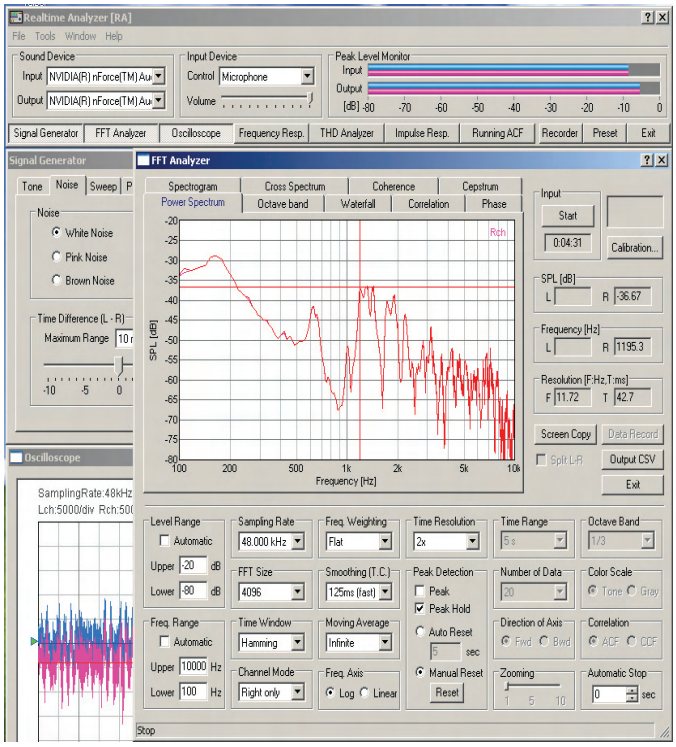


Fig. 12. The magnitude measured sound pressure [6] in the rectangular loudspeaker enclosure (dimensions: width 18 cm, height 27 cm, depth 15.4 cm and wall thickness of 0.85 cm) with a loudspeaker type BK 138 A4 TM (mechanical resonance frequency 90 Hz ± 20Hz, and bandwidth to 15 kHz)

A PC with an adequate sound card is necessary for the measurement. Measuring microphone must be small in size (e.g. electret microphone) and must be placed where extreme values of sound pressure for the characteristic frequencies are expected. For example, if the microphone is placed in a corner of the box, a peak sound pressure will be expected (Figs. 3, 4, ..., 10) due to the box's specific frequencies.

Measurements were made with broadband loudspeaker type BK 138 A4 TM with mechanical resonance frequency of 90

Hz ± 20Hz, and bandwidth of 15 kHz [7], mounted in a rectangular box with dimensions: width 18 cm, height 27 cm, depth 15.4 cm and wall thickness of 0.85 cm.

The program allows the data from the measured values of sound pressure in dB to be stored in tabular and text format for further analysis. The natural frequency of the rectangular box, calculated in accordance with mathematical dependence (5) is presented in Fig. 11.

To examine the modal structure of the bare enclosure, box response at a corner was measured.

Frequency response of sound pressure, measured in the frequency domain from 20 Hz to 100 Hz corresponds to a closed box loudspeaker system (high pass filter).

Local maximum for frequencies around 664 Hz corresponding to the first order own frequency (Table I, № 1 and Fig. 2) in *x* axial wave (1, 0, 0). The second natural frequency 1.043 kHz (Table I, № 2) in *y* axial wave (0, 1, 0) raised slightly the slope of this characteristic.

The peak frequencies of the response curve agree very well with the modal frequencies (Table I) from normal mode theory.

The few modes missing in the measured response are mostly the degenerate modes.

III. Conclusion

Reflected or standing waves may be present in the box of a loudspeaker due to internal reflections [8].

The method – Green's function theory used in room acoustics can be applied to the analysis of a loudspeaker box.

The comparison of experimental results (Fig. 12) with a specific theoretical sound box shows the influence of resonant frequencies on their own characteristics and the possibility to use the established theory of rectangular rooms.

The results obtained in this work can be used for theoretical analysis, design and production of boxes for loudspeakers.

Acknowledgement

This paper was developed in the frames of a scientific research project "Multi-dimensional Computer Analysis in Communications", sponsored by Ministry of Education, Youth and Science, Bulgaria, 2011.

References

[1] Waig T. Chu, ASHRAE Research Project 624-RP, "Laboratory Methods of Testing for Rating Low Frequency Sound Emission of Air Conditioning Equipment", 1993.

[2] Finn Jacobsen, *The sound field in a reverberation room*, Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Note 31261, 2010

[3] Philip M. Morse, *Vibration and Sound*, New York, McGraw-Hill, 2nd. ed., 1948.

[4] Yu Luan, Finn Jacobsen, "A method of measuring the Green's function in an enclosure", Journal of the Acoustical Society of America, vol. 123, issue. 6, pages: 4044-4046, 2008

[5] Екатеринослав С. Сираков, "Собствени резонансни честоти на правоъгълно озвучително тяло", Национална конференция с международно участие „Акустика 1", Варна, Технически Университет-Варна, Катедра „Радиотехника", 2009.

[6] http://www.ymec.com/products/dssf3e/

[7] http://www.vissokogovoriteli.plc.bg/

[8] Glen Ballou, *Electroacoustic Devices: Microphones and Loudspeakers*, Focal Press, 2009

One Approach for Increasing the Efficiency of Algorithms for Metadata Extraction from Static Images

Gergana Markova[1], Oleg Asenov[2], Margarita Todorova[3]

Abstract - **The current work presents an approach for increasing the efficiency of algorithms for extraction of data from static images aiming at increasing the volume and content of the generated metadata and providing the opportunity for searching by content in large libraries of static images.**

Keywords-**Efficiency, processing, static images, algorithms.**

I. INTRODUCTION

Extraction of metadata from static images with different parameters and properties is an important information process which gives the opportunity for quick theme search in large libraries of images. The parameters and properties of the image are a function of the source for input. The difference between the image parameters and algorithm requirements for processing leads to a limited volume and content of the generated metadata.

Automated search in multiple static images imposes the application of algorithms for metadata extraction and carrying out the search itself in the multitude of the generated metadata. Practice imposes processing of static images represented in different formats, resolution and sizes [1], [2], [6].

The efficiency of a certain algorithm for metadata processing and extraction from a static image is defined by a number of quality characteristics of the representation of the video information generating significant metadata [2], [3].

Let's assume we can always "deliver" for a selected algorithm for image processing and metadata extraction the photograph with the necessary resolution and quality. In this case the generating of the metadata will be reliable because the necessary quality of the input information has been supplied. But creation of digital images is not a determined process [4], [5]. A clear discrepancy is seen between the specific requirements regarding the input information, ensuring the efficiency of a certain algorithm for format processing and the properties of the real input information which is to be processed. This discrepancy leads to the necessity of developing an intermediate functional layer that will ensure the concordance between the primary formats and the "successful" formats, specific to each processing algorithm. In this way the efficiency of the algorithm itself is ensured on a level of restrictions regarding the processed image and also on the concordance stage the images for which it is impossible to provide the minimum input restrictions for a certain algorithm are rejected.

[1]Gergana Markova, Faculty of Mathematics and informatics, arh.G. Kozarov str 3, bild 3, 5000Veliko Tarnovo, Bulgaria, E-mail: gerganavioletova@abv.bg.

[2]Oleg Asenov,Faculty of Mathematics and informatics, arh.G. Kozarov str 3, bild.N3, 5000 VelikoTarnovo,Bulgaria, E-mail: olegasenov@abv.bg

[3]MargaritaTodorova, Faculty of Mathematics and informatics, arh.G. Kozarovstr 3, bild.N3, 5000 VelikoTarnovo, Bulgaria, E-mail: marga_get@abv.bg

II. LAYOUT

The current work represents an approach for increasing the efficiency of algorithms for metadata extraction from static images through the following sequence of stages:

Stage 1. Let the set of input parameters, through which an image is characterized for the *i*-th algorithm is called SIRA*i* (Set Image Recognition for Algorithm*i*). An analysis is carried out for the classification of the static image processing algorithms in terms of the set of input parameters for each of the included in the system *i*-th algorithm for the *j*-th image through SIRA*i*.

Stage 2. A processing is carried out and classification of the incoming stream of input static images in terms of SIRA*i* and the profile of the *j*-th input image is formed, which we define as PIP*j*.

Stage 3. A comparative analysis is performed for the efficiency of the forthcoming processing of the *j*-th image from the *n* processing algorithmsavailable in the system. Thus, the *j*-th image is analysed in terms of the possibility for processing of each *i*-th algorithm *(i=1, n).*

Stage 4. On basis of the analysis the type of transformational function F*p* (SIRA*i*, PIP*j*) is formed and, through it, derivative images of the basic image *j* are generated and we define its profile with PIP*ij*, meeting the requirements of the input information SIRA*i* for the *i*-th algorithm.

Stage 5. Processing of the PIP*ij* images from the corresponding *i*-th algorithm and extraction of metadata PMD*ij* for the derivative image PIP*ij*.

Stage 6. Recording the metadata in a database for the context of the image *j*, processed with the set of algorithms *i*, (*i*=1,*n*).

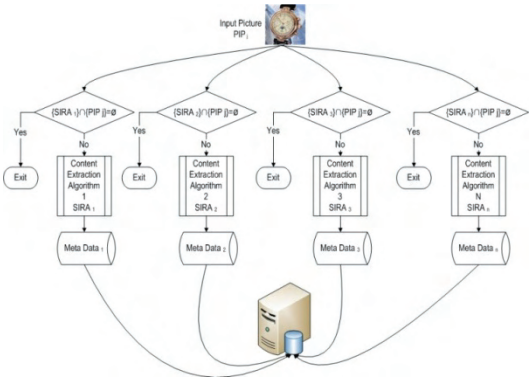


Fig.1. Architecture of a system for metadata extraction without the application of the approach for increasing the efficiency.

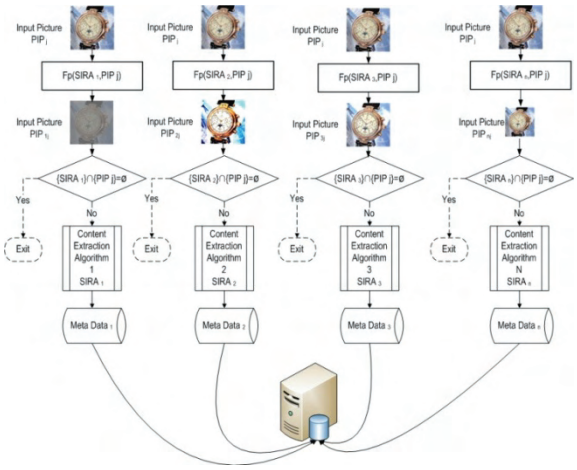


Fig.2. Architecture of a system for metadata extraction with the application of the approach for increasing the efficiency.

In Fig.1 and Fig.2 are the graphic representation of the necessity for increasing the efficiency of the algorithms for metadata extraction from static images. Fig.1 represents the architecture of a system for metadata extraction without the application of the approach for increasing the efficiency.

The check for compatibility of the *j*-th input static image in terms of the calls of the input information for the *i*-th algorithm. SIRA*i*∩PIP*j* is finding the intersection of the two sets of parameters. If this section is an empty set {ø}, the processing is discontinued and metadata extraction is not possible through the application of the *i*-th algorithm. The more different the two sets are, the less the probability for extraction of reliable metadata is possible. Fig.2 represents an improved version of the structural scheme for metadata extraction from static images. Through the function for preliminary preparation of the image F*p*(SIRA*i*,PIP*j*), the primary image with parameters PIP*j* is transformed into an image, which meets the requirement parameters of the processing algorithm. The result of the transformation application F*p*(SIRA*i*,PIP*j*) is the derivative image PIP*ij*, which is much closer, in terms of parameters, to the initially known input restrictions of the *i*-th algorithm. Thus the probability for the check to result into {ø} is minimal and extraction of metadata by the *i*-th algorithm is expected.

An example of practical realization.

Let's assume that algorithm Ak is an algorithm for recognition and reading of automobile license plates. For the current example we use an algorithm of Adaptive Recognition Hungary - CARMEN® Freeflow Number Plate Recognition Engine [5]. We apply the algorithm to the input image in Fig.3. The developers of the algorithm have defined SIRAк in the documentation in the following way: "The CARMEN® Freeflow offers general-purpose license plate reading from digital pictures of any tpye of sources. Whether it is a megapixel-size, 12bits/pixel high dynamic range, high resolution digital image or a 384x288 size colour CCTV video picture provided as input, (either from file or memory) CARMEN® provides ballanced reading of car plates." [5]. It is evident from Fig.3 that the image is 24 bits/pixel high dynamic range and a resolution of 3072x2340. The result: the license plate of the automobile is not recognized and read due to the discrepancy of the profile of the j-th input image PIPj in terms of SIRAк. If we apply, in accordance with Fig.2, an architecture for metadata extraction with the suggested approach for increasing the efficiency of the image PIPj (Fig.3), we form the derivative j-th image, suitable for processing with the algorithm PIPiк (Fig.4). For PIPiк we only change the resolution parameter from 3072x2340 to 460x365. The result is a 100% recognized license plate. Thus, without changing the content of the image, and only through a change of micro-parameters, such as resolution and number of pixels in representing the colour, a higher efficiency of metadata extraction is ensured.

The example given shows the necessity of looking into the problems of compatibility of the input graphic information with the algorithms for metadata extraction and the development of approaches for increasing the efficiency through specific micro-processing.

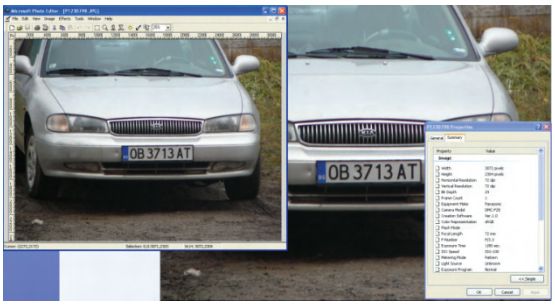


Fig.3. Input image – PIPj

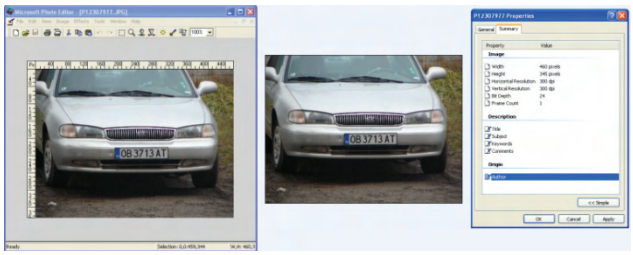


Fig.4. Derivative image PIPjk

III. Conclusion

The application of this approach for increasing the efficiency of the algorithms for metadata extraction from static images will ensure a more complete usage of the possibilities for generating content and classification of the static images through metadata, as well as performing quick multi-parametric searches in the classified array of images.

The presented approach is open, it allows the introduction of automation and dynamic complement of the set of algorithms and expanding the volume and content of the static image extracted metadata. Consequently the future work will be further development of the transformational function F*p*

and achievement of higher efficiency of the metadata extraction algorithms.

An experiment with the practical realization will be provided in the future in order to prove the application of the algorithm in different areas. Also comparison with similar approaches will be realized.

REFERENCES

[1] A.Bovik,Handbook of Image and Video processing, Elsevier Academic Press, 2005.

[2] J. Wolf, Metadata image processing, 2008.

[3] W. Pratt, Digital Image Processing: PIKS Inside, Third Edition, New York, 2001.

[4] Z. Ma, Intelligent databases: Technologies and Application, Northeastem University, China, 2007

[5] http://www.anpr.net/anpr_09/anpr_freeflow.html

[6] http://www.aurigma.com/docs/gm/JPEGFileFormat.htm

This paper is financed by project: Creative Development Support of Doctoral Students, Post-Doctoral and Young Researches in the Field of Computer Science, BG 051PO001-3.3.04/13, EUROPEAN SOCIAL FUND 2007–2013г. OPERATIONAL PROGRAMME “HUMAN RESOURCES DEVELOPMENT”

Driver Distraction and In-vechile Information System

Marjana Čubranić-Dobrodolac [1], Svetlana Čičević [2], Momčilo Dobrodolac [3], Andreja Samčović[4]

Abstract - **As computers and other information technology move into cars and trucks, distraction-related crashes are likely to become an important problem. Distraction is a well-established causal factor in road accidents. The range of system often termed In-Vechile Information Systems (IVIS) are the focus of this paper.**

Keywords - **distractions, drivers, IVIS, risk**

I. INTRODUCTION

Over the past century, automation has gradually moven its way into most aspects of our daily lives. Presently we use automation in many contexts, both passively (by using products/services provided using automation), and actively (by interacting with automation directly). The majority of researchers and authors prefer to refer to the whole range of driving automation technology as intelligent transport systems (ITS). ITS can be broadly regarded as falling into two distinct categories: advanced driver assistance systems (ADAS) and in-vehicle information systems (IVIS). IVIS include route guidance systems, traffic information systems, vehicle monitoring systems, audio/video devices, vehicle communication systems and driver convenience services (e.g. personal digital assistants – PDA`s, phone related services, hands-free equipment, driver identification systems). The ONS omnibus survey examined UK drivers' and passengers' attitudes towards transport [1]. In one section of this survey, they asked whether the car/van that participants used most often had a satellite IVNS installed. They also collected a range of demographic information (including age, gender, socio-economic group, gross annual income and driving frequency). The survey showed that an equal proportion of male and female drivers reported using an IVNS (7%) and that they were used by drivers of all ages, although the highest using age bands were 26-44 years (9%) and 45-54 years (9%). The GFK survey showed that in Germany, the highest purchasing age bands were 40-60 years (43%), closely followed by those under the age of 30 years (32%) and over 60 years (25%).

[1]Marjana Čubranić-Dobrodolac is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Beograd, Serbia, E-mail: marjana@sf.bg.ac.rs.;

[2]Svetlana Čičević is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Beograd, Serbia, E-mail: s.cicevic@sf.bg.ac.rs.;

[3]Momčilo Dobrodolac is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Beograd, Serbia, E-mail: m.dobrodolac@sf.bg.ac.rs

[4] Andreja Samčović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Beograd, Serbia, E-mail: andrej@sf.bg.ac.rs

II. LITERATURE REVIEW

Distraction of the driver is one of the major causes for car accidents. About 20 % of injury crashes in 2009. involved reports of distracted driving [2]. In 2009, 5,474 people were killed in U.S. roadways and an estimated additional 448,000 were injured in motor vehicle crashes that were reported to have involved distracted driving [3]. Wang, Knipling, and Goodman [4] have analyzed 1995 CDS data to compare distraction-related crashes to other crashes by crash type. They raport that distraction - related crashes account for about 13 percent of crashes nationally. Their analyses by crash type and distraction showed that distraction-related single vehicle crashes (both run–off-the-road and on-road) account for about 18.1 percent of single vehicle crashes and 41.2 percent of all distraction-related crashes. Thus, the single vehicle run off the road crash The age group with the greatest proportion of distracted drivers was the under 20 age 16 percent of all drivers younger than 20 involved in fatal crashes were reported to have been distracted while driving [2].

A. Types of distraction

To understand distractions, some knowledge of theories of human attention is warranted. According to the Multiple Resource Theory [5], people are considered to have a variety of resources (visual, auditory, cognitive, and biomechanical) they can allocate to a task or combination of tasks. Overload can occur when the task demand exceeds at least one of the resources or, in less common cases, the capability to switch between tasks. So for example, people cannot read two high data rate, no redundant streams of text separated by a large visual angle because their eyes cannot be directed towards two widely separated locations at once. Similarly, people cannot retain nonchunkable long strings of digits in memory because that would overload the cognitive resource, specifically, short-term memory.

Four inter-related subcategories have been identified: visual distraction; auditory distraction; biomechanical distraction and cognitive distraction [6]. Auditory distraction – occurs when the driver momentarily or continually focuses their attention on sounds or auditory signals rather than on the road environment. Biomechanical distraction - this involves movements of the driver's body away from the standard posture required to perform the physical tasks associated with safe driving. Cognitive distraction – includes any though the road network safely and their reaction time is reduced. Visual distraction – comes in tree forms: driver's visual field is blocked when a driver focuses visual attention on something other than the road ahead (navigation system, a loss of visual attentiveness, also known as "looked but did not see").

The experimental approach on studying driver distraction has been an area of interest in human factors research since the 1980s. Driving simulation studies have been frequently used in order to avoid real crash risk [7]. A popular paradigm in this line of research has been based on the measurement of driver workload and driving performance at the level of operational control of the vehicle [8]. The basic problems with interpreting the results of these experiments often reside in the not-self paced and time-pressured tasks, and subsequently in the absence of participants' possibilities to prioritize the driving over secondary tasks. The external validity of the conclusions can often be questioned [9]. These studies are valuable for revealing capacity limitations of the drivers in a dual-task situation. However, they do not necessarily tell us if the drivers are able to overcome their capacity limits with tactical behaviors in real traffic to maintain a sufficient level of driving performance. Recently, new perspectives and models for studying driver distraction on multiple levels have been proposed [10]. Lee, et al. [8] introduced the model of driver distraction comprising of breakdowns at the operational, tactical and strategic levels of control in dual-tasking while driving based on Michon's [11] three-level model of driving behavior. This model induces new types of challenges for experimental research; how can breakdown in control be measured on the levels of tactical and strategic control? These are not necessarily in direct relation to task workload or to the lapses of vehicle control at the level of operational control.

The effects of visual time sharing on driving performance have been extensively studied and are fairly well understood. When visual attention is diverted from the road (by a secondary task or by visual occlusion), the driver cannot give any tracking response, which results in periods with fixed steering wheel angle [12]. During these periods, heading errors build up which result in lane weaving and, sometimes, lane exits. Many studies have found a strong relationship between visual demand and reduced lane keeping [13]. There is considerable evidence that the driver's lane keeping performance is guided by time-based safety margins, representing the "safe boundaries" that the driver aims to stay within [14], an idea first proposed by Gibson and Crooks [15]. Physiological effects of IVIS performance have primarily been studied in the context of mental workload and stress, where usually no distinction is made between visual and cognitive load. Especially cardiac activity, measured in heart rate and heart rate variability, has been proved to be sensitive to mental workload and stress. Another common physiological workload indicator is skin conductance, which has been proved sensitive to task complexity. Today, with the advent of hands-free and voice-based solutions for the phone and other functions, the number of mainly cognitively loading in-vehicle tasks (with no visual diversion from the road ahead) is increasing. This has generated much interest in the effects of cognitive load on driving. Numerous studies have reported degrading effects of purely cognitively loading tasks in terms of reduced event detection performance. These include results from artificial detection tasks such as the Peripheral Detection Task [16], as well as more realistic tasks such as detection of critical events in simulation [13]. Recarte and Nunes [17] investigated the effect of cognitively loading tasks on visual behavior and event detection performance. They found reduced event detection as well as a concentration of gaze towards the road centre during certain cognitively loading tasks, such as word production and complex conversation. Similar results where obtained by Harbluk and Noy [18]. They found that the number of saccades decreased and that the percentage of time spent looking at the central region of the road increased with task complexity. If a driver's attention is drawn away from the primary task of driving, or they are overloaded beyond their capabilities, crash risk is elevated.

B. Factors of distraction

Advances in computer and communication technology over the last two decades have led to the development of a wide array of advanced in-vehicle information systems, collectively called telematics. The proliferation of in-vehicle technology has generated concern that these systems, singly and in combination, might cause an increase in driver distraction [19]. One of the most widely available in-vehicle advanced technologies is the route guidance system. These systems provide the driver with information about a route to a destination supplied by the driver. Because these systems use vehicle location technology, such as GPS, route directions can be timed to correspond with the driver's information needs as he or she drives. There is little information about the incidence of route-guidance systems in vehicles or the frequency with which they are used. Analysis of the crash databases yielded no instances in which use of a route-guidance system was indicated as a contributing factor in distraction-related crashes [20]. In addition, natural use studies of various route guidance systems have found no adverse effect on traffic safety, nor any increase in self-reported distraction [2, 4, 9, 19]. Despite these results, there is general agreement in the literature that the function of destination-entry is quite distracting if it involves visual displays and manual controls [13]. While most destination-entry would probably occur in a stationary vehicle, Green [3] has pointed out that there are several scenarios in which a driver might engage in destination-entry while driving, and in turn be at greater risk for a distraction- related crash: driver is in a hurry and enters the destination after starting the trip; driver changes his or her mind about the destination after starting trip; driver gets other information, such as a radio traffic report, then decides to change the route; driver entered the wrong destination; and the driver does not know the exact destination prior to departure and enters the actual destination later. Thus, there are several scenarios in which use of a route-guidance system could lead to distraction-related crashes.

Evidence obtained from simulated driving [21] and on-the-road driving [23, 24] has shown that use of a mobile phone can lead to decrements in tasks required for safe driving. There is general agreement in the literature that the most distracting activities involving cellular phone use are dialing and receiving phone calls [24, 3]. In addition, use of hand-held phones tend to be associated with greater decrements in driving performance than hands-free phones, but the

conversations tend to be equally distracting, especially when the information content is high [25]. Evidence is also mounting, although still far from conclusive, that the use of cellular phones increases crash risk. In their analysis of the CDS data, Stutts, Reinfurt, and Rodgman [26] found that cellular phone use or dialing was implicated in about 1.5 percent of distraction-related crashes. One would expect this percentage to increase as the predicted use of cellular phones increases. More recent work in Virginia has found that about 5 percent of distraction-related crashes involve cellular phones [5]. Using a cell phone use while driving, whether it's hand-held or hands-free, delays a driver's reactions as much as having a blood alcohol concentration at the legal limit of .08 percent [26]. Reed and Robbins [27] conducted a simulator study to investigate the impact of text messaging while driving. Results show that participants' driving behavior was impaired, particularly reaction time (35% slower when writing a text message) and the ability to maintain lateral vehicle control.

Lee et al. [9] investigated driving performance while operating a speech-based e-mail system. Results show a 30% increase in reaction time to a braking lead vehicle when the speech-based task was carried out. Furthermore, subjective workload increased significantly while performing the e-mail task compared to a baseline condition. A follow-up study of Jameson et al. [28] confirmed these results and demonstrated a significant reduction of the time to collision in the distraction condition. The vast majority of motor vehicles are equipped with entertainment systems that include radios, cassette players, and/or compact-disc (CD) players. Operation of these systems usually involves manual manipulation of buttons, knobs, and media, as well as visual input, leading to a potential for physical, cognitive, and visual distraction. Analyses by several researchers have shown that adjusting an entertainment system is one of the leading in-vehicle triggering events for distraction-related tow-away crashes [26, 4] distraction-related police-reported crashes [29], and distraction-related fatal crashes [30].

III. CONCLUSION

The potential for novel IVIS tasks to dangerously distract is a significant safety concern. It can be difficult to legislate against driver distraction, in contrast to other impairments, such as when driving under the influence of alcohol. Distraction, unlike alcohol or fatigue, is likely to affect drivers only intermittently over the cours of journey [31]. Furthermore, drivers may even choose to engage in distracting task during periods where overall accident ris is low, for example when on quite, straight roads, or when stationary, such as when waiting at traffic lights. As driving processes become more automated, IVIS may even be important in avoiding potential driver under-load. Consequently, banning drivers from engaging in all IVIS task whilst a vechile is in motion is in motion is neither realistic nor practical.

A future research aim for driver distraction in general might migh be to combine accident analysis studies with evidence from driving performance evaluations, in order to establish the absolute risk posed by IVIS tasks. Risk is a factor of demand of performing a task, the prevalence of the system amongst the driver population, the frequency of use, and the driving environment at the instant of interaction. Consequently, determining the overall risk of a distraction source must consider exposure to the source, in addition to its distracting effects.

ACKNOWLEDGEMENT

This paper is supported by Serbian Ministry of Science and Technological Development with Project 36022.

REFERENCES

[1] Svahn, F. 2004. "In-Car Navigation Usage: An End-User Survey on Existing Systems," in: *Proceedings of the 27th Information Systems Research Seminar in Scandinavia*. Falkenberg, Sweden.

[2] NHTSA (2000). Driver distraction expert working group meetings: Summary and proceedings. NHTSA, Washington DC.

[3] Green, P. (2000). Crashes induced by driver information systems and what can be done to reduce them (SAE paper 2000-01-C008). In Proceedings of Convergence 2000 Conference (SAE publication P-360). (Warrendale, PA: Society of Automotive engineers).

[4] Wang, J.-S., Knipling, R.R., & Goodman, M.J. (1 996). The role of driver inattention in crashes: New statistics from the 1995 Crashworthiness Data System. In 4oth Annual Proceedings Association for the Advancement of Automotive Medicine. (pp 377-392). AAAM: Des Plaines, IA.

[5] Wickens, C. D. [1984]. Processing resources in attention. In R. Para-suraman & R. Davies (Eds.), Varieties of attention (pp. 63–101). New York: Academic Press.

[6] Young, K., Regan, M., & Hammer, M. (2003). Driver distraction: A review of the literature (No. 206). Clayton: Monash University Accident Research Centre.

[7] Horey, W.J., Wickens, C.D., 2006. Examining the impact of cellphone conversations on driving using meta-analytic techniques. Hum. Factors 48, 195–205.

[8] Lee, Y. C., Lee, J. D., & Boyle, L. N. (2007). Visual attention in driving: The effects of cognitive load and visual disruption. Human Factors, 49(4), 721-733.

[9] Just MA, Keller TA, Cynkar J. (2008). A decrease in brain activation associated with driving when listening to someone speak. Brain Research. 1205:70-80.

[10] Sheridan, T. B., Christian, C. K., Dierks, M. M. and Roth, E. M. (2009). Factors Contributing to Surgical Risk. In Marilyn Sue Bogner (Ed.) Human Error in Medicine. Second Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

[11] Michon, J. (1985). A critical review of driving behaviour models: what do we know and what we should do? In L. S. Evan (Ed.), Human behaviour and traffic safety (pp. 485-520). New York

[12] Godthelp, H., Milgram, P., and Blaauw, G.J. (1984). The Development of a Time-Related Measure to Describe Driving Strategy, Human Factors, 26 (3), 257-268.

[13] Greenberg, Neil. J.A. Carr, and C.H. Summers, Ethological Causes and Consequences of the Stress Response. *Integrative and Comparative Biology*. 42(3) (2002), pp. 508-516.

[14] Godthelp et al. (1984). The development of a time-related measure to describe driving strategy. Human Factors, 26, 257.

[15] Gibson JJ (1979). The ecological approach to visual perception. Houghton Mifflin, Boston.
[16] Olsson, G. (2010). Process Control, in Information Technology in Water/Wastewater Utilities, WEF Manual of Practice No. XXX, Water Environment Federation, USA
[17] Recarte, M. A., & Nunes, L. M. (2002). Mental load and loss of control over speed in real driving. Towards a theory of attentional speed control Transportation Research, 5, 111–122.
[18] Harbluk, J.L., Noy, Y.I., Trbovich, P.L. & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. Accident Analysis & Prevention, 39, 372-379.
[19] Westat (2000). NHTSA Driver Distraction Expert Working Group Meetings: Summary and Proceedings, Rockville, Maryland: Westat (http://www-nrd.nhtsa.dot.gov/Pdf/Crash Avoidance/FinalDDWorkingGroupProceedings11-10-00.pdf).
[20] Stevens, A. and Minton, R. (2001) In-vehicle distraction and fatal accidents in England and Wales. Accident Analysis and Prevention, 33, 539–545.
[21] McKnight, A.J. &McKnight, A.S. (1993). The effects of cellular phone use upon driver attention. Accident Analysis & Prevention, 25, 259-265.
[22] Brookhuis, K.A. & De Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics, 36*, 1099-1110.
[23] Tijerina L, Kiger S, Rockwell TH, Tornow C. (1995a). Final Report-Workload Assessment of In-Cab Text Message System and Cellular Phone Use by Heavy Vehicle Drivers on the Road. Report No. DOT-HS-808-467.Washington, DC: US Department of Transportation.
[24] Alm, H. & Nilsson, L. (2001). The use of car phones and changes in driver behaviour. International Journal of Vehicle Design, 26, 4-11.
[25] Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone . Psychological Science, 12, 462-466.
[26] Stutts, J.C., Reinfurt, D.W., & Rodgman, E.A. (2001). The role of driver distraction in crashes: An analysis of 1995-1999 Crashworthiness Data System data. In 45th Annual proceedings Association for the Advancement of Automotive Medicine. (pp 287-301). AAAM: Des Plaines, IA.
[27] Recarte, M. A., & Nunes, L. M. (2002). Mental load and loss of control over speed in real driving. Towards a theory of attentional speed control. Transportation Research, 5, 111–122.
[28] Jameson, T. L., Hinson, J. M., & Whitney, P. (2004). Components of working memory and somatic markers in decision making. Psychological Bulletin & Review, 11, 515–520.
[29] Glaze, A.L. & Ellis, J.M. (2003). Pilot Study of Distracted Drivers. Virginia Commonwealth University: Richmond, VA.
[30] Stevens, A. and Minton, R. (2001) In-vehicle distraction and fatal accidents in England and Wales. Accident Analysis and Prevention, 33, 539–545.

Sensor Web Architecture for Data Management in Power Supply Companies through Web GIS

Sanja Bogdanović-Dinić[1], Nataša Veljković[2], Leonid Stoimenov[3]

Abstract –**Retrieving and managing data in real time from Supervisory Control and Data Acquisition (SCADA) systems through Web GIS applications has become a necessity for every power supply company. This paper presents one way of dealing with these requests using Sensor Web concept. The solution is given as integration of GINISED and GINISSENSE systemsinto a unique system which enables data retrieval, monitoring and visualization using power supply network elements as data sources.**

Keywords– **Sensor Web, SCADA systems, Web GIS application, Power Distribution Companies**

I. INTRODUCTION

Sensor networks gained popularity over the past few years due to technological advances in sensor technology. Sensors are smaller, lighter, more reliable and portable. They are capable of monitoring and measuring features of observed phenomena and can be placed anywhere. Networks of interconnected sensors are used for intelligent gathering of sensor measurements. Sensor measurements are sent through the network to the control centres where are being processed and analyzed. The results can indicate if there is some critical situation in the field, allowing operators to react in time and prevent or mitigate the catastrophic consequences. In order to perform more detailed and comprehensive analyzes, the process of data gathering should be based on some intelligent rules and pursued by intelligent hardware components. The Sensor Web concept precisely presents an intelligent sensor network, comprised of sensor pods which can have built-in intelligent modules enabling them to make decisions while measuring. This practically means that a sensor pod can alert control centre operator only in cases when measured values exceed critical limits. Sensor pods within such network can communicate in order to share data and check the status of other pods. Sensor Web has been very interesting exploration field to many researches and is thus very popular.

Recently, there have been attempts for adding a visual dimension to Sensor Web, by combining it with Geographic Information Systems (GIS). GIS, as an information technology, which combines geographic locations of natural and artificial objects as well as other types of data in order to generate interactive visual maps and reports, is often used in combination with Sensor Web [1]. If used as a data source in a Web-based GIS application, Sensor web gains a visual dimension. The value of information gained from different types of sensors and systems attached as data sources in Sensor Web is increased greatly by adding GIS component that contributes to it in a geographical sense.

For the needs of Power Distribution company Jugoistok Niš, CG&GIS Lab, Faculty of Electronic Engineering in Niš, with the support of Ministry of Science of Republic of Serbia, developed a geo-information system GINISED [2]. GINISEDis a specialized geo-information system which allows recording, processing, analyzing and graphic presentation of specialized information about the electric power supply network, such as spatial data, temporal data, image and multimedia [3]. Recently, we have added new functionality to GINISED, concerning data retrieval, visualization and user notification on defined parameters received from Supervisory Control and Data Acquisition (SCADA) systems. This is done by integrating GINISED with GINISSENSE. GINISSENSEis Sensor Web based architecture for monitoring real-time data, and for reacting when possible danger is noticed. It can be applied to various environmental problems, since it enables connection toheterogeneous data sources. It is based on Sensor Web concept, and it isfully designed accordingly to Open Geospatial Consortium (OGC) specifications and recommendations for Sensor Web.

The rest of the paper is organized as follows. Section 2 describes Sensor Web concept. In Section 3, GINISSENSEarchitecture is presented. One possible application of GINISSENSEarchitecture in power supply companies, for data management trough Web GIS application is presented in Section 4. Conclusion is given in Section 5, followed by list of cited papers.

II. SENSOR WEB

Sensor Web is a relatively new concept which describes a type of sensor network especially well suited for environmental monitoring. This concept was first used by Kevin Delin of NASA in 1997, who defined it as a system of wireless, intra-communicating, spatially distributed sensor pods that can be easily deployed to monitor and explore new environments[4].

The main characteristic of Sensor Web is that all data collected by one sensor can be shared and used by all other sensors in the network, enabling sensor communication and collaboration.

[1]Sanja Bogdanović-Dinić,[2]Nataša Veljković and [3]Leonid Stoimenov are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: {sanja.bogdanovic.dinic, natasa.veljkovic, leonid.stoimenov}@elfak.ni.ac.rs.

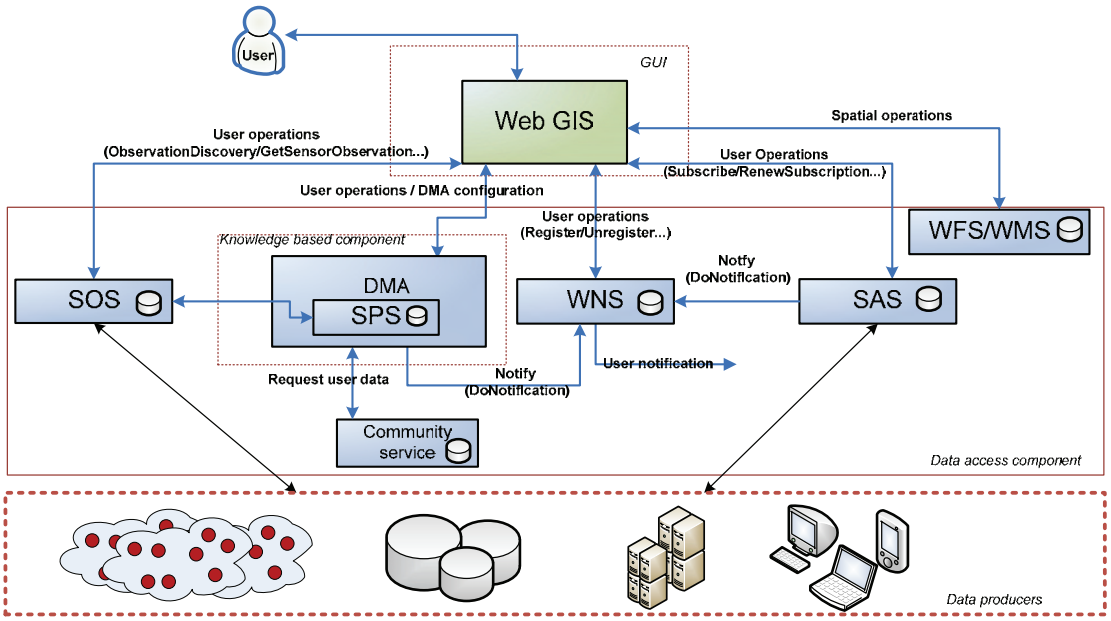


Fig. 1. GINISSENSESWE architecture

Sensor Web is thus an intelligent sensor network of collaborating sensor nodes capable of self- maintenance to some level. Another important characteristic of a Sensor Web is availability of sensors' measurements through the Web. This enables the development of Web systems for accessing and online processing of real-time sensor data. The Open Geospatial Consortium (OGC), as a leading organization in the field of developing new standards for geo-spatial and location services, has developed a set of standards and specifications named Sensor Web Enablement (SWE) [5].SWE represents a recommendation for implementing a Sensor web system and is comprised of four Web Services specifications: Sensor Observation Service (SOS), Sensor Planning Service (SPS), Sensor Alert Service (SAS) and Web Notification Service (WNS), and three modelling languages specifications: Sensor Markup Language (SensorML), Transducer Markup Language (TML) and Observations and Measurements (O&M). Web services are responsible for communicating with sensors, collecting their measurements and polling them when necessary. Modelling languages are used for modelling observations and measurements as well as for describing sensors. The proposed specifications can be applied in various situations: tracking floods, water contaminations, air pollution, traffic density, power supply networks etc.

The CG&GIS Laboratory at the Faculty of Electronic Engineering in Nis has been exploring the field of Senor Web for several years and has developed GINISSENSE architecture for applications in the field of environmental protection.

III. GINISSENSE

GINISSENSE is an architecture based on OGC SWE specifications [5]. GINISSENSEenables creation of systems for monitoring, acquisition, control, on-demand measurements and analysis of data received from heterogeneous data sources. Data sources are typically sensors or sensor networks, set in critical areas, with ability to measure different phenomena and deliver data to other systems for the purposes of further data processing.

The GINISSENSESWE architecture has the following components (as illustrated in Fig. 1): Data producers (sensors), Data access component (Web services), Knowledge based component (DMA) and Graphical user interface (Web GIS). Components of the architecture communicate using various protocols, media, topology, etc. The most common communication means are the Internet, satellite, mobile-phone or radio networks.

Data producers are any devices (sensors) or applications, capable of harvesting or measuring physical phenomena.Typically these are sensors, but can also be databases, archives, other systems and applications, etc.

Data access component is in charge for collecting and processing data from different sources (sensors). This primarily includes real time sensor data, spatial data necessary to display sensor position and objects of interest on the map, as well as data collected by users who contribute with gathered information regarding objects of interest. For each data type, there is a separate database used for data storing. Data access component comprises of seven different Web services. SOS, WNS, SAS and SPS are responsible for planning, acquisition, analysis and notifyingusers about sensor observations. Web Map Service (WMS) and Web Feature Service (WFS) are used for accessing geographical data. Community service is an external service for gathering and retrieving data from environmental friendly users. Using this service, a user can submit a photo with a brief description about water pollution, landfills or other environmental disasters. From this service point of view, humans are data producers.

Knowledge based component is a component used for comparing and analyzing data obtained from different

sources, making action plans, running on demand or automated actions and proposing action plans to system operators.

Graphical User Interface (GUI) is a Web GIS application [6]. Besides the basic GIS functionality, the Web GIS application provides users with support for accessing Web services layer. Spatial location, from which the sensor information is obtained, is very important in the analysis, which is why GIS is used as default.

IV. USING SCADA AS DATA SOURCE IN GINISSENSE ARCHITECTURE

GINISSENSESWE architecture, presented in Section 3, can be successfully applied on electric power supply network for communicating with SCADA systems, getting SCADA measurements and informing users about critical events. The CG&GIS Laboratory within the Faculty of electronic engineering in Nis has developed SCADA module, which relays on GINISSENSE SWE architecture, andhas integrated it withGINISEDsystem [2]. GINISED is a Web GIS solution that provides user with interactive geographical representation of electrical substations' locations and enables getting SCADA information about these substations: elements and bays connected to a substation as well as real-time measurements for selected element. A user can get a visual representation of element's measurements for a concrete date and specified time range along with spreadsheet overview. They can also subscribe for receiving notifications about certain element's measurements via email address or SMS and define subscription criteria (e.g. notify me if a value exceeds its limit).

For realization of described scenario several components should be included: SAS and WNS services within GINISSENSESWE and SCADA service and SCADA module within GINISED.

A. GINISSENSESAS: Accessing SCADA Data

SAS is by definition used for sensor advertisement and user subscription. In this usage scenario, SCADA represents primary data source and as such it should advertise its measurements with SAS. The advertising process is standardized by OGC SAS specification and it implies that data source sends a detailed document with meta-information as well as measurements information. Since SCADA elements are the components that perform measuring, they all should be listed in this document. Basic Advertise request elements are sensor description and message structure. They are both structured in accordance with SensorML specification [7]. Sensor description contains information about a data source, e.g. SCADA, which includes inputs, outputs, parameters, processes and methods, along with relevant metadata. Message structure describes data format used for data encoding. After successful data source registration SAS generates a unique identifier for the data source and creates an XMPP channel that will be used by that source for data advertising. Each time SCADA has new measurements, it will push an alert message into XMPP channel. SAS receives the message and puts data into a measurement database making it available for other system components. An example of an alert message, that shows measurement coming from an element with id 77 attached to SCADA with id 18, is given in Fig. 2.

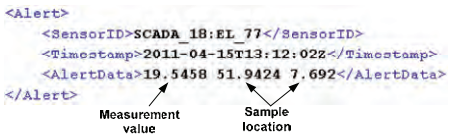


Fig. 2. Alert message

SAS enables user subscription for receiving measurements. Invoking a Subscribe request, a user sends information about SCADA elements that is interested in, as well as contact information (email address or SMS) and notification criteria (critical values, value ranges etc.). SAS then must register subscribed user with WNS in order to use this service for user notifying. Upon receiving the sensor alert message, SAS performs filteringsubscribed users upon the type of sensor or measured phenomenon andwhen the conditions are matched, it sendsnotifications via WNS.

B.GINISSENSEWNS: User Notification

GINISSENSE Web Notification Service enables asynchronous message interchanges between clients and other GINISSENSE services [8]. WNS provides interface for user registrations. It is possible to register a new, single user, or to form a group of already registered users. Registered users receive notifications from WNS using email or SMS as communication protocols.WNS is invoked by some other service within the architecture; in this case by SAS when new measurements are available and user defined conditions for notification process are fulfilled.

C. GinisED: SCADA module

SCADA module enables communication with SCADA system from Web GIS. It gives a visual representation of available electrical substations on an interactive map and allows user to choose on by clicking its icon on a map. SCADA module is than induced, calling SCADA service in the background. If selected SCADA system is advertised with SAS there will be data for display. SCADA service communicates with database that contains information about bays and elements connected to SCADA and database that contains elements' measurements. It retrieves this information and returnsit to SCADA module, which than presentsit to the user. The module also enables user to retrieve measurement information from any element for a specified date and time period. In that case SCADA service gathers required information and sends it back to the module where it is presented as spreadsheet, as well as graphically in the form of chart with indicated measurement values.Fig. 3 represents one usage example of this module. SCADA service is the one responsible for communicating with GINISSENSE in the background. The communication is, for now, based on reading databases populated by SAS.

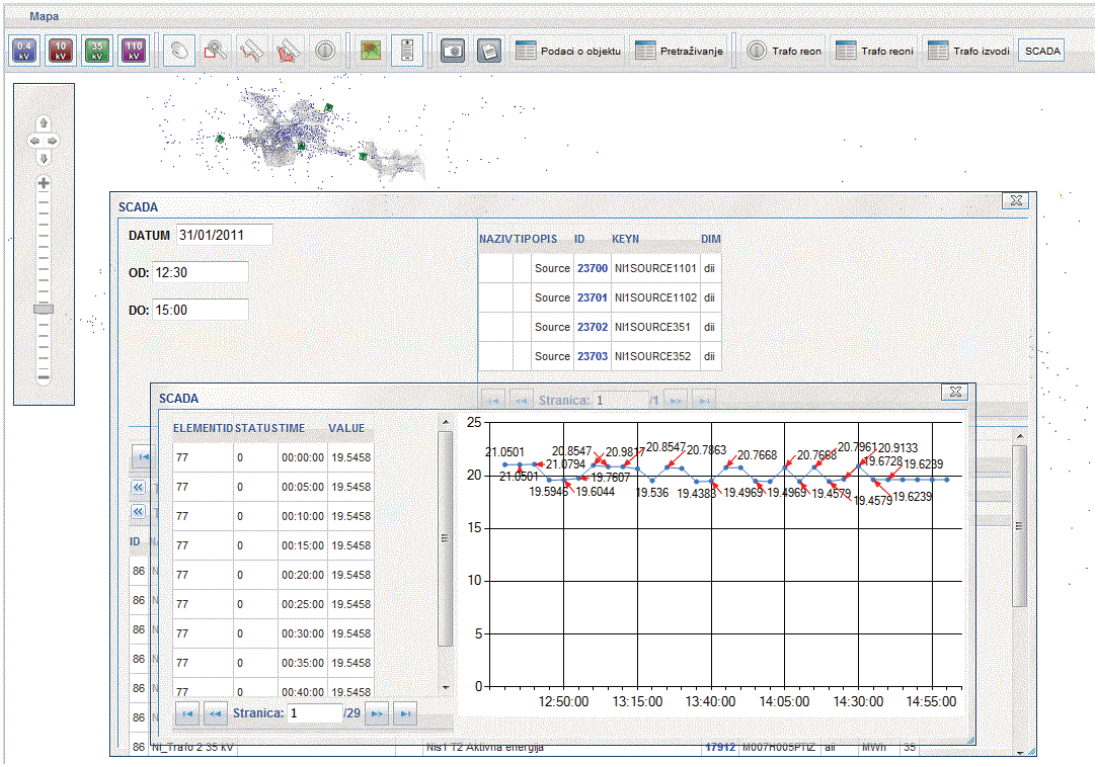


Fig. 3. GINISEDSCADA module

V. CONCLUSION

The technology today provides possibility of developing efficient monitoring systems that can be installed anywhere and used for all types of environmental issues. Sensor Web has become a primary concept in this area and a standard for developing such systems.On the other hand, visual representation of all types of data on interactive maps is crucial in monitoring applications. Combining GIS and Sensor Web we can develop powerful applications in this area.Power Supply Companies have a necessity to track, collect, analyze and visualize data and measurements coming from electrical network, especially to have an insight into elements' behaviour over certain period. GINISSENSE architecture provides possibility to use electrical network as data sourceand apply all Sensor Web based functionalities to data retrieved from components attached to electrical substations in the network. That further means that we can use GINISSENSE in GINISED system in order to collect measurements obtained by SCADA components, perform different kinds of analyses and visualize their changes using charts or other graphictools.

In this paper, we have presented GINISEDsystem integrated with GINISSENSE resulting in a powerful application for visualization and monitoring of data retrieved from SCADA systems and provided an efficient data management tool for Power Supply Companies.

REFERENCES

[1] N. Veljković., M. Bogdanović., S. Bogdanović-Dinić., L. Stoimenov., "GinisSense - Visualizing Sensor Data", in Proceedings of SGEM conference, Albena, Bulgaria, 19.-25. June, pp. 1119-1126, 2010.

[2] L. Stoimenov, S. Đorđević-Kajan, D. Stojanović, M. Kostić, A. Vukašinović, A. Janjić, "Geographic Information System for evidencing, maintenance and analysis of electric power network", YU INFO 2006, Kopaonik, 2006 (in Serbian).

[3] L. Stoimenov, A. Stanimirović, M. Bogdanović, N. Davidović, A. Krstić, "GinisED – Geo-Information System for Support of Evidencing, Maintenance, Management and Analysis of Electric Power Supply Network", 3rd Small Systems Simulation Symposium, pp. 23-26, 2010.

[4] K. A. Delin, S. P. Jackson, "The Sensor Web: A new instrumentconcept", Jet Propulsion Laboratory,California Institute ofTechnology, SPIE Symposium on integrated Optics, 2001.

[5] I. Simonis, "Sensor Web Enablement Architecture", OpenGIS BestPractice Paper, OGCDocument Number: 06-021r4, Version:0.4.0, 2008.

[6] M. Bogdanovic, N. Davidovic, I. Antolovic, A. Stanimirovic, D. Stojanovic, L. Stoimenov, "WebGIS application for viewing and analysis of electric power supply network geodata", YUINFO, Kopaonik, 2008. (in Serbian)

[7] M. Botts, (Ed.), "OpenGIS® Sensor Model Language Implementation Specification", OGC Document Number: 05-086r2, 2006.

[8] S. Bogdanovic-Dinic, N. Veljkovic, L. Stoimenov,"Web Notification Service of GinisSense architecture", YUINFO, Kopaonik, 2010.(in Serbian)

Measuring Points System for Wayside Dynamic Control of Vehicles on Serbian Railway Network

Života Đorđević[1], Slavko Vesković[2], Simo Mirković[3], Slaviša Aćimović[4], Aleksandar Radosavljević[5]

Abstract –**For the purposes of the control of train weight status and the control of good wagons technical state SR developed the concept for wayside measuring stations system for rail vehicle dynamic control on the railway lines. This paper gives an overview of the two measurements: The first system is designed for monitoring the temperature condition of rolling bearings, wheels, and brake discs of all railway vehicles. The second system is a dynamic scale for measuring the axle loads, unbalanced loading, and detection of flat places on the surfaces of rolling wheels.**

Keywords – **dynamic control, railway vehicles, wheels, rolling bearings, flat places**

I. INTRODUCTION

An extraordinary event in railway operations implies an event which impedes or makes service impossible, endangers human lives and destroys railway property and goods in transportation. Extraordinary events can be classified according to their causes and consequences: crashes, accidents and natural disasters (Table I) [1].

By observing the accident database of one infrastructure manager, one can see a high number of small derailments at shunting yards and less on normal track between stations, but with high degree of loss (Table II, III).

Automation of wayside train monitoring leads to higher estimation accuracy and minimizes human necessity in railway operation. Consequently, the impact of human factors is becoming less important in case of extraordinary events. Current worldwide tendencies are to minimize the influence of human factors in extraordinary events which would be impossible without the application of modern systems for railway operation control and wayside monitoring equipment [2,3, 4]. The new system can dynamically detect 75% of the causes of cars' exclusion from railway network traffic (Table IV).

On the other hand, for railway vehicle maintenance it is also very important to act in a timely manner. Furthermore, from the aspect of safety and reliability, the wheel sets are the second most important assembly, just after the braking system. Poor wheel conditions can often lead to derailments, whereas early detection of wheel faults brings numerous benefits to infrastructure owners and also to railway operators. In order to increase safety, improve rolling stock maintenance and to protect infrastructure public enterprise "Serbian Railways" plans to create a system for wayside monitoring. Implementation of wayside monitoring on the Serbian Railways network will decrease the influence of human factors in vehicle inspection and faults will be identified in a timely manner.

The first installation for wayside train monitoring on Serbian Railways will be located near Batajnica station and will be a result of close collaboration with Austrian Federal Railways and the Vienna University of Technology. The installation will have two TK99 measuring groups for hot-box, hot wheel and hot disc detection and dynamic scale G-2000 for wheel set weighing, flat spot detection and detection of uneven loading of wagons. Both devices are developed and assembled by Infrastructure of Austrian Federal Railways (ÖBB - Infrastruktur AG).

In the scope of early defect detection, it is indispensable to have equipment capable for contactless recognition of overheating, flat spots and uneven loading, accurate data processing and transfer to remote places. Identification of freight cars which are loaded out-of-gauge should be left for the final phase of the project since such equipment is still in experimental use on the other railway networks (i.e. BLS, ÖBB) [5]. Equipment for detecting gauge overload should be restricted to locations close to border stations and before tunnels with lower contact wire where lorries are transported by Ro-La-trains [6]

II. LOCATION AND EQUIPMENT DISPLACEMENT IN MEASURING STATION AT BATAJNICA

The installation location is going to be km 24+776 (from km 24+734 to km 24+818) on the left side of the double track line No. 5 from Belgrade to Šid - border line Serbia/Croatia will collect data from all railway vehicles coming from the West and North of Serbia (Fig. 1) [7].

The installation will consist of modules TK 99 (module 3) and G-2000 (module 4) placed on track and cabin (module 1) at km 24+776 on the left side of line 5, next to the left track. The outside installation part has axle counters RSR180 (2x3, 6 in all – Fig. 2), two TK99 modules (Fig. 3), one on each track, and one scale G-2000.

[1]Života Đorđević is with the Serbian Railways, Nemanjina 6, 11000 Belgrade, Serbia, E-mail: zivota.djordjevic.@srbrail.rs

[2]Slavko Vesković is with the University of Belgrade - Faculty of Traffic and Transport Engineering, V. Stepe 305, 11000 Belgrade, Serbia, E-mail: veskos@sf.bg.ac.rs

[3]Simo Mirković is with the Institute of Transportation, Nemanjina 6, 11000 Belgrade, Serbia, E-mail: smirkovic@sicip.co.rs

[4]Slaviša Aćimović is with the University of Belgrade - Faculty of Traffic and Transport Engineering, V. Stepe 305, 11000 Belgrade, Serbia, E-mail: slavisa@sf.bg.ac.rs.

[5]Aleksandar Radosavljević is with the Institute of Transportation, Nemanjina 6, 11000 Belgrade, Serbia, E-mail: radosavljevica@sicip.co.rs

TABLE I - EXTRAORDINARY EVENTS

Type of Extraordinary Event	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*
Accidents	21	33	30	27	11	18
Operating incidents	435	447	430	383	349	375
Accidents and Incidents at Level Crossings	198	209	193	131	168	181
Total	654	689	653	541	528	574

TABLE II - TYPES OF ACCIDENTS

Types of Accidents	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*
Collisions	1	1	1	-	-	-
Overtaking	1	2	4	-	-	-
Car Derailments	16	26	23	26	11	16
Derailments and Overtaking at Shunting yards	-	-	-	66	79	1
Collisions, Overtaking and Derailments of Maintenance Vehicles	2	3	1	-	-	-
Other Accidents	1	1	1	1	-	1
Total	21	33	30	93	90	18

TABLE III - TYPES OF OPERATING INCIDENTS

Types of Operating Incidents	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*
Collisions Avoided	7	4	6	7	3	4
Overtaking Avoided	2	6	4	1	2	2
Signal Passing	15	12	14	11	10	8
Derailments and Overtaking at Shunting yards	76	83	68	66	72	86
Collisions, Overtaking and of Maintenance Vehicles	-	-	1	-	-	-
Other Operating Incidents	335	342	337	298	262	275
Total	435	447	430	383	349	375

TABLE IV–THE NUMBER OF CARS EXCLUDED FROM TRAFFIC IN 2010.

Passenger Freight	Overheated axle-box	Wheelset	Suspension spring	Frame and bogie	Buffer	Drawgear	Brake disks	Σ
P	89	581	29	352	288	54	1121	2514
T	148	1524	367	782	625	235	2447	6128
Σ	237	2105	396	1134	913	289	3568	8642
%	2.74	24.36	4.58	13.12	10.56	3.34	41.29	100

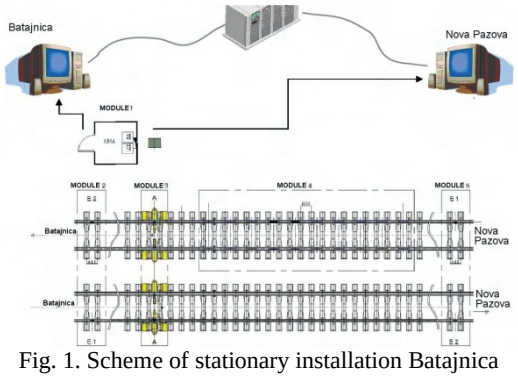


Fig. 1. Scheme of stationary installation Batajnica

All six axle counters are identical, only their positions are different (km 24+734 - module 2, km 24+776 - module 3 and km 24+818 - module 5, one on each position on both tracks). The mid counters will be positioned at the same locations as TK99 modules and the end counters will be at a distance of 42 m in each direction. The TK99 measures wheel set bearings temperatures, wheel body temperatures and brake disc temperatures.

Dynamic scale G-2000 will be positioned from km 24+779 to km 24+789,80 on the left track on railway line No. 5 Belgrade – Šid. Strain gauges are sensors which are used for measurements in rail deformations caused by vehicle wheels passing over them. Strain gauges are positioned at a distance of 1.2 m directly on the rail neck between sleepers, in total 10 strain gauges on each rail, 20 in all on that one track. A PC inside a cabin (module 1) which is connected to G-2000 (module 4) calculates the axle loading for each wheel set upon rail strain measured between sleepers and wheel flat spots from impulse forces exerted over the rails.

The cabin equipment (module 1) will consist of power supply electronics for TK99 sensors (scanners), 2 PCs for data storage, calculation and transfer, and UPS units for 15 minutes power supply in case of power supply interruption. The cabin will be thermally isolated and will have a base of 2.40x2.40 m. All electronic equipment will sit on two movable racks. The distance between the front of the cabin and the centre of the track will be 6 m. The power supply will be 5 kW max. power, 230V/50 Hz, from a catenary transformer. Connections between cabin PCs and train inspectors PC terminals in Batajnica and Nova Pazova stations will be done by modems.At Batajnica and Nova Pazova stations the computers will be located at movements inspector's office. They will be served by authorizes personnel from the Department of technical vehicle services of the ŽS.

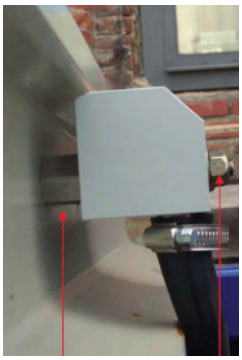


Fig. 2 Counter pick-up device.

III. Experimental Methodology

Generally, the trains are registered in both travelling directions; weighbridge axles' temperature is measured contactlesson the left and on the right side. Simultaneously, the wheel center and brake disks temperature is also measured. Every body that has a temperature above absolute zero emits electromagnetic radiation proportionally to its temperature. The wavelength of that radiation is within the range 0.7 – 1000 μm. The range of interest for technical measurements is 0,7 μm to 14 μm (Fig. 4). Since in this spectrum range vapor and carbon-dioxidecause transmission losts, only certain measurement windows can be used. Typical measurement windows are 1,1...1,7 μm, ...2,5 μm, 3...5 μm, and 8...14 μm (Fig. 5). However, these measurement windows have different radiation maximums depending on the object temperature (Fig. 6). The most suitable range for measuring axle and wheels' temperature (0-600°C) is 3 to 5 μm. For this range we use infrared (IR) detectors with thermo-electrical cooling, since they have response time shorter than 5 μs.

The most important characteristic of the device for detection of overheated axle bearing (HOA) is double checking of axle bearing' pair of wheels. Both bearings of the same axle are checked both vertically upwards and horizontally from the outside. By doing this, besides bearings of any construction kind opened downwards, any axle-bearing type with reconstructed bearing is certainly detected (ex. Y25, Y31). Modular structure of the devce allows for different layout of measurement points on wheels and disks (FOA and SOA). The temperature of disk brake is measured verticallydownwards. The temperature of wheels is measured by using a special sensor in the flange of wheel area.

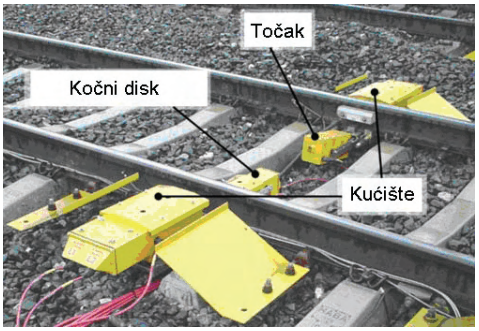


Fig. 3 –The device for axle weighbridge's wheel (axle), wheels and brake disks temperature measurement

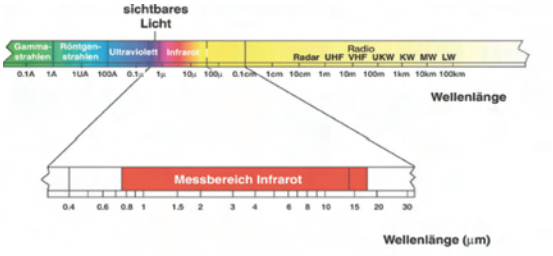


Fig4. Measurement area from 0,7 μm to 14 μm
(Sichtbares Licht=Visible light;
Wellenlange=wavelength;
MessbereichInfrarot=infrared measurement area)

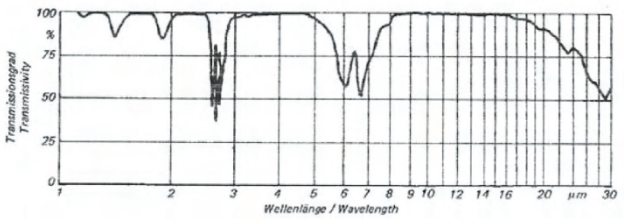


Fig 5. Transmission degree.

The G2000 system uses gauges which follow rail deflection caused by the wheel seating force. Sensors are put onto the rail, between sleepers (with axle distance of 1.2 m), so that one stretch of rails has ten sensor pairs. A computer loads measurement data when the train passes through measurement point and calculates axle loads of every wheel and the size of flat places, if there are any. Sensor device is mounted on the rail side, along neutral line. Gauges are welded onto the rail. The complete sensor device has four parts, mounted very near the measurement module (one pair on the inner rail side, and the outer pair on the outside, Fig. 7).

IV. Warnings and Alarms

In a case of any irregularity detected on trains passing over the installation, pictogram alarms will be shown on each monitor connected to the stationary system network (Fig. 8). Pictograms will be accompanied by exact values represented in data tables. In accordance to procedures, the train will be stopped and the faulty car will be removed from the train.

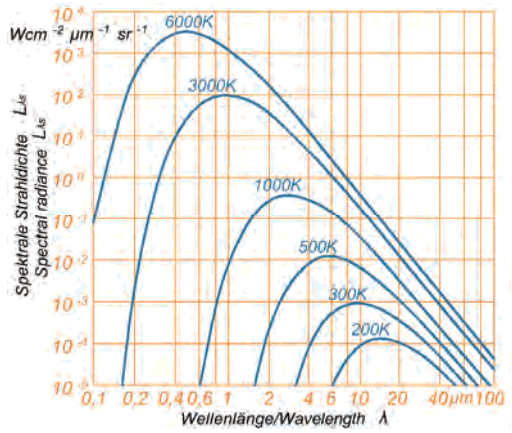


Fig.6 Spektraldensity of L ray λs

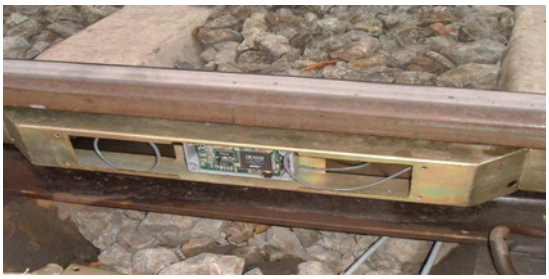


Fig. 7 Measurement device in protective box; sensors are on the left and on the right.

Displayed values will include train identification number, date and train time, direction, train speed, and mass, axle number, total number of axles, measured temperature, track number, fault identification and description, etc.

Timely detection of technical irregularities and avoiding additional damage offers diverse advantages in railway traffic flow: better security, avoiding traffic disturbances, lowering costs due to reduced number of traffic accidents, lowering risk in dangerous goods traffic, lowering the risk of tunnel accidents, better traffic quality, higher speed and axle load, longer periods between inspection and vehicle maintenance, optimization of traffic control, less frequent active railway maintenance, smaller superstructure load.

V. Conclusions

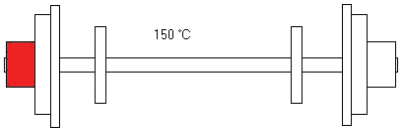
As it is clearly shown in the project sensitivity analysis, the project is highly resilient to all variations of input parameters and also to expected divergences. The facts indicate investment return and necessity of immediate project realization which should significantly reduce costs for both infrastructure and vehicle maintenance. Procena direktnih efekata od uvođenja ovog sitema ide i do 500.000 € godišnje. Results for the investment estimate are [8]:

Internal rate of return (IRR) = 11.99%,

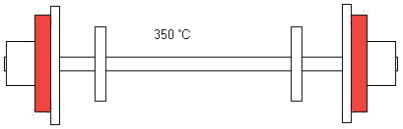
Net present value (NPV) = 603.719. €

By embedding measurement stations for wayside dynamic control of the technical state of railway vehicles, Serbian railways will join current European transportation system by improving the reliability and quality of service. Batajnica measurement station is going to be one of the first steps in this direction.

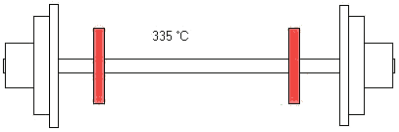

a) Overloaded wheelset and wheel flat spot



b) Overheated bearing detection



c) Irregular wheel temperature



d) Irregular brake disc temperature

Fig. 8

Acknowledgement

This paper is part of the research financed by Ministry of Science, Republic of Serbia in Project: MNTR36012 – “Research of technical-technological, staff and organisational capacity of Serbian Railways, from the viewpoint of current and future European Union requirements”.

References

[1] Institute of Transportation CIP, General design of check point system for wayside monitoring on network of Serbian railways, Belgrade, 2008.

[2] Karner, J., Maly, T., Schöbel, A., TK99-the austrian solution for hot box detection, XIII Scientific-Expert Conference on Railways, Nis, 09-10 October 2008, Proceedings, pp. 57-60.

[3] Schöbel, A., Karner, J., Components for wayside train observation in Austria, XII Scientific-Expert Conference on Railways, Nis, 19-20 October 2006; Proceedings, pp. 25-27.

[4] LeDosquet, G., Pawellek, F., Müller-Boruttau, F., Lasca: automatic monitoring of the running qaulity of railway vehicles, RTR No. 2, pp. 34-39, 2007.

[5] Bocchetti, G., Lancia, A., Mazzino, N., Valcada, A., TCCS - train conformity check system, 7th International Conference & Exhibition On Rail Technology Rail-Tech Europoe 2009, 2009.

[6] Stadlbauer, R., Schöbel, A., Pisek, M., Check point: solutions for automatic supervision with coupling to automatic train protection systems, 2nd International Railway Symposium, Istanbul, 15-17 October 2008, Proceedings, pp. 1035-1044.

[7] Djordjević, Ž., Karner, J., Schöbel, A., Mirković, S., Batajnica checkpoint for wayside train monitoring, XIV Scientific-Expert Conference on Railways, Niš, 07-08 October 2010; Proceedings, pp. 189-192.

[8] Radosavljević, A., Đorđević, Ž., Mirković, S., Concept for Wayside Train Monitoring At Serbian Railways - Pilot Project Batajnica, RTR Special, pp. 6-11, 2011.

Estimation of NO_2Immission Concentrations from Teko-B Power Plant and Measuring Locations Selection

Jelena Malenović-Nikolić[1], Goran Janaćković[2]

Abstract – **Thermoenergetic systems are major sources of pollution in gaseous, liquid and solid physical state. Continuous monitoring and control of the impact of energy production and processing is essential for the preservation of human health and environmental protection. In this paper data on the impact of gaseous products of combustion of coal, power plant Kostolac B (TEKO B) are presented, and air quality in the environment is analyzed. Based on the concentrations of emissionsnitrogen dioxide imission concentrations around point source of pollutionare estimated. It is proposed location of measuring stations at the place where the maximum level of pollutionis achieved, and in the direction of dominant winds.**

Keywords– **Air pollution, monitoring, power plants, immission prediction, Gaussian model.**

I. INTRODUCTION

Environmental pollution has initiated an extensive study of cross-border transfer of pollutants. The implementation of such studies is only possible international cooperation of many countries. Stockholm United Nations Conference on Environment (1972) has approved the basic principles of building a global monitoring system and recommended the organization of the cells to measure the quantity of pollutants [1]. In the framework of the United Nations on environmental issues (1973-1974) are developed basic principles of creating a global environmental monitoring system [1].

To be a part of that global environmental monitoring system, there is a need to monitor environmental quality parameters, such as air pollution, at local or regional level. By means of collaboration and data exchange among different local stakeholders, the environmental pollution problems can be more easily identified and solved.

II. PROBLEM DESCRIPTION

Thermo-energetic systems are major sources of pollutantsin gaseous, liquid and solid physical condition. Continuous monitoring and control the influence of the production and processing energy is essential for the preservation of human health and the environment. There is the impact of gaseous products of combustion of coal, power plant Kostolac B (TEKO-B), on the quality of air in the local environment.

Air quality monitoring and analysis of thermal power plants on the environment is carried out according to the Regulations on limit values, methods of measuring emissions, establishing criteria for measuring points and data records [2].

Air pollution has influence on people and living species. There is also influence on buildings. The archaeological site Viminacium is in the vicinity of pollution source.On the Fig. 1 is presented the location of that archaeological placein relation to the power plant, and the Fig. 2 presents the map where is shown thermal power plant, archaeological location and nearby settlements.

Viminacium (Viminacivm) was a major city (provincial capital) and military camp of the Roman province of Moesia, and the capital of Moesia Superior. It is an archaeological site located near the power plant TEKO B (Roman baths are less than 2 km from the source of pollution, as shown in Fig. 2), which dates from the reign of Hadrian (about 117th BC), when the settlement got the status of the city, and later became a Roman colony. The city was the capital of the Roman province of Moesia Superior, and Roman legion VII was stationed there.

Fig.1. The distance between the power plant and Viminacium

Plans for future Viminacium include the idea that it becomes one of the cultural centres of Serbia. The observed area is significant from the historical and cultural standpoint, so the planned construction and future work of TEKO-B require special attention, in order to protect thenearby historicalmonument.

III. MONITORING SYSTEM

Technical monitoring system consists of a set of measures and activities undertaken with the aim of monitoring and improving environmental quality. Within it there are mechanisms for keeping data on physical and chemical effects in selected samples, changes in biological processes in the surrounding environment, the level of pollution of air, water and soil and changes that occur in flora and fauna, and that are caused by anthropogenic and natural influences. The task of

[1]Jelena Malenović-Nikolić is with the University of Niš, Faculty of Occupational Safety, Čarnojevića 10a, 18000 Niš, Serbia, E-mail: jelena.malenovic@znrfak.ni.ac.rs.

[2]Goran Janaćković is with the University of Niš, Faculty of Occupational Safety, Čarnojevića 10a, 18000 Niš, Serbia, E-mail: goran.janackovic@znrfak.ni.ac.rs.

monitoring is to collect information on the state of the environment, and on the basis of these forecasts to provide quality basic elements of the environment, and to suggest the necessary prevention or maintenance measures.

Fig.2. Location of analysed pollution source

Monitoring of air environment includes the installation of measuring points in the area where the expected maximum concentration of pollutants. The air pollutants can be found due to the accumulation of large amounts of waste water and solid waste in the country, which may make the discharge of groundwater to the surface, due to an excessive load of soil. Following the most damaging by-products of burning fossil fuels (coal and, as one of them) in thermal power plants are sulphur oxides, nitrogen and carbon, radioactive materials, soot, ash, dust and slag.

At selected measuring points is necessary to continuously measure the concentration of sulphur dioxide, carbon monoxide, carbon dioxide, nitrogen dioxide, soot and suspended particles [3].Monitoring can be carried out from time to time (occasionally) or periodically (at regular time intervals), depending on the compelling need (hourly, daily, monthly, etc.). Continuous measurement is aimed to monitor the performance of environmental protection systems at large distances.

The system for monitoring the sources of air pollution is a system designed that after sampling records and transmits the required measured values (parameters) to the destination.

IV. Calculation of Nitrogen Dioxide Immission Concentrations

By analysing the results of monitoring set out in the report of the Mining Institute - Belgrade, for the period from 2008 to 2011, it isfound that there aredays when nitrogen dioxide immission levels are above the acceptedvalue [4-6]. These conclusions formed the basis to determine the maximum value of emission and compare with the limit values.

In this paper is applied the Gaussian method for forecasting the impact of nitrogen dioxide on the air quality. Bearing in mind the pace of reaction of nitrogen oxides, their interaction and reactions with components in the atmosphere,it cannot be ignored their impact, especially if it is known that nitrogen dioxide remains in the air and up to 4,5 years. Nitrogen dioxide in the presence of moisture easily enters the reaction and form nitric acid.

Gaussian model is the most common type of model, in which is assumedthat the propagation of air pollution is based on Gaussian distribution, which means that the propagation of pollutants is normal probability distribution [7]. Gaussian models are commonly used to predict the distribution of air pollutants originating from the ground or elevated sources. The primary algorithm used for Gaussian model is based on generalized dispersion equations for a continuous point source [8].

Simplified algorithm for estimation ofimmission concentrationson certain distance from the source is shown in Figure 3, while the calculation results are shown in Figure 4.

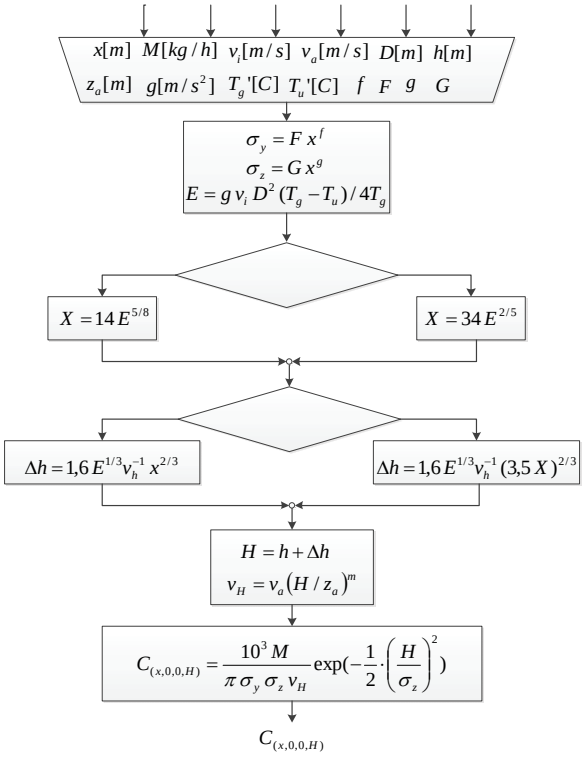


Fig.3. Simplified algorithm for calculation of immission concentrations of nitrogen dioxide

Immission concentration of pollution is determined by the following equation:

$$C_{(x,0,0,H)} = \frac{10^3 M}{\pi \sigma_y \sigma_z v_H} \exp\left(-\frac{1}{2}\cdot\left(\frac{H}{\sigma_z}\right)^2\right) \qquad (1)$$

whereM isthe intensity of emission from sources (g/s); σ_y , σ_z arecoefficients which includes fluctuations in horizontal and vertical planes; v_H is thewind speed at

effective height of source (m/s); and H is effective source height (m).

Calculated immission concentrations of nitrogen dioxide are shown on Table I. On this table are presented the values of immission concentrations for nitrogen oxides at a distance of 100 to 10,000 meters from sources of pollution.

TABLE I
IMMISSION CONCENTRATIONS OF NITROGEN DIOXIDE FOR FIVE DIFFERENT EMISSION MEASUREMENTS(C_1-C_5)

X (m)	C_1 (µg/m3)	C_2 (µg/m3)	C_3 (µg/m3)	C_4 (µg/m3)	C_5 (µg/m3)
100	1.1E-20	1.5E-20	1.0E-20	1.1E-20	9.5E-21
200	7.54E-5	8.68E-5	7.61E-5	5.83E-5	5.21E-5
300	0.1807	0.1992	0.1858	0.1278	0.1174
400	3.2702	3.5292	3.3914	2.2134	2.0606
500	12.7164	13.5525	13.2562	8.3876	7.8696
600	26.0879	27.5742	27.2880	16.9168	15.9534
700	39.1424	41.1303	41.0419	25.0773	23.7352
800	49.5392	51.8276	52.0370	31.4533	29.8507
900	58.6335	62.5066	61.0942	38.6478	36.2869
1000	69.3496	73.2559	72.5378	44.8677	42.3591
1100	76.4495	80.1999	80.1957	48.7720	46.2363
1200	80.4598	83.9572	84.5913	50.7764	48.2909
1300	82.0710	85.2779	86.4378	51.3512	48.9613
1400	81.9304	84.8432	86.4124	50.9108	48.6404
1500	80.5699	83.2032	85.0763	49.7840	47.6430
1600	78.3987	80.7752	82.8633	48.2168	46.2068
1700	75.7187	77.8638	80.0952	46.3866	44.5044
1800	69.6363	71.3928	73.7546	42.3990	40.7525
1900	72.7470	74.6860	77.0043	44.4187	42.6580
2000	66.4920	68.0872	70.4597	40.3858	38.8454
2500	52.0281	53.0583	55.2274	31.3386	30.2175
3000	40.8787	41.5919	43.4344	24.5075	23.6636
3500	32.6685	33.1907	34.7318	19.5282	18.8720
4000	26.6004	26.9998	28.2918	15.8700	15.3455
4500	22.0407	22.3566	23.4488	13.1318	12.7029
5000	18.5478	18.8044	19.7368	11.0397	10.6823
5500	15.8214	16.0344	16.8382	9.4099	9.1073
6000	13.6563	13.8362	14.5357	8.1175	7.8577
6500	11.9099	12.0640	12.6780	7.0761	6.8506
7000	10.4814	10.6152	11.1583	6.2251	6.0274
7500	9.2985	9.4158	9.8996	5.5209	5.3460
8000	8.3079	8.4117	8.8454	4.9315	4.7757
8500	7.4701	7.5626	7.9538	4.4333	4.2934
9000	6.7551	6.8382	7.1927	4.0082	3.8820
9500	6.1400	6.2150	6.5379	3.6427	3.5281
10000	5.6068	5.6750	5.9704	3.3259	3.2215

For the purpose of immission concentration prediction it is created an application TEKO-Air, in which are used the following basic parameters: emission speed (g/s), the amount of pollution sources (m), inner diameter of pollution sources (m), temperature of gas at the source (K), and ambient temperature (K). It is considered a point source of pollution, and it is taken into account the coefficient of dispersion for rural areas. It is a simple site, without the influence of wetlands, and it is considered of all the weather condition information(stability class and wind speed). For further calculation of immission concentration is used Eq. (1), and other equations presented in Fig. 3.

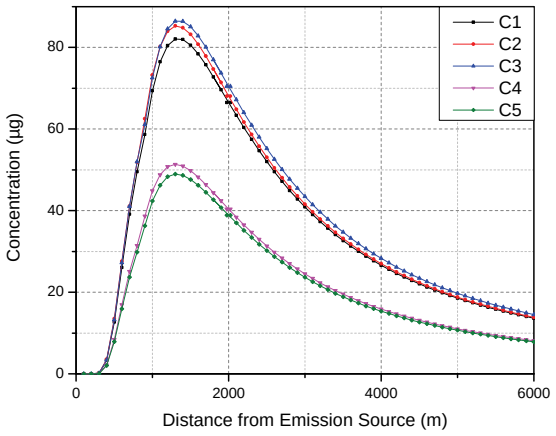


Fig.4. Calculated immission concentration of nitrogen dioxide

Fig.5shows the wind rose for the area of Požarevaccity, according to measurements of nearest hydro meteorological station Gradište. This wind rose is used during the consideration ofair pollution propagation.

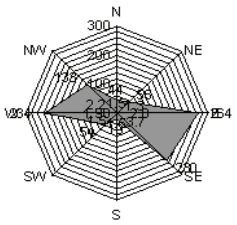


Fig.5. Wind rose for the city of Požarevac

In this area the dominant wind direction is the east and southeast, and the second one is west and west-north west direction. The strongest winds are from the east (27%), with mean speed values over 4 m/s.

Based on the presented results, the problem of increased emission levels of pollutants is solved thanks to the setting up proper chimney height, but there are still occasionally increased concentrations of nitrogen dioxide. Serbian Air Protection Act defines daily limit for concentrations of nitrogen dioxide for non-urban areas (0.07 mg/m^3) and urban areas (0.085 mg/m^3) [9].

V. MEASURING LOCATIONS

Monitoring stations can be stationary (for systematic and long-term measurements), semi-mobile (for continuous monitoring using mobile laboratories to cars) and mobile (for monitoring the current concentration of emission sources, under the smoke plume).

Based on the results obtained by measuring the immission concentrations of nitrogen dioxide, to increased levels of pollution are occasionally exposed surrounding small towns: Kostolac, Bradarac, Klenovik and Petka.

Based on the calculation of immission concentrations of nitrogen dioxide it is necessary to set the measuring points at

approximately 1300 m (the radius of the circle presented) from the source of pollution in the direction of dominant winds, as it is shown in Fig. 6.

Fig.6. Distance where is calculated maximum immission concentration in the vicinity of the thermal power plant

As it is previously mentioned, Roman baths are less than 2 km from the power plant and the air pollution has strong influence on thosearchaeological remains.

VI. Conclusion and Future Research

In this paper is calculated immission concentration of nitrogen dioxide near thermal power plant TEKO-B. For that purpose is used Gaussian model of atmospheric dispersion. It is concluded that the maximum immission concentration is approximately 1300 m distant from the source of pollution, measured in the direction of dominant winds, and that is proposed location for measuring point.Daily limit for concentration of nitrogen dioxide can be exceeded, and all small towns that are in the range of 3 km can be exposed to that increased pollution. All these locations are potential measuring points in the future functional network, in order to create the real view on the state of the environment. Monitoring should take place in several stages: data collection, transmission of measured values using the network, data processing, decision making, alarm activation and taking certain measures. Among all the phases there should be communication, thus the system should be interoperable.

References

[1] J. Malenović - Nikolić, *Indicators of sustainable thermal energy systems based on the coal surface mines*, Master thesis, University of Niš, Faculty of Occupational Safety, Niš, 2009. (in Serbian)

[2] Regulation on limit values, methods of measurement and emissions, establishing criteria for measuring points and data collecting, Official Gazette of RS, no. 30/99, 1999. (in Serbian)

[3] M.R. Beychok, *Fundamentals of Stack Gas Dispersion*, 4th ed., Author-published, pp. 124, 2005.

[4] "Report on periodical measuring of harmful and dangerous substances in the air near TE Kostolac - B, Block 1", Technical report, Mining Institute, Belgrade, 2008. (in Serbian)

[5] "Report on periodical measuring of harmful and dangerous substances in the air near TE Kostolac - B, Block 1", Technical report, Mining Institute, Belgrade, 2009. (in Serbian)

[6] "Report on periodical measuring of harmful and dangerous substances in the air near TE Kostolac - B, Block 1", Technical report, Mining Institute, Belgrade, 2010. (in Serbian)

[7] C.H. Bosanquet and J.L. Pearson, „The spread of smoke and gases from chimney", Trans. Faraday Soc., 32:1249, 1936.

[8] A. Afanasev, C. Fomin, *Monitoring and methods for environmental control*, MNEPU, Moscow, 1998. (in Russian)

[9] Law on Air Protection, Official Gazette of RS, 36/09, 2009. (in Serbian)

Optical Flow Based Algorithm for Vehicle Counting

Nikola Dojčinović and Jugoslav Joković

Abstract **–General approach for vehicle counting based of videos of traffic flow is presented and discussed. Presented approach provides a framework for different vehicle counting algorithms. Particular implementation of algorithm is provided for conventional traffic monitoring systems. Algorithm reduces number of calculations, neglecting orientation information of moving vehicles, without performing vehicle tracking. Decision criteria is presented for vehicle extraction ensuring reliable vehicle counting in noisy videos.**

Keywords – **Vehicle counting, optical flow, motion detection**

I. Introduction

Traffic congestion is a serious issue confronting many urban centers. This issue is being addressed traditionally by increasing the supply of the roads. Due prohibited costs, such a solution is being abandoned. Instead, contemporary solutions focus on optimizing the throughput of existing roads. Therefore, methods for gathering real-time information about traffic flow are key. First solutions were based on inductive-loop detectors [1] buried underneath roads. Such detectors could only count vehicles travelling over them, without additional information about type and direction. Installation and maintenance of inductive-loop based systems is very expensive and requires traffic stopping.

More recently traffic monitoring systems are based on more promising camera networks. Less disruptive and less costly, camera networks based monitoring systems are very easy to install and to maintain. Key advantage of camera network based over inductive-loop based systems are additional information about traffic flow that camera network system can provide.

One of primary goals of traffic monitoring systems is classification of traffic patterns based on information about traffic flow. Therefore, camera network system must provide enough information about traffic flow. There are two main set of methods for extraction of traffic information from optical information gathered from camera network.

First set of methods directly measures a dynamics aggregated over regions of image in time. Traffic scenes are treated as instances of a dynamic texture [2], i.e. as spatiotemporal image patterns best characterized in terms of the aggregate dynamics of a set of constituent elements, rather than in terms of the individuals.

Another set of methods is based on a combination of segmentation and tracking. Individual objects are being detected and analysed. The general procedure ([3]-[13]) is consisted from the following three steps: (i) motion detection, (ii) tracking of detected vehicles and (iii) combining trajectory information to derive overall description of traffic flow. There are two major problems associated with this approach[14]: (i) segmentation problems due varying environmental conditions (e.g. lightning, strong wind), occlusions and low resolution imagery resulting in small pixel support of vehicle targets and (ii) tracking issues related to correspondence problems and occlusions.

In this paper solution for indirect traffic frequency calculation through temporal vehicle counting is presented. General algorithm for vehicle counting is proposed. Specific contribution of this paper is new method for vehicle counting based on optical flow estimation. New decision criteria for object determination is presented. This method involves no tracking of individual vehicle, and therefore number of calculations is significantly reduced.

II. General Description of Vehicle Counting Algorithm

Algorithms for vehicle counting are oriented toward analysis of individual objects extracted from videos of traffic flow. If whole scene is analysed, without of objects extraction, we could only estimate, but could not get exact number of vehicle on the road. One property of vehicle that uniquely distinguishes vehicles on roadway scenes is their motion. All non-moving objects on the scene can be classified as non-vehicle. In practice, there are non-moving vehicles on the roadway (e.g. stopped or parked vehicle), but they have no significance for traffic frequency estimation because they have no direct influence on traffic congestion.

Motion of the vehicles in the scene is best described with motion vector, because it contains information both of direction and displacement of the vehicles. In literature, there are several methods for motion vector estimation between consecutive frames. Block matching set of methods are based on simple and straightforward, and yet very efficient, algorithm presented in [16]. Fundamental idea of block matching is to find non-overlapped, equally spaced, fixed size, small rectangular blocks of image that match best by some predefined matching criteria. Methods for motion vector estimation from block matching set defer one to another mostly in procedure for searching of best matching blocks. Various new methods are proposed, among them coarse fine three-step search [17], conjugate direction search [18]. Some of them are using multi-resolution block matching [19] or subsampling [20]. Although there are significant limitations of block matching [21], it has been by the most popularly utilized motion estimation technique in video coding and it has been adopted by major video coding standards (ISO MPEG-1 and MPEG-2 [22], ITU H.261, H.263 and H.264 [23]).

Optical flow methods provide relatively more accurate motion vector estimation than block matching. Starting point of optical flow estimation is modelling of typical motion recording system. 2D representation of 3D scene over parametric transformation is used. This model is valid only if difference in depth caused by motion is small relatively to the distance of object from camera objective. Another prerequisite

Authors are with Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dojca@elfak.rs , jugoslav.jokovic@elfak.ni.ac.rs

for good modelling is constant scale of object in the scene. Change of object dimensions (object re-scaling) is detected as motion, e.g. object dimensions increasing are detected as depth reduction, moving toward camera, and vice versa, object dimension decreasing is detected as moving from the camera. Let I(x, y, t) be intensity level of pixel in frame t, positioned in x-th row and y-th column of image matrix. If displacement induced in pixel (x, y) is presented as pair (p(x, y, t), q(x, y, t)), assumption of grey level constancy between two frames :

$$I(x + p(x,y,t), y + p(x,y,t), t + 1) = I(x,y,t) \quad (1)$$

If presumption that movement is small relatively to dimension of the frame is used, it is considered infinitesimal, so its valid expanding of left side of equation (1) to its first order Taylor expansion around (x, y, t) and neglecting all nonlinear terms yields:

$$I(x + p, y + q, t + 1) = I(x,y,t) + pI_x + qI_y + I_t \quad (2)$$

where

$$I_x = \frac{dI(x,y,t)}{dx}, \quad I_y = \frac{dI(x,y,t)}{dy}, \quad I_t = \frac{dI(x,y,t)}{dt}, \quad (3)$$

$$p = p(x,y,t), \quad q = q(x,y,t)$$

Equations (1) and (2) yields the well-known Horn-Schunck constant [24]:

$$pI_x + qI_y + I_t = 0. \quad (4)$$

We look for motion (p , q) which minimizes the error function at frame t in the region of analysis R:

$$Err^{(t)}(p,q) = \sum_{(x,y)\in R}\left(pI_x + qI_y + I_t\right)^2 \quad (5)$$

Now we perform the error minimization over the parameters of one of the following motion models [25]:

1. **Translation model**: Motion is described with 2 parameters, *p(x, y, t) = a, q(x, y, t)=d*, where is assumed that the entire object have a single translation. First order derivate minimization is performed by setting $Err^{(t)}(p, q)$ derivatives with respect to a and d to 0. We get two linear equation with two unknowns, *a* and *d*. In related papers [26] and [27], every small window is assumed to have a single translation.
2. **Affine model:** Motion is described with 6 parameters, *p(x, y, t) = a + bx +cy, q(x, y, t)=d + ex + fy*. Using derivation of $Err^{(t)}(p, q)$ with respect to the motion parameters and setting to zero yields six linear equations with six unknowns, *a, b, c, d, e, f* [26][28]
3. **Model of planar surface moving (a pseudo projective transformation):** Motion is described with 8 parameters [26][29], $p(x, y, t) = a + bx +cy + gx^2 +hxy$, $q(x, y, t)=d + ex +fy +gxy + hy^2$. In this case we have eight linear equations with unknowns, *a, b, c, d, e, f, g* and *h*.

Depending of motion model and minimization, optical flow estimation algorithms differ in number of calculations and accuracy of estimation. Therefore, choosing a right motion model is of great significance. If it is possible to predict which motion model will suit particular motion, it is possible to significantly reduce number of calculations, without estimation accuracy loss. Model of planar surface is most general model and will give accurate results for every motion. But, in cases when motion can be described accurate enough with translation model, it is not needed to use more complex ones. It requires much more calculation to solve system of eight linear equations with 8 unknowns (Model of planar surface) than system of two linear equations with two unknowns (Translation model).

If complex motions are analysed, like 3D motions or translation movement captured with moving camera, it is more appropriate to use complex motion models. In literature, there are frameworks for motion vector estimation for all complex model of movement. In [26] is described a hierarchical framework for the computation of motion information. Framework consists from “global model that constrains the overall structure of the motion estimated, a local model that is used in the estimation process, and a coarse-fine refinement strategy”, with application to specific examples.

Number of calculation needed for optical flow estimation directly affects to algorithm temporal properties. For whole algorithm it is essential to count vehicle in real time. This means that optical flow must be estimated in less than time between two frames, so post-processing and analysis can be done.

As intermediate result, matrix of image blocks (pixels) which represent inter-frame calculated movement vectors occurs. Movement vectors are result of object movement in the scene. However, some of movement vector are false positive. Their origin from moving objects which are not vehicles (birds, trees, etc.) and/or noise. To save further computations, false positives are removed with adaptive threshold and filtering.

As early stated, motion vectors are used for extraction of objects from background, and it is a common practice [22][23]. Object is considered as group of neighbourhood image blocks (pixels) with common motion vector property. Depending on particular application, different motion properties are used for object determination. In following section one example of determination property selection will be shown.

Detected objects are than tracked trough the scene and counted. Simple scenes with one roadway do not require tracking of vehicle for counting. In complex scenes, like crossroads, roundabout or roadway intersections, vehicles has to be tracked in order not to be counted more times than one. Tracking algorithms are beyond the scope of this paper.

III. Implementation of Vehicle Counting Algorithm

Regarding to exposure in previous section, general model of system for implementation of vehicle counting algorithm is presented (fig 1). Crocked blocks are optional. After video acquisition, pre-processing is used to optimize video for motion detection (stabilization, adjusting, reconstruction, etc). Role of post-processing block is to filter out small artefacts from false positive motions. Object tracking is, also, optional, depending on previous information about vehicle movement orientation or nature of the scene. Particular implementation is optimized for vehicles counting on straight road. Presumption is that traffic flow is oriented from top of the scene to scene bottom, or vice versa. Cause for this presumption is common positioning of cameras in traffic monitoring systems, above the road (fig 2).

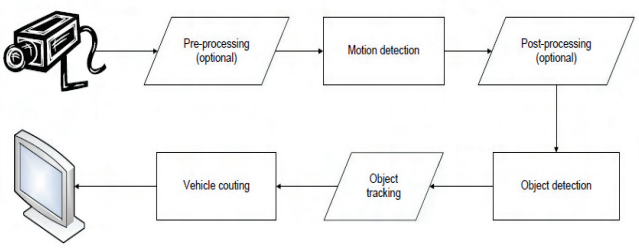


Fig. 1. Scheme of vehicle counting system

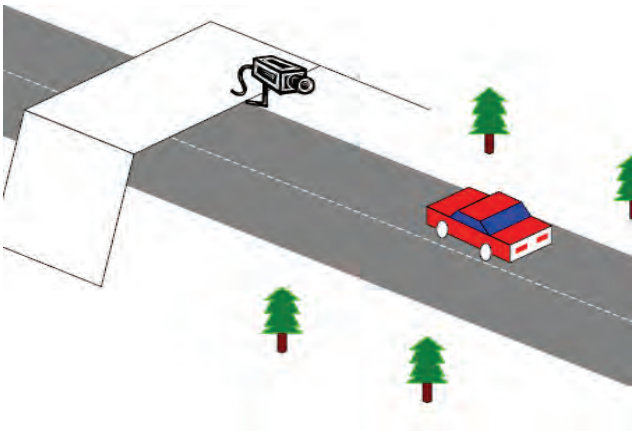


Fig. 2: Common camera positioning

By choosing appropriate motion vector estimation method, curtain facts must be considered. Minimizing number of calculations, motion vector are estimated over smallest possible regions, both in spatial and temporal domain. However, small image blocks carry little motion information, and therefore such motion estimations are inaccurate. Increasing of temporal region to two or more frames improves accuracy of calculated motion vectors. Simple increasing of spatial region can include more than one motion vector in single region. Vehicles have only one motion vector, so we choose spatial region not to contain two vehicles. In other words, region of motion vector estimation must be smaller than smallest vehicle in video.

Fact that vehicle has only one motion vector implies usage of simplest motion model (translation model). In order to reduce number of calculations, Horn&Schunk method is chosen[24].

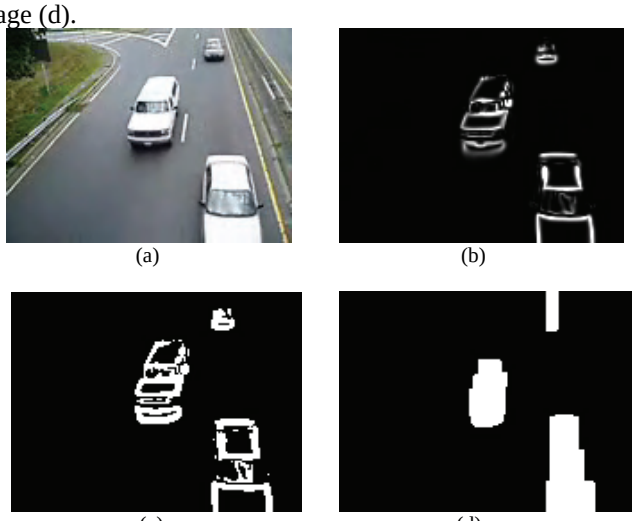
In post-processing, matrix of amplitudes of motion vectors is calculated. Considering presumed orientation of vehicles, only information of motion vector intensity is used for segmentation. Threshold is calculated as spatial mean square value in current frame

$$T_{msv} = k\sqrt{\frac{\sum_{(x,y)\in t} I(x,y)^2}{numel(t)}} \quad (6)$$

in frame *t*, where *numel(t)* in number of blocks (pixels) in frame t and k scaling parameter. Segmentation over threshold calculated this way is subject to bad segmentation in frames where there are no vehicles in the scene. Absolutely small movements, but relatively greater than threshold, will be false segmented. Therefore, it is needed to insert some inertia in threshold calculation, by applying temporal threshold calculation, over past few frames. Let $T_{msv}(t)$ denotes threshold calculated in frame *t*, and let *n* be number of past frames over which calculation is done. Spatiotemporal threshold can be, then, calculated as:

$$T_{temp} = \sqrt{\frac{\sum_{i=0}^{n} T_{msv}(t-i)^2}{n}} \quad (7)$$

In Figure 3 are presented 40th frame on sequence "viptraffic.avi" (a)(standard Matlab sequence), motion vectors image (b), segmented image with spatiotemporal threshold calculated over 3 consecutive frames (c) and post-processed image (d).



(a) (b) (c) (d)

Fig. 3: Intermediate representations of 40th frame of original sequence "viptraffic.avi"(a), motion vector representation (b), segmented image (c) and post-processed image (d).

After segmentation, intermediate results are post-processed in order to remove false positive motions and to fill in space within closed contours (fig 3.c). Amount of information infects accuracy of motion vector estimation. As much information are contained in region, estimation is more accurate. High frequency details, like edges, contain information about contour of the objects, and therefore their motion vectors are estimated more accurate. Motion vectors of uniformly distributed regions are estimated less accurate, so that in segmented image object contours are presented. Performing operations of morphological erosion and closing, small artifact are filtered out and contours are filled in (fig 3.d).

In bimodal image (fig 3.d) objects are detected, as groups of neighboring positively segmented blocks (pixels). Not all detected objects are vehicles. Performing post-processing as described in previous section, small artifacts are filtered out, but procedure will have opposite impacts on large false detected blocks. This blocks are filled in, like vehicles contours (fig 4). What differs them from vehicle objects is their spatial distribution. While vehicle objects are homogeneous, consistent, non-vehicle objects are jagged, spatially displaced from centers (fig 4). Therefore, decision criteria based on object homogeneity is employed for object classification on vehicle objects and non-vehicle objects.

For every object unique property of homogeneity, h_c, is calculated as ratio of object area and area of square around the object (8). If coordinates of up, left corner of the square around the object are (x_1, y_1) and down, right coordinates are (x_2, y_2)

Fig. 4: Red square non-vehicle object and vehicle object right from it

$$h_c = \frac{area(object)}{abs[(x_2 - x_1)(y_2 - y_1)]} \quad (8)$$

with maximum value 1 (squared objects).

If h_c is greater than ratio threshold, object is considered as vehicle. Vehicles are counted over counting area (fig 5). Area is defined with start and stop line. Orientation is normal to orientation of the traffic flow and width is set to be greater than greatest expected object displacement, so that no vehicle can pass through counting area uncounted.

(a) (b)

Fig. 5: original frame (a), counting area with counter of number of vehicles currently present in it (b)

IV. Conclusion

Vehicle counting based of video information of traffic flow happens to be more accurate, less expensive and easily applicable, than traditional inductive-loop based. Proposed algorithm, provided for commonly used, standard traffic monitoring systems, counts vehicles in real time, neglecting motion orientation, and therefore significantly reduces number of calculations. In case of noisy videos, decision criteria ensures reliable vehicle counting and enables vehicle shape information. Further calibration can enable vehicle velocity estimation. Although it is not a optimal solution, due great cost of computation, this algorithm has provided a solid framework for further development.

References

[1] Traffic Detector Handbook, page 338. Institute Transportation Engineers, 1990.

[2] D. Chetverikov and R. Peteri. A brief survey of dynamic texture description and recognition. In CORES, pages 17–26, 2005.

[3] D. Koller, K. Daniilidis, and H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. IJCV, 10(3):257–281, 1993.

[4] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell. Towards robust automatic traffic scene analysis in real-time. In ICPR, pages I:126–131, 1994.

[5] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real time computer vision system for measuring traffic parameters. In *CVPR*, pages 495–501, 1997.

[6] A. Lipton, H. Fujiyoshi, and R. Patil. Moving target classification and tracking from real time video. In WACV, pages 8–14, 1998.

[7] T. Tan, G. Sullivan, and K. Baker. Model-based localization and recognition of road vehicles. IJCV, 27(1):5–25, 1998.

[8] R. Cucchiara, M. Piccardi, and P. Mello. Image analysis and rule-based reasoning for a traffic monitoring system. Trans. ITS, 1(2):119–130, 2000.

[9] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi. Traffic monitoring and accident detection at intersections. Trans. ITS, 1(2):108–118, 2000

[10] Y. Jung, K. Lee, and Y. Ho. Content-based event retrieval using semantic scene interpretation for automated traffic surveillance. Trans. ITS, 2(3):151–163, 2001

[11] B. Maurin, O. Masoud, and N. Papanikolopoulos. Monitoring crowded traffic scenes. In ICITS, pages 19–24, 2002

[12] D. Magee. Tracking multiple vehicles using foreground, background and motion models. IVC, 22(2):143–155, 2004

[13] A. Cavallaro, O. Steiger, and T. Ebrahimi. Tracking video objects in cluttered background. CirSysVideo, 15(4):575–584, 2005

[14] K.G. Derpanis and R.P.Wildes. Classification of Traffic Video Based on a Spatiotemporal Orientation Analysis

[15] Yun Q. Shi and Huifang Sun. Image and Video Compression for Multimedia Engineering, 2nd ed, CRC Press, 2008

[16] J.R. Jain and A.K. Jain, Displacement measurement and its application in intraframe image coding, *IEEE Transaction on Communications*, COM-29, 12, 1799-1808, December 1981

[17] T. Koga, K. Linduma, A. Hirano, Y. Ilijima and T. Ishiguro, Motion compensated interframe coding for video conferencing, Proceeding of NTC'81, New Orleans, LA, pp. G5.3.1-G5.3.5, December 1981.

[18] R. Srinivasan and K.R.Rao, Predictive coding based on efficient motion estimation, Proceedings of ICC, pp. 521-526, May 1984

[19] D. Tzovras, M.G. Strintzis and H. Sahinoulou, Evaluation of multiresolution block matching techniques for motion and disparity estimation, Signal Processing: Image Communication, 6, 56-67, 1994

[20] M. Bierling, Displacement estimation by hierarchical block matching , Proceeding of Visual Communication and Image Processing, SPIE 1001, pp. 942-951, 1988.

[21] S. Lin, Y.Q. Shi and Y.Q. Zang, An optical flow based motion compensation algorithm for very low bitrate video coding, Proceeding of 1997 IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, Germany, pp.2869-2872

[22] www.iso.org

[23] www.itu.int

[24] B.K.P. Horn and B.G. Schunck, Determining optical flow, Artificial Intelligence, 17:185-203, 1981

[25] M. Irani, B. Rousso and S. Peleg, Computing Occluding and Transparent Motions, International Journal of Computer Vision, 1994.

[26] J.R. Bergen, P. Anandan, K.J. Hanna and R. Hingorani, Hierarchical model-based motion estimation, In European Conference on Computer Vision, pp. 237-253, Santa Margarita Ligura, May 1992.

[27] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In Image Understanding Workshop, pp. 121-130, 1981.

[28] J.R. Bergen, P.J. Burt, R. Hingorani, P. Jeanne and S. Peleg, Dynamic multiple-motion computation. In Y.A. Feldman and A. Bruckstein, editors, Artificial Intelligence and Computer Vision: Proceeding of the Israeli Conference, pp. 147-156, Elsevier, 1991.

Uncertainty Assessment of Electric Probe in Electromagnetic Field Monitoring System

Mirjana Trobok[1], Nikola Djuric[2]

Abstract –**Development of the electromagnetic field monitoring and control systems represents one of the major innovations in the range of methodologies for evaluating, through the measurements, the so-called environmental electromagnetic pollution. This phenomenon have caused the alerting of the public and the agencies for non-ionizing radiation protection, and therefore there is a need for constant and accurate surveillance of electromagnetic fields. This paper presents initial consideration of uncertainty assessment of theelectric probe, which used in the electromagnetic pollution monitoring system. This system is based on the wireless sensor networkand our team develops it within the program of technological development of the Republicof Serbia, for period of 2011–2014.**

Keywords – **EM pollution, monitoring network, uncertainty.**

I. INTRODUCTION

The fast growing penetration of radiofrequency and microwave radiating devices into everyday and occupational life of population emerges the public theme of the so-called electromagnetic pollution of the environment. Electromagnetic pollution has become sensitive and highly important scientific and research subject at an international level requiring analysisof the electromagnetic radiation on biological systems[1]-[2].

As support for those research efforts various methodologies for evaluating, through the measurements, have been considered. In addition, several agencies and standardization bodies [4]-[7]established guidelines for limiting the electromagnetic fields (EMFs) exposure that provide protection against known adverse health effects[8].

In this paper we consider a monitoring system based on a wireless sensor network [9]-[11], for automated, remote and selective monitoring of the overall level of EMFs. The proposed system collects measurement data and compares them with national prescribed limits on the daily basis[12]-[19].Moreover,the results of measurements are instantly available through Internet, providing information to the relevant institutions in the area of environmental protection against the electromagnetic pollution [12], [20], caused by a number of sources of EMFs.

The proposed system is an advanced solution available to meet the growing demands for monitoring the EMFs and continuous informing about the EMF distribution in the areas connected with human activities and their exposure to EMF radiation. The system has been supportedby Ministry of Sciences and Technological Developments of the Republicof Serbia[21]in period of 2011–2014.

This system is intended to be used in the various aspects, especially for EMF exposure evaluation in the zones of increased sensitivity, that are defined in legaldocument"*Rules for non-ionizing radiation sources of interest, types of sources, the manner and period of their investigation*" [16], as shown in Fig. 1.

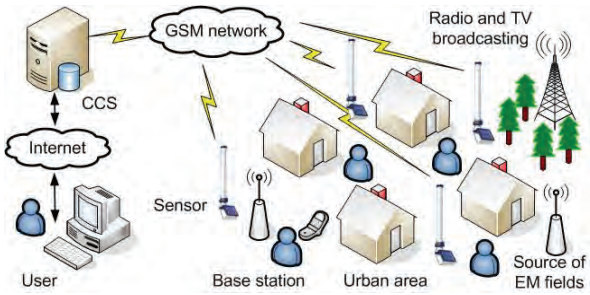


Fig. 1 Example of monitoring system utilization

The zones of the increased sensitivity are residential areas where people can spend 24 hours a day, i.e. schools, homes, pre-schools, maternity hospitals, hospitals, tourist facilities, play-grounds and areas of un-built land intended for specified purposes [16].

The monitoring system will perform measurements in real environment, containing the fixed and movable reflective structure and will be exposed to different weather conditions, as show in Fig. 2.

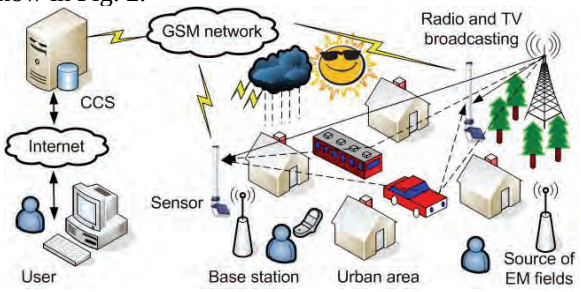


Fig. 2 Example of the area conditions

Generally, the measurements have to meetseveral requirements proposed by specific guidelines or standards[22]. On the other hand, performing non-ionizing exposureassessment using real measurement data requires consideration of the uncertainty of measurement[23].

In this paper, a methodology forestimating the overall uncertainty ofthe electric field probe in the monitoring system is proposed. Estimating and reporting measurement uncertainty are of great importance,especially when the measured values are very close to the established limits of human exposure to non-ionizingEMFs.

[1]Mirjana Trobok is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad, E-mail: trobok@uns.ac.rs

[2]Nikola Djuric is with the Faculty of Technical Sciences, University of Novi Sad, Trg D. Obradovica 6, 21000 Novi Sad,E-mail: ndjuric@uns.ac.rs

II. GENERALLY ABOUT UNCERTAINTY

Metrology suggests that each measurement result should be stated with accompanied uncertainty. Only than result is complete and acceptable, and can be compared with other results of same quantity or with reference values.

Uncertainty of measurement represents quantitative indication of the quality of measurement result. It defines range of values that could be attributed to the measured quantity.

General approach for uncertainty evaluation are imposed by *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (GUM) [24], where two types of uncertainty assessment are introduced

- Type A – determined using statistical methods on a series of repeated readings and
- Type B – determined using non-statistical methods, i.e. information from past experience of the measurements, from calibration certificates, from manufacturer specifications, from calculations, from published information, etc.

In order to evaluate measurement uncertainty, all the possible contributors of uncertainty have to be considered[25]. It has to be pointed out that uncertainty is a result both of incomplete knowledge of the value of the measurand and of the factors influencing it [26].

Some of the possible sources of uncertainty are:

- The uncertainty of the equipment used for calibration of a specific measuring instrument,
- Uncertainties resulting from measuring equipment determined with accredited calibration procedure, together with any drift or instability in their values,
- The uncertainties due to the instabilities of the measuring equipment during the measurement,
- The measuring procedure followed to estimate the measured quantity,
- Differences due to different staff carrying out the same type of measurement,
- The effects of environmental conditions (i.e. temperature, humidity) in the measurement setup [27].

GUM assumes that the uncertainty of a measurement result can be evaluated based on a mathematical model of the measurement, which describes afunctional relationship between the measurand *y*and the input (influence) quantities x_i.If the input quantities are designated as x_1, x_2, … x_n, then the mathematical modelcan be written as

$$y = f(x_1, x_2, ..., x_n). \quad (1)$$

In some cases, this relation is complicated and it is not possible to explicitly write it down.Besides, a functional relationship *f*can be determined experimentally or exist only as a numerically evaluated algorithm.

Contribution of all the possible sources of uncertainty have to be expressed in terms of standard uncertainties $u(x_i)$ (standard deviation) based on the associated probability distributions (extended consideration on calculating uncertainties depending on respective distribution is presented in [24]).

Moreover, uncertainty must be expressed in the same units as measurand before they are combined. It is achieved by multiplying standard uncertainty by sensitivity coefficients, denoted asc_i. Mathematically, sensitivity coefficients are obtained from partial derivatives of the model function *f* with respect to the input quantities[24], or they can be estimated experimentally. Sensitivity coefficients describe how change in the estimation of input quantities influences the estimate of measurand[27].

If all input quantities areindependent combined uncertainty is calculated according to

$$u(y) = \sqrt{\sum c_i^2 u(x_i)^2} \quad (2)$$

Expanded uncertainty, usually denoted as*U*,is the combined uncertainty multiplied by a coverage factor*k*. A particular value of coverage factor gives a particular confidence level for the expanded uncertainty.

Commonly, the overall uncertainty is scaled by the coverage factor $k = 2$, giving a level of confidence of approximately 95 %. ($k = 2$ is correct if the combined standard uncertainty is normally distributed, which is usually a fair assumption.The detail reasoning behind this is explained in [24]).

III. UNCERTAINTY OF THE PROBE

The proposed monitoring system is designed to investigate the overall level of the EM field and population exposure to the EM field at the particular location on the daily basis. In order to determine electric field strength,the electric field quad-band probe is used[28]-[29].

Calibration certificate data and manufacturer specification of field probe are stated in Table I.

TABLE I
MANUFACTURER SPECIFICATION AND CALIBRATION DATA OF PROBE

Quantity	Wide Band 0.1 to 3000 MHz	EGSM 900, 1800 and UMTS Bandpass
Meas.Range	(0.2 to 200) V/m	(0.03 to 30) V/m
Meas. Resolution	0.01 V/m	
Flatness @ 6 V/m	(1 - 200 MHz) ±0.8 dB (150 kHz - 3 GHz) ±1.5 dB	+0.5/-2.5 dB
Anisotropy	±0.8 dB (typical ±0.6 dB) @ 50 MHz, 3 V/m	±0.8 dB (typical ±0.6 dB)
Temperature Error	0 – 50ºC = ±0.3 dB -20 – 0ºC = -0.1 dB/ ºC	
Calibration Uncertainty	1 dB	1 dB

In case of EMFs measurement, the employed instrumentation and the measurement technique, as well as environmental conditions, contribute to measurement uncertainty. To evaluate uncertainty ofelectric field strength measurement, only standard uncertainty type B is taken into account. In order to determine standard uncertainty of type A, series of measurements should be provided. As our system is still in the developing phase, we are not able to give concrete results of measurement and to calculate the type A of uncertainty, therefore it will be omitted. The most important uncertainty sources,

TABLE II
UNCERTAINTY BUDGET OF MEASUREMENTS OF ELECTRIC FIELD STRENGTH MEASURED IN EGSM 900, EGSM 1800 AND UMTS BANDPASS MODE USING ELECTRIC PROBE

No		Source of uncertainty	Partial uncertainty [dB]	Type	Probability distribution	Divisor	Sensitivity coefficients	Standard uncertainty [%]
1.	Measurement Equipment	Calibration	1	B	normal	2	1	5.93
2.		Resolution of measurement	1.34	B	rectangular	1.73	1	9.32
3.		Flatness	2.5	B	rectangular	1.73	1	18.10
4.		Anisotropy	0.8	B	rectangular	1.73	1	5.47
5.		Temperature variation	0.3	B	rectangular	1.73	1	2.02
6.	Meas. Method	Antenna position in the field with high spatial variation	0	B	rectangular	1.73	1	0
7.		Spatial averaging	-	B	rectangular	1.73	1	-
8.	Environmental Parameters	Perturbation by the environment	1.5	B	rectangular	1.73	1	10.5
9.		Variations in the emitted power sources	1	B	rectangular	1.73	1	6.88
10.		Reflection of the major mobile sources close to sources of radiation	0	B	rectangular	1.73	1	0
Combined standard uncertainty [%]								25.32
Coverage factor								2
Expanded uncertainty [%]								50.64

their type and probability distributions are shown in Table II.

As all units of standard uncertainty are expressed in terms of the measurand and the functional relationship between the input quantities and measurand is given as linear summation, then all sensitivity coefficients are unity value (c_i=1), otherwise partial derivatives should be calculated [30].

Since all input contributors can be considered unrelated, the combined uncertainty is obtained as summation in squares of the individual uncertainty contributors.

Finally, the expanded uncertainty is obtained by multiplying the combined uncertainty by coverage factor k=2 which corresponds to the confidence level of 95 % as recommended by standard [25].

In this paper the worst case scenario is considered, therefore the maximum errors are taken into account when estimating uncertainty, using the parameters for EGSM 900, EGSM 1800 and UMTS bandpass mode.

As it can be seen from Table IIthe relative standard uncertainty due to resolution is significant. This value is calculated atelectric field strength of 0.03 V/m, at the detection threshold of the probe. At upper measurement range limit, 30 V/m, relative standard uncertainty due to resolution is less than 0.01 %. Therefore, knowing the exact measuredvalue of electrical field strength, the uncertainty contribution due to resolution will be significantly reduced,so that it can even be neglected.

The relative standard uncertainty due to flatness is dominant. When calculating uncertainty due to temperature variations, it is assumed that temperature is higher than 0 ºCin a real environment, therefore temperature error of ±0.3 dB is used.

Considering the measurements method, in this initial investigation the antenna position in the field with high spatial variation was omitted as a sources of uncertainty. This was done because in some first implementations the sensors will be positioned in a same plane. In addition, for the same reason the spatial averaging was not taken into account.

These two parameters are accounted as a possible source of the uncertainty and will be considered in applications where sensors are place on different heights (for example in case of monitoring electric field from the base station for mobile communication [11], [31]).

Furthermore, considering the environmental parameters as sources of uncertainty, reflection of the major mobile sources close to the sources of radiation, was omitted as influential source because we assumed that our system will be implemented in area where reflection can be neglected. We are fully aware of this assumption, but without precise description of the site where sensors are positioned, this source can not be properly estimated.

IV. CONCLUSION

This paper presents an information network, as support for automatic and continuous monitoring of the overall level of EMFs. To ensure the EMF measurement validity, it has to be conducted according to the standards and uncertainty estimation has to be associated.This is important for data interpretation, especially when a compliance statement with a specification or a legal exposure limit is needed.

In order to determine uncertainty, all sources of uncertainty have to be identified and their contribution to overall uncertainty has to be estimated.

In this paper, a concept for estimating uncertainty, when EMF measurements are performed using electric field sensor, is presented.

In a process of evaluation of measurement uncertainty we consider different contributors divide in a three main categories:

- uncertainties which derive measuring equipment,
- method and
- environmental parameters.

The contribution of each source of uncertainty of measurement is registered by the name, probability distribution, and sensitivity coefficient and uncertainty value.

ACKNOWLEDGEMENT

This paper has been supported by Ministry of Sciences and Technological Development of the Republic of Serbia, under the grant for project TR 32055.

REFERENCES

[1] EU Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) – "Possible effects of Electromagnetic Fields (EMF) on Human Health", 2007, http://ec.europa.eu/-health/ph_risk/committees/04_scenihr/docs/scenihr_o_007.pdf,

[2] EU Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) – "Health and electromagnetic fields", 2009, http://ec.europa.eu/health/archive/ph_determinants/-environment/emf/brochure_en.pdf,

[3] International Commission on Non-Ionizing Radiation Protection (ICNIRP) – "Exposure to high frequency electromagnetic fields, biological effects and health consequences (100 kHz-300 GHz)", 2009, http://www.icnirp.de/documents/RFReview.pdf,

[4] World Health Organization – WHO – http://www.who.in,

[5] EU Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) – http://ec.europa.eu/health/scientific_committees/emerging/index_en.htm,

[6] International Commission on Non-Ionizing Radiation Protection (ICNIRP) – http://www.icnirp.de,

[7] European Committee for Electrotechnical Standardization (CENELEC) – http://www.cenelec.eu,

[8] International Commission on Non-Ionizing Radiation Protection (ICNIRP) – "Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz)", 1998, http://www.icnirp.de/documents/emfgdl.pdf,

[9] N. Djuric, "Razvoj informacione mreze za kontinualno i udaljeno isptivanje elektromagnetskih polja," submitted for the journal Telekomunikacije 2011,

[10] N. Djuric, M. Prsa, K. Kasas-Lazetic, "Information network for continuous electromagnetic fields monitoring," submitted for PES 2011.

[11] N. Djuric, M. Prsa, K. Kasas-Lazetic, "Serbian System for Remote Monitoring of Electromagnetic Fields," submitted for MPS 2011, Cluj-Napoca, Romania , 17-20 May 2011,

[12] Ministry of Environment and Spatial Planning of the Republic of Serbia – http://www.ekoplan.gov.rs/src/index.php.

[13] V. Bajovic, N. Djuric, D. Herceg, "Serbian laws and regulations as foundation for electromagnetic field monitoring information network," submitted for PES 2011.

[14] "Law on Non-Ionizing Radiation Protection", the law of Republic of Serbia, no. 36/09,

[15] "Regulation on the limits exposure of non-ionizing radiation", the law of Republic of Serbia, no. 104/09,

[16] "Rules for non-ionizing radiation sources of interest, types of sources, the manner and period of their investigation", the law of Republic of Serbia, No. 104/09,

[17] "Regulation for conditions to be met by institutions and companies that performs testing of radiation level", the law of Republic of Serbia, No. 104/09,

[18] "Regulation for conditions to be met by institutions and companies that performs systematic testing of non-ionizing radiation", the law of Republic of Serbia, No. 104/09,

[19] "Regulation on the contents and the form of reports of systematic testing of non-ionizing radiation in the environment", the law of Republic of Serbia, No. 104/09.

[20] Municipal Agency for the Environmental Protection – City of Novi Sad – http://www.environovisad.org.rs,

[21] Ministry of Sciences and Technological Developments - http://-www.nauka.gov.rs/,

[22] Serbian Accreditation Body – http://www.ats.org.rs,

[23] M. Trobok, N. Djuric, "Uncertainty assessment of sensor in information network for electromagnetic field monitoring," submitted for conference PES 2011,Nis, 2011,

[24] JCGM 100:2008, Evaluation of measurement data — Guide to the expression of uncertainty in measurement (GUM),

[25] EN 50413 – Basic standard on measurement and calculation procedures for human exposure to electric, magnetic and electromagnetic fields (0 Hz - 300 GHz),

[26] G104 - A2LA "Guide for Estimation of Measurement Uncertainty In Testing", July 2002.

[27] G. Basso, "Uncertainty in the measurement of electromagnetic field with isotropic broadband sensor and selective E&H field analyzer", www.narda-sts.com,

[28] M. Milutinov, N. Djuric, B. Vukobratovic, "Multi-band area monitor sensor in information network for electromagnetic fields monitoring," submitted for conference PES 2011,Nis,

[29] NARDA Safety Test Solutions, "Continuous, remotemonitoring and loggingof electromagnetic fields",www.narda-sts.com,

[30] D. Stratakis, A. Miaoudakis, C. Katsidis, V. Zacharopoulos, T. Xenos, "On the Uncertainty Estimation Of ElectromagneticField Measurements Using Field Sensors: A GeneralApproach", *Radiation Protection Dosimetry* (2009), Vol. 133, No. 4, pp. 240–247,

[31] V. Bajovic, N. Djuric, D. Herceg, "Serbian Program of Systematic Testing of Non-ionizing Radiation in the Environment," submitted for MPS 2011, Cluj-Napoca, Romania , 17-20 May 2011.

Session PO4:

PO4 – CONTROL SYSTEMS AND ROBOTICS

Algorithm for Modal Control of Dual-Mass Electromechanical System

Nikola Nikolov[1], Vitan Dimitrov[2], Mariela Alexandrova[3], Ivaylo Penev[4]

Abstract–**Created an algorithm for modal control of dual-mass electromechanical system which is composed of a DC motor, power electronic converter voltage and working machine. The created algorithm is based on discrete mathematical system description in state-space. It is provide fast and smooth acceleration of the working machine to set speed and current in the motor acceleration remains less than 2Inom.**

Keywords–**Algorithm for modal control, dual-mass electromechanical system, DC motor, power electronic converter, working machine, discrete systems, state space, MATLAB-implementation.**

I. INTRODUCTION

The structural scheme of dual-mass electromechanical system consisting of a power converter, a DC motor (at constant magnetic flux Φ=const) with moment of inertia J1 and working machine with an equivalent moment of inertia J2, is shown in Figure 1 [1, 4, 5].

Mathematical description of the structural scheme shown above can be obtained based on equations of the processes in the DC motor and dual-mass equations of mechanical part.

In this case, considering DC motors with parallel excitation, which has catalog data:

- ✓ $U_{nom}=220\ V;\ P_{nom}=0.3\ kW;$
- ✓ $n_{nom}=1000\ tr/min;\ I_{nom}=2\ A;$
- ✓ $r_a+r_p=16.6\ \Omega$ – *resistance of the armature windings;*
- ✓ $N=3384$ – *number of active conductors;*
- ✓ $2a=2$ – *number of parallel branches;*
- ✓ $P_p=1$ – *number of pairs of poles;*
- ✓ $\Phi.10^{-2}=0.31\ Wb;$
- ✓ $N_{max}=2000\ tr/min;$
- ✓ $J=J_1=0.042\ kgm^2;$
- ✓ $m=38.0\ kg.$

The variables introduced in Fig. 1 have the following values:

- ✓ $R_a=20.8828\ \Omega;$
- ✓ $T_a = \frac{L_a}{R_a} = 0{,}0126\ s$, **where**

$$L_a = \frac{r.U_{nom}}{P_{nom}.\omega_{nom}.I_{nom}} = 0{,}2626\ H\ ;$$

- ✓ $r=0.25\Omega;\ \omega_{nom} = \frac{2pn_{nom}}{60} = 104{,}7198\ rad/sec\ ;$
- ✓ $J_1 = J = 0{,}042\ kgm^2\ ;$
- ✓ $J_2 = 0{,}5J_1 = 0{,}021\ kgm^2\ ;$
- ✓ $cF = \frac{U_{nom} - R_a I_{nom}}{\omega_{nom}} = 1{,}7020\ Vs\ ;$
- ✓ $M_{nom} = \frac{P_{nom}}{\omega_{nom}} = 2{,}8648\ Nm\ ;$
- ✓ $c_{12} = 0{,}5M_{nom} = 1{,}4324\ Nm$ - *coefficient of hardness to elastic connection;*
- ✓ $M_{c1} = 0{,}1M_{nom} = 0{,}28648\ Nm\ ;$
- ✓ $M_{c2} = 0{,}9M_{nom} = 2.57832\ Nm\ .$

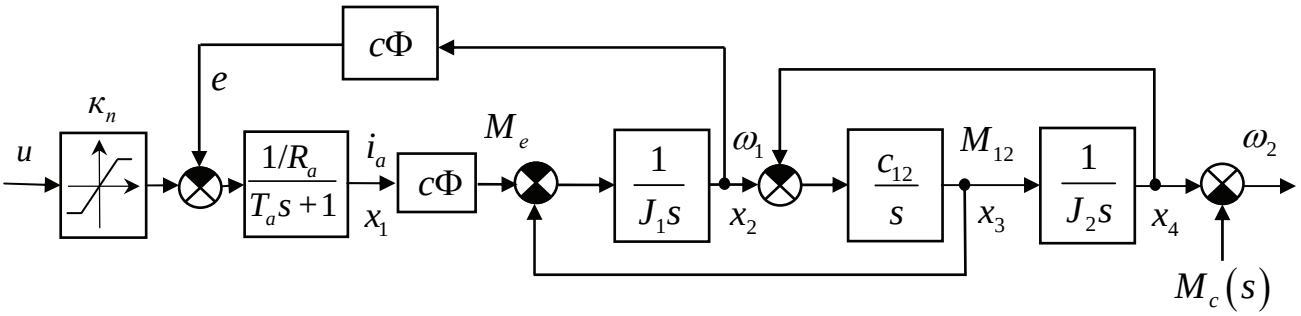


Fig. 1 Structural scheme of dual-mass electromechanicalsystem

The type of transition process of dual-mass electromechanical system is shown in Figure 2. It was simulating on the basis of the structural scheme of Figure 1.

[1]Nikola Nikolov, E-mail: nn_nikolov@abv.bg,
[2]Vitan Dimitrov, E-mail: vit_dim@yahoo.com,
[3]Mariela Alexandrova, E-mail: meri_alexandrova@mail.bg,
[4]Ivaylo Penev, E-mail: ivailopenev@yahoo.com,
are with the Faculty of Computer Science and Automation, in Technical University of Varna, Bulgaria.

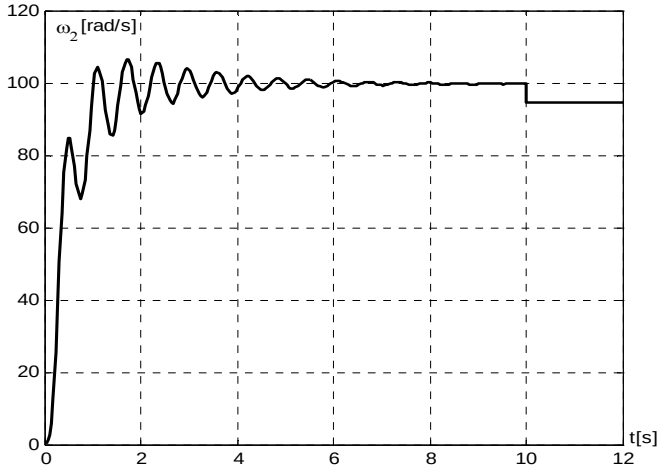


Fig. 2Transition process of dual-mass electromechanical system without modal control.

As a result of internal electromotive feedback, when input voltage is constant ($u = const$) the output speed is establishe. With increasing hardness of the elastic connection increases the frequency of fluctuations and reduces their amplitude. With increasing moment of inertia of the second mass fluctuations subside more slowly. When had appeared of disturbance (in the moment t = 10s) the speed of dual-mass system is lower.

II. Discrete Mathematical Model of Dual-mass Electromechanical System

The algorithm for modal control is going to synthesize based on discrete mathematical model of the dual-mass system in state-space. Therefore, it first is going to made mathematical description of the continuous system in state-space.

To be able the algorithm is using for practical purposes, elements of the state vector is going to chose real physical values: $x_1 = i_a$ - armature motor current; $x_2 = \omega_1$ - rotor speed the motor; $x_3 = M_{12}$ - mechanism elastic torque и $x_4 = \omega_2$ - speed of the working machine.

$$\left|\begin{array}{l} \dot{x}_1 = \dfrac{cF}{T_a R_a}\left(k_n u_{зад} - cF x_2\right) - \dfrac{1}{T_a} x_1; \\ \dot{x}_2 = \dfrac{1}{J_1}\left(x_1 - x_3 - M_{c1}\right); \\ \dot{x}_3 = c_{12}\left(x_2 - x_4\right); \\ \dot{x}_4 = \dfrac{1}{J_2}\left(x_3 - M_{c2}\right). \end{array}\right. \quad (1)$$

Based on (1) is obtained mathematical description of continuous dual-mass system in state-space:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t); \\ y(t) &= \mathbf{c}^{\mathrm{T}}\mathbf{x}(t) + du(t), \end{aligned} \quad (2)$$

and the individual matrices are the following

$$\mathbf{A} = \begin{bmatrix} -1/T_a & -cF/T_a R_a & 0 & 0 \\ cF/J_1 & 0 & -1/J_1 & 0 \\ 0 & c_{12} & 0 & -c_{12} \\ 0 & 0 & 1/J_2 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} -79{,}3651 & -6{,}468451 & 0 & 0 \\ 40{,}52381 & 0 & -23{,}8095 & 0 \\ 0 & 1{,}4324 & 0 & -1{,}4324 \\ 0 & 0 & 47{,}619 & 0 \end{bmatrix};$$

$$\mathbf{b} = \begin{bmatrix} k_n/R_a T_a \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 65{,}8437 \\ 0 \\ 0 \\ 0 \end{bmatrix};$$

$$\mathbf{c}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix};\ d = 0.$$

Based on the mathematical description of the continuous system (2) mathematical description is made of discrete dual-mass system in state-space for a sample time $T_0 = 0.1\ s$:

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{b}u(k); \\ y(k) &= \mathbf{c}^{\mathrm{T}}\mathbf{x}(k) + du(k), \end{aligned} \quad (3)$$

and the individual matrices are the following

$$\mathbf{A} = \begin{bmatrix} -0.0160 & -0.0331 & -0.1530 & -0.0016 \\ 0.2071 & 0.4235 & 1.7665 & -0.0146 \\ -0.0576 & -0.1063 & 0.1923 & 0.1036 \\ 0.0197 & -0.0291 & -3.4441 & 0.1777 \end{bmatrix};$$

$$\mathbf{b} = \begin{bmatrix} 0.3526 \\ 6.0168 \\ 0.0272 \\ 8.6667 \end{bmatrix};$$

$$\mathbf{c}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix};\ d = 0.$$

III. Algorithm for Modal Control of Dual-mass Electromechanical System

Algorithm which is presented below was developed based on algorithm for modal control in [3]. Modal control is realized by synthesizing the feedback vector **k**. The vector **k** is defined so that the poles of the closed-loop system to moved in a circle with specified radius η known as zone of stability.

Algorithm for modal control of dual-masssystem is shown in a structural form with the scheme of Figure 3 and is described step by step below.

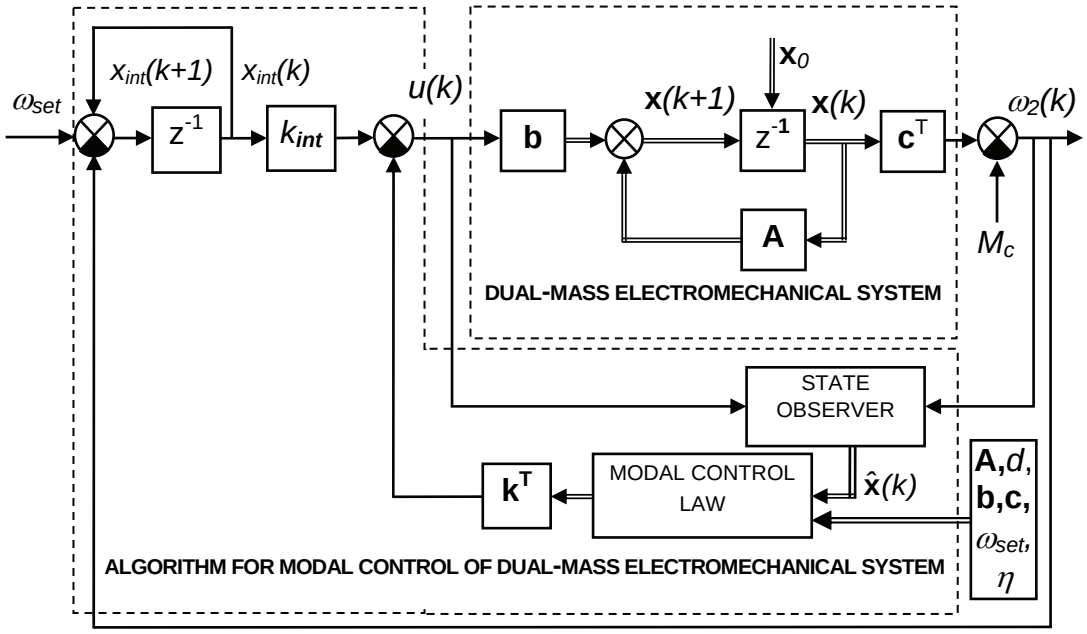


Fig. 3 Algorithm for modal control of dual-mass electromechanical system in a structural form.

Step 1 Introduction of: the matrix **A**, the vector **b**, the vector **c**, the scalar *d* and *zone of stability* η. Zone of stability specifies the radius of the circle in which to locate the poles of the closed system;

Step 2 Check order of the system-*n* through the length of the vector **b**;

Step 3 Determination the eigenvalues of the matrix **A** ($\lambda_1, \lambda_2, ..., \lambda_n$) and corresponding to them own vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$;

Step 4 Formation the matrix of own vectors **Q**

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}; \mathbf{q}_i = \begin{bmatrix} q_{i1} & q_{i2} & \cdots & q_{in} \end{bmatrix}; \quad i = 1,2,...,n;$$

Step 5 Generation of *n*-dimensional vector $\boldsymbol{\xi}$, of random numbers with normal distribution between *0* and *1*

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 & \xi_2 & \cdots & \xi_n \end{bmatrix}^T;$$

Step 6 Determination of vector eigenvalues of the closed system

$$\boldsymbol{\mu} = \begin{bmatrix} \eta \\ \eta \\ \vdots \\ \eta \end{bmatrix} - \left| \left(\begin{bmatrix} 0.5 \\ 0.5 \\ \vdots \\ 0.5 \end{bmatrix} - \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \right) 0.1 \right| = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix};$$

Step 7 Determination of the vector

$$\mathbf{g} = \mathbf{Q}^T\mathbf{b} = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \end{bmatrix}^T;$$

Step 8 Calculation of the elements of the vector **d**

$$d_i = \frac{\prod_{j=1}^{n} (\lambda_i - \mu_j)}{g_i \prod_{j=1, j\neq i}^{n} (\lambda_i - \lambda_j)}, \quad i = 1,2,...,n;$$

$$\mathbf{d} = \begin{bmatrix} d_1 & d_2 & \cdots & d_n \end{bmatrix}^T;$$

Step 9 Defining the vector of feedback

$$\mathbf{k}^T = \mathbf{d}^T\mathbf{Q} = \begin{bmatrix} k_1 & k_2 & \cdots & k_n \end{bmatrix};$$

Step 10 Determination of the innovation vector addition **f** and matrix of the state observer **F**

$$\mathbf{f} = \text{acker}(\mathbf{A}, \mathbf{c}, \mathbf{w}), \ \mathbf{w} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T;$$

$$\mathbf{F} = \mathbf{A}\text{-}\mathbf{f}\,\mathbf{c}^T;$$

Step 11 Estimation the current state vector $\hat{\mathbf{x}}$(k)

$$\hat{\mathbf{x}}(k+1) = \mathbf{F}\hat{\mathbf{x}}(k) + \mathbf{b}u(k) + \mathbf{f}y(k), \ \hat{\mathbf{x}}(0) = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T, \ k = 0,1,2,...;$$

Step 12 Calculation of the integral variable x_{int}

$$x_{int}(k+1) = x_{int}(k) + \omega_{set} - \omega_2(k), \ k = 0,1,2,...;$$

Step 13 Calculation of control voltage

$$u(k) = \begin{bmatrix} k_{int} & -\mathbf{k}^T \end{bmatrix} \begin{bmatrix} x_{int}(k) \\ \hat{\mathbf{x}}(k) \end{bmatrix}, \quad k = 0, 1,2,...;$$

Step 14 Repeating steps 11, 12 and 13 during the control process.

IV. MATLAB-IMPLEMENTATION OF ALGORITHM FOR MODAL CONTROL

```
clc,clear
%State Space Modal Control for DC Motor ***********
omega_set=100; stab_zone=0.27; k_int=0.02; Mc=2.57832;
takt_disturb=80; endprocess=120;
%************************************************
A=[-0.0297   -0.0544    0.1308   -0.0094;
    0.3408    0.6087   -1.7352    0.1446;
    0.0493    0.1044    0.5422   -0.1204;
    0.1180    0.2891    4.0035    0.6867];
b=[ 0.6496;   2.5117;   0.1632;   0.2463];
c=[ 0; 0; 0; 1]; d=0;
n=length(b);
[Aq,Az]=eig(A');
ERROR(1:n,:)=Aq(:,1:n).'*A-Az(1:n,1:n)*Aq(:,1:n).';
if abs(max(ERROR))<0.0001
   q(1:n,:)=(Aq(:,1:n)).'; z=diag(Az);
```

```
if max(abs(z))>stab_zone
      mu=(stab_zone-(0.5-randn(1,n))*0.1);
end
    p=[q(1:length(q(:,1)),:)*b];
   dm(1:length(z))=0;
    Pden=1;
    Pnum=1;
for i=1:length(z)
for j=1:length(z)
            Pnum=Pnum*((z(i)-mu(j)));
if i~=j
            Pden=Pden*((z(i)-z(j)));
end
end
      DEN=p(i)*Pden;
      dm(i)=-(Pnum/DEN);
      Pden=1;
        Pnum=1;
end
   k=dm*q;
else
   error('Discrepancy between eigenvalues and own
vectors')
end
if abs(imag(k))<0.001
   K=real(k);
end
w=zeros(n,1); f=(acker(A',c,w))'; F=A-f*c';
takt=1
x_int(takt)=0; xo(:,takt)=zeros(n,1);
xr(:,takt)=zeros(n,1); y(takt)=0; disturb=0;
key1=1;
while key1==1
   takt=takt+1
   x_int (takt)=x_int (takt-1)+y_point-y(takt-1);
   u(takt)=[ki K]*[x_int (takt); xo(:,takt-1)];
   xo(:,takt)=F*xo(:,takt-1)+b*u(takt)+f*y(takt-1);
   xr(:,takt)=A*xr(:,takt-1)+b*u(takt);
if takt>takt_disturb
       disturb=Mc;
end
   y(takt)=c'*xr(:,takt)-disturb
   I(takt)=xr(1,takt);
if takt>=endprocess
       key1=0;
end
```

Simulation study is done through the above shown implementations of MATLAB. Results of the study is shown in Figure 4.

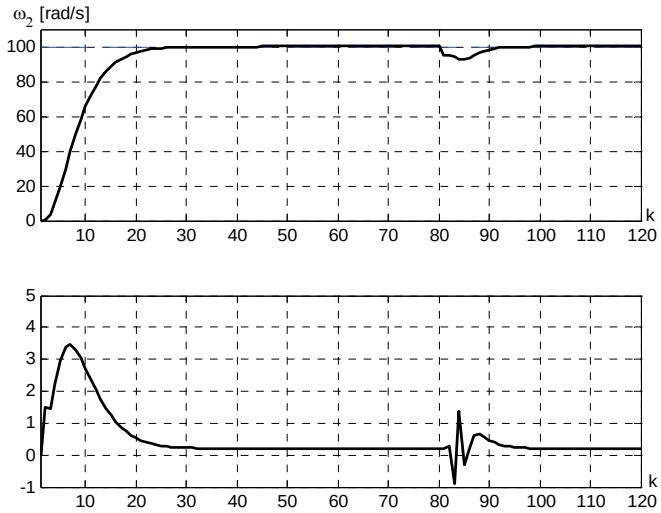


Fig. 4Results of the simulation study with MATLAB-implementation.Above-the output speed, below-armature current.

V. Conclusion

The results, shown in Figure 4 were obtained in assigned speed of working machine ω_{set}=100 rad / s; integrating factor k_{int}=*0.02*; and poles of the closed system are located in a circle of radius = *stab_zone=0.27*. In *k=80* (8 th second) appears disturbance in the form of torque-Mc of the shaft of the working machine.

If you compare results of Figure 2 and Figure 4 shows that in the control of the proposed algorithm, acceleration is smooth and three times faster. There are no fluctuations which are caused by elastic connections. Inrush current exceeds nominal only *1.9* times (I_{in} = *1.9*I_{nom}). Integral component in the law for control eliminated the disturbance and return the system back to set speed.

From the experiments it became clear that the algorithm is highly sensitive to the location of the poles of the closed-loop system (vector **μ**) and the value of the coefficient of integration k_{int}. Here **μ** and k_{int} are determined experimentally. Theoretical problem of determining **μ** and k_{int} is not solved.

VI. References

[1] Mikhov M.R., T.T.Georgiev, Optimal Modal Control of Dual-Mass Electromechanical Systems, Proceedings EDS'08 IMAPS CS International Conference; Invited Papers Brno, Czech Republic, September 10-11, 2008, pp.47-55.

[2] Nikolov N., Algorithm for Synthesis of Modal Adaptive State Controller, International Conference "AUTOMATICS AND INFORMATICS'07", proceedings vol.1, p.I-13÷I-16, october 3-6, Sofia, 2007.

[3] Nikolov N, Study of algorithms for synthesis of modal adaptive state controllers, dissertation-PhD, TechnicalUniversity of Varna, Bulgaria, 2009.

[4] Башарин А.В., В.А.Новиков, Г.Г. Соколовский, Урпавление электропроводами, Ленинград, Энергоиздат, 1982г.

[5] Ключев В.И., Теория на електрозадвижването, София, Техника, 1989.

Creative Development Support of Doctoral Students, Post-Doctoral and Young Researches in the Field of Computer Science, BG051PO001-3.3.04/13, European Social Fund 2007-2013 Operational Programme "Human Resources Development".

Analysis of the Opportunities for Energy Efficiency Improvement of Electric Vehicle Regenerative Braking

Zhelyazko Kartunov[1], Dimitar Bojchev[2] and Borislav Traykov[3]

***Abstract* – In current paper we made an analysis of the energy efficiency regenerative braking process.**

It was applied a method of series connected high efficiency DC-DC converters to the electric drive machine. More over in the paper is analyzed system behavior, controlled by proportional-integration regulator. Finally is suggest an algorithm for realizing of all process with single chip microcontroller and defined the limitation of regenerative braking efficiency.

***Keywords* – Regenerative braking, Efficiency, DC-DC converter, Regulator.**

I. INTRODUCTION

A regenerative brake is an energy recovery mechanism which slows a vehicle by converting its kinetic energy into another form, which can be either used immediately or stored until needed. This contrasts with conventional braking systems, where the excess kinetic energy is converted to heat by friction in the brake linings and therefore wasted.

In the speed range in which the braking energy is most dissipated, the operating efficiency of the electric motor, functioning as a generator, may be of most concern.

The braking energy dissipated in the low-speed range, such as below 15 km/h in all the typical driving cycles, is insignificant. This result indicates that we need not attempt to obtain high operating efficiency at low speeds in the design and control of regenerative braking. In fact, it is difficult to regenerate at low speeds, because of the low motor electromotive force (voltage) generated at low motor rotational speeds. Thus, the operation of the hybrid brake system must be at speeds higher than a minimum threshold value. Electric regenerative braking should be applied primarily to recapture as much braking energy as possible. At speeds lower than this threshold, mechanical braking should be primarily applied to ensure the vehicle's braking performance [1].

[1]Zhelyazko Kartunov is with the Faculty of Transport, Technical University of Sofia, Kliment Ohridski blvd. 8, 1000 Sofia, Bulgaria, E-mail: g_kartounov@tu-sofia.bg.

[2]Dimitar Bojchev is with the Faculty of Telecommunications Technical University of Sofia, Kliment Ohridski blvd. 8, 1000 Sofia, Bulgaria, E-mail: dbojchev@tu-sofia.bg.

[3]Borislav Traykov is with the Faculty of Transport, Technical University of Sofia, Kliment Ohridski blvd. 8, 1000 Sofia, Bulgaria, E-mail: btraykov@tu-sofia.bg.

II. FORMULATION OF THE PROBLEM

One of the main problems in the implementation of electronic circuits for managing the process of regenerative braking is a practical impossibility to realize DC-DC converter operated without loss. This leads to restriction of the returned energy in to the battery in the process of recovery.

In general, regenerative braking depends on the initial speed, from which start the braking, the mass of the car, the resistance forces and slope of the road. It is important to note that this examination is done only in terms of energy balance.

$$E_r = \eta_g \eta_t \eta_B \left\{ \left[\int_{t0}^{t1} V_0 . \sum \overrightarrow{F_S} \, dt \right] \pm E_P - E_K \right\} \quad (1)$$

where E_r - regenerative braking energy, m - mass of the vehicle, V_0 - initial speed of the vehicle, $\sum \overrightarrow{F_S}$ - the sum of resistance forces affecting the vehicle, E_P - potential energy of the vehicle, E_K - kinetic energy of the vehicle, η_g - efficiency of the machine (in generator motor mode), η_t - efficiency of the DC-DC converter, η_B - efficiency of battery charging.

From the analysis made here shows that a significant contribution to energy loss is DC converter. This problem is especially relevant at relatively low speed of the car the car and consequently a small amount of energy they could recuperate. Practical in this case regenerative process is ineffective and inappropriate.

In general, losses in the DC converter can be divided into two groups – permanent loss and losses, depending on the converted power. Assuming that the power loss increases linearly in terms of power conversion, it can be written the following expression:

$$P_Z = P_C + k_Z . P_{in} \quad (2)$$

Consideration is simplified in terms of that interest in this analysis is the process in a small amount of regenerated power.

$$\eta_t = \frac{P_{in} - P_C - k_Z . P_{in}}{P_{in}} \quad (3)$$

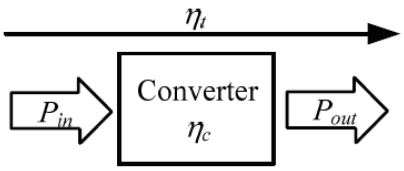
The main idea is published in [2]. The total efficiency is obtained by including the DC-DC converter in part of the input power. It can be calculated by the expression:

$$\eta_t = \frac{P_1 + P_2.\eta_c}{P_1 + P_2} \tag{4}$$

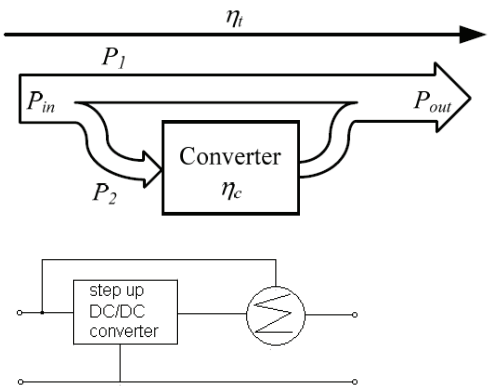
$$\eta_c > \frac{P_1 + P_2.}{P_2}\eta - \frac{P_1}{P_2} \tag{5}$$

The total efficiency of the proposed converter is obtained by Eq. (4) if the loss of the direct mode power P_1 can be neglected. Therefore, the total efficiency obtained by using the proposed concept is improved, as shown by Eq. (5).

The basic idea is shown in Fig. 1.



(a) Conventional concept.



(b) Proposed concept.

Fig. 1. Power flow diagrams.

It can be seen that the resulting efficiency is higher then the same of DC-DC converter.

III. Functional Scheme of Converter with PI Controler

For aims of current work it is used proportional integration regulator. The derivate component of regulator is not included because the reference is constant. It is important to be noted that the maximal energy have to be regenerated in the battery and the ultracapacitor. For this reason the regulator reference is always in maximal value. Thus the schematic diagram is shown in Fig. 2.

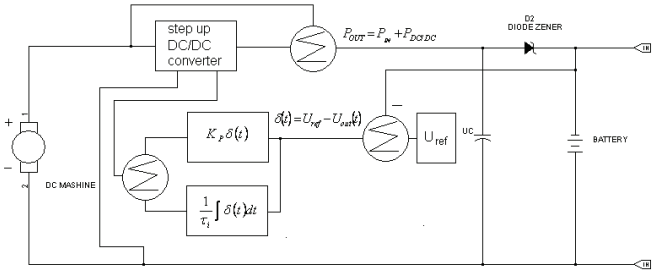


Fig. 2. Functional diagram of regeneration.

A. Expression of output energy.

In general case the regenerated energy can be expressed as:

$$E_r = E_{in} + E_{DC/DC} \tag{6}$$

E_r - energy of regenerative braking, E_{in} - input energy

$E_{DC/DC}$ - output energy from DC/DC converter

It can be written following expressions:

$$P_{out} = U_{out}.I_{out} \tag{7}$$

P_{out} - output power

were:

$$I_{out} = I_{bat} + I_{UC} \tag{8}$$

I_{out} - output current, I_{batt} - battery charging current,

I_{UC} - ultracapacitor current

and

$$U_{out} = U_{UC} = U_{batt} + U_D \tag{9}$$

U_{out} - output voltage, U_{batt} - battery voltage,

U_{UC} - ultracapacitor voltage, U_D - diode voltage

The output signal of regulator will be:

$$U_M = K.\delta(t) + K_i \int \delta(t)dt \tag{10}$$

U_M - managing voltage, output signal of PI regulator,

K - proportional value of regulator

K_i - integration value, $\delta(t)$ - output error versus time

where:

$$\delta(t) = U_{out}(t) - U_{ref} \tag{11}$$

After substituting Eq. (11) in Eq. (10) the total output voltage of DC-DC converter became:

$$U_{DC/DC} = k_1.k_{tr}.\begin{Bmatrix} K\left[U_{out}(t) - U_{ref}\right] + \\ K_i \int \left[U_{out}(t) - U_{ref}\right] dt \end{Bmatrix} \quad (12)$$

Where:

k_{tr} - coefficient of step up transformation of DC/DC converter, k_1 - coefficient of transformation of U_M related to the coefficient of Pulse Wide Modulation

Finally the expression of output energy becomes:

$$E_{out} = \int_{t0}^{t1} \left\{ k_1.k_{tr}.\begin{Bmatrix} K\left[U_{out}(t) - U_{tef}\right] + \\ K_i \int \left[U_{out}(t) - U_{ref}\right] dt \end{Bmatrix} \right\}.(I_{batt} + I_{UC})dt + E_r \quad (13)$$

Where:

t_0 - start moment of regeneration, t_1 - final moment of regeneration

Of Eq. (13) it can be seen that the regenerative braking is possible until recovered energy becomes equal of DC-DC converter losses, e.j.:

$$P_C + k_Z.P_{in} = \left\{ k_1.k_{tr}.\begin{Bmatrix} K\left[U_{out}(t) - U_{tef}\right] \\ + K_i \int \left[U_{out}(t) - U_{ref}\right] dt \end{Bmatrix} \right\}.(I_{batt} + I_{UC})dt + P_{in} \quad (14)$$

B. Algorithm for realizing of proposed concept.

The proposed concept can be realized with single chip microcontroller. In this way the scheme will become more energy efficient and relatively cheaper then others decisions. In this case one of the basic problems is numerical integration. This mathematical operation needs of relatively large system resource. This is an important question for optimal practical realization. In [3] is represented a simplest algorithm for numerical integration, suitable for the goal of current work. Thus all mathematical operations of Eq (13) can be represented by sum and subtraction.

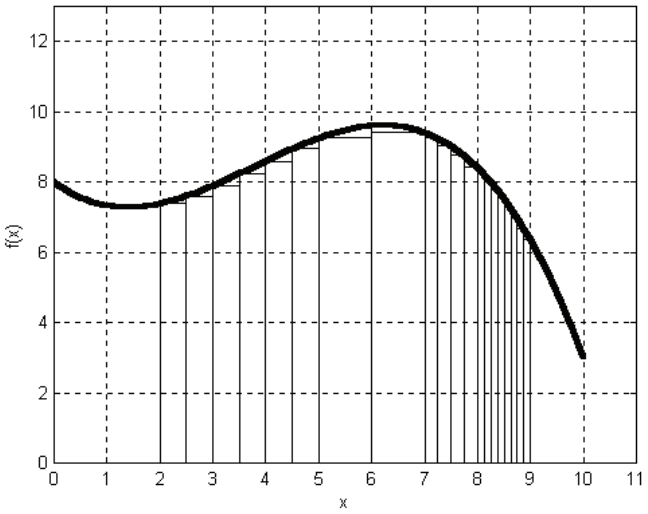


Fig. 3. Numerical integration by using a variable step discretization.

IV. CONCLUSION

In current paper it was made an analysis of possibilities for increasing energy efficiency of regenerative braking process. Then it was suggest a new circuit consisted proportional integration regulator and step up DC-DC converter. It was applied a method of separating input energy [2] and in this way the energy efficiency of the suggested system is higher then the DC-DC converter. Moreover the suggested scheme allows common work of ultracapacitor and battery. Then it was made an expression Eq. (14) witch define the limiting conditions for regenerative braking. Finally it was made recommendation for composing an algorithm for realizing suggested concept with single chip microcontroller.

ACKNOWLEDGEMENT

The results are obtained under a project funded by the grant for scientific research in TU - Sofia.

REFERENCES

[1] M. Ehsani, Y. Gao, A. Emadi “Modern Electric, Hybrid Electric and Fuell Cell Vehicles”, CRC PressTaylor & Francis Group, United States of America, ISBN:978-1-4200-5398-2, 2010.

[2] J.-I. Itoh, T. Fujii, "A New Approach for High Efficiency Buck-Boost DC/DC Converters Using Series Compensation", Nagaoka University of Technology, IEEE Power Electronics Specialists Conference, ISSN:0275-9306, 2008

[3] R. Miletiev, Al. Bekiarski “A Simplest Algorithm for Numerical Integration by Single Chip Microcontroler” .Technical University of Sofia proceeding V.58, ISSUE 1, ISSN:1311-0829, 2008

On-Line Identification of Time-Varying Systems

Mariana Todorova[1]

Abstract – **A method for on-line identification of continuous linear time-varying systems is applied based on the block-pulse functions. A suitable m-file is proposed. Numerical examples are considered and the efficiency of the identification algorithm in case of noise influence is examined.**

Keywords– **On-line identification, Linear time-varying systems, Block-pulse functions.**

I. INTRODUCTION

Applications of the linear time-varying (LTV) systems include rocket dynamics, time-varying linear circuits, satellite systems, pneumatic actuators, ets. LTV structure is also often assumed in adaptive and standard gain-scheduled control systems. The motivation for this work is an investigation of a method for on-line identification of continuous LTV systems based on the block-pulse functions. The time-varying linear systems identification immediately in the continuous-time domain is a quite complicated problem. To avoid the direct measurements of time derivatives, the following identification methods have been proposed:

1. Methods based on Poisson moment functionals [5, 11];
2. Methods based on orthogonal functions and polynomials [3, 4, 6 - 10].

The second methods are non-recursive. The first methods can be applied for recursive identification, but the involved Poisson filter chains must be realized physically by analogue devices and the system structure may be complicated.

Z. Jiang and W. Schaufelberger [1] proposed a new method for recursive identification of continuous time-varying linear systems. Based on the relation between entries of block pulse integral operational matrices, regression equation models of continuous time-varying linear systems can be obtained and recursive algorithms developed for discrete time model identification can be applied to estimate the time-varying parameters without much modifications.

In this paper the recursive method above is applied and Matlab file is proposed. Numerical examples are considered.The efficiency of the identification algorithm in case of noise influence is examined andcorresponding conclusions are made.

[1]Mariana G. Todorova is with the Faculty of Computing and Automation, TechnicalUniversity, "Studentska" str. № 1, 9010 Varna, Bulgaria, E-mail: mgtodorova@yahoo.com

II. BLOCK PULSE ORTHOGONAL FUNCTION PROPERTIES

Before commencing consideration of the problem, some pertinent properties of block – pulse function (BPF) will be briefly reviewed.

A block – pulse function [2,3,4, 7] is defined over a time interval $t \in [0,T]$ as $\{B_i(t)\}, i = \overline{1,m}$, where:

$$\left.\begin{aligned} B_i(t) &= 1, \quad for\ t \in [0; T/m] \\ B_i(t) &= \begin{cases} 1, & for\ t \in [(i-1)T/m;\ iT/m] \\ 0, & elsewhere \end{cases} \\ & for\ i = 2, 3, ..., m \end{aligned}\right\}. \quad (1)$$

They posses the orthogonal property

$$\int_0^T B_i(t).B_j(t).dt = \begin{cases} T/m, & if\ i = j \\ 0, & if\ i \neq j \end{cases}. \quad (2)$$

A function y(t), absolutely integrable in the region $t \in [0,T]$, may be approximated to

$$y(t) \cong \sum_{i=1}^{m} y_i.B_i(t) = Y_M{}^T.B_M(t), \quad (3)$$

where:

Y_M - block – pulse function coefficient vector of the function $y(t)$;

$$B_M(t) = [B_1(t)\ \ B_2(t)\ \ ...\ \ B_m(t)]^T.$$

The orthogonal function $B_M(t)$ has the property

$$\underbrace{\int_0^t ... \int_0^t}_{r-times} B_M(t).dt^r = P_M^r.B_M(t). \quad (4)$$

where P_M is $(m \times m)$ integral operational matrix.

Applying the property (4) to y(t) yields

$$\begin{aligned} &\underbrace{\int_0^t \int_0^t ... \int_0^t}_{r-times} y(t).dt^r = \\ &= \underbrace{\int_0^t \int_0^t ... \int_0^t}_{r-times} Y_M^T.B_M(t).dt^r = Y_M^T.(P_M)^r.B_M(t) \end{aligned} \quad (5)$$

III. On-Line Identification of Second Order Time- Varying Linear System

Consider the following N-th order differential equation

$$\sum_{i=0}^{N} a_i(t) y^{(i)}(t) = \sum_{i=0}^{N} b_i(t) u^{(i)}(t) \tag{6}$$

where:

$u(t)$, $y(t)$ – system input and output respectively;

$a_i(t)$ and $b_i(t)$ - time functions which are described by n-th order power polynomials with constant coefficients, as follows:

$a_i(t) = a_{i,0} + a_{i,1}t + \ldots + a_{i,n}t^n$;

$b_i(t) = b_{i,0} + b_{i,1}t + \ldots + b_{i,n}t^n$.

The identification problem for continuous LTV systems described by the equation bellow is to determine $a_{i,j}$ and $b_{i,j}$ ($i = \overline{0,N}$; $j = \overline{0,n}$) from the system input $u(t)$ and output $y(t)$. This problem can be solved straightforward based on the block-pulse regression equation [1]

$$\sum_{i=0}^{N} \sum_{j=0}^{n} a_{i,j} z_{i,j,l} = \sum_{i=0}^{N} \sum_{j=0}^{n} b_{i,j} v_{i,j,l}\ , \tag{7}$$

where $z_{i,j,l}$ and $v_{i,j,l}$ are linear combinations of the block pulse coefficients y_{l+r} and u_{l+r} ($r = \overline{0,N}$) respectively.

The relation between $z_{i,j,l}$ and y_{l+r}, and relation between $v_{i,j,l}$ and u_{l+r} respectively have the forms:

$$z_{i,j,l} = \sum_{s=0}^{\min(i,j)} \left((-1)^s \binom{i}{s} \left[\sum_{k=0}^{N} \sum_{r=0}^{N-k} (-1)^k \binom{N}{k} \frac{j! h^{N-i+j}}{(N-i+j+1)!} p_{N-i+s,\, j-s,\, l+r,\, l+N-k y_{l+r}} \right] \right) \tag{8}$$

$$v_{i,j,l} = \sum_{s=0}^{\min(i,j)} \left((-1)^s \binom{i}{s} \left[\sum_{k=0}^{N} \sum_{r=0}^{N-k} (-1)^k \binom{N}{k} \frac{j! h^{N-i+j}}{(N-i+j+1)!} p_{N-i+s,\, j-s,\, l+r,\, l+N-k u_{l+r}} \right] \right) \tag{9}$$

Equation (8) can be written as:

$$z_{i,j,l} = \sum_{s=0}^{\min(i,j)} (-1)^s \binom{i}{s} \frac{j!}{(j-s)!} x_{s,l} \tag{10}$$

where

$$x_{s,l} = \sum_{r=0}^{N} y_{l+r} \left[\sum_{k=0}^{N} (-1)^k \binom{N}{k} \frac{(j-s)! h^{N-i+j}}{(N-i+j+1)!} p_{N-i+s,\, j-s,\, l+r,\, l+N-k} \right] \tag{11}$$

By substitution u_{l+r} for y_{l+r} and $v_{i,j,l}$ for $z_{i,j,l}$ the value $v_{i,j,l}$ can be given in the same way. The values $x_{s,l}$ (eq. (11)) can be computed from the N-th order difference of entries in each row of the submatrix and from the block-pulse coefficients of y(t). The value $z_{i,j,l}$ is a linear combination of the obtained values $x_{s,l}$. After all the terms of $z_{i,j,l}$ and $v_{i,j,l}$ are obtained, the block-pulse regression equation (7) can be constructed, and the algorithms for discrete-time model identification can be applied to the time-varying linear system without much modification.

Consider a system modelled by the following time- varying ordinary differential equation

$$\frac{d^2 y(t)}{dt^2} + a_1 \frac{dy(t)}{dt} + a_2(t) y(t) = b_0(t) u(t)\,, \tag{12}$$

where:

$a_2(t) = a_{2,0} + a_{2,1}t + a_{2,2}t^2$;

$b_0(t) = b_{0,0} + b_{0,1}t$.

Given a record of output y(t) and input u(t) signals, shown in Fig.1, the problem is to estimate the unknown parameters $a_1; a_{2,0}; a_{2,1}; a_{2,2}; b_{0,0}; b_{0,1}$ of the system (12).

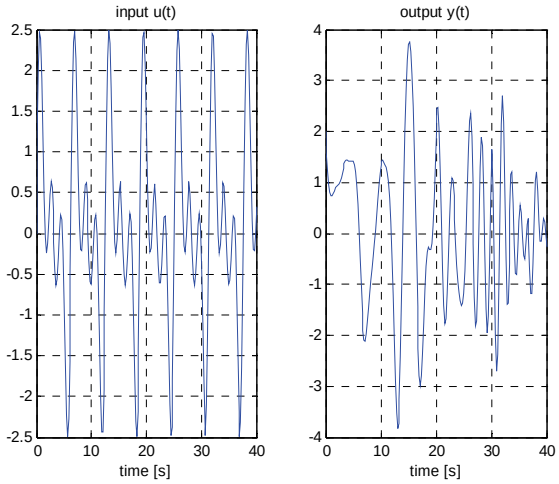


Fig. 1. Output and input signals

M – file is created in Matlab based on considered algorithm. The results obtained from the recursive estimation are given in Fig. 2.

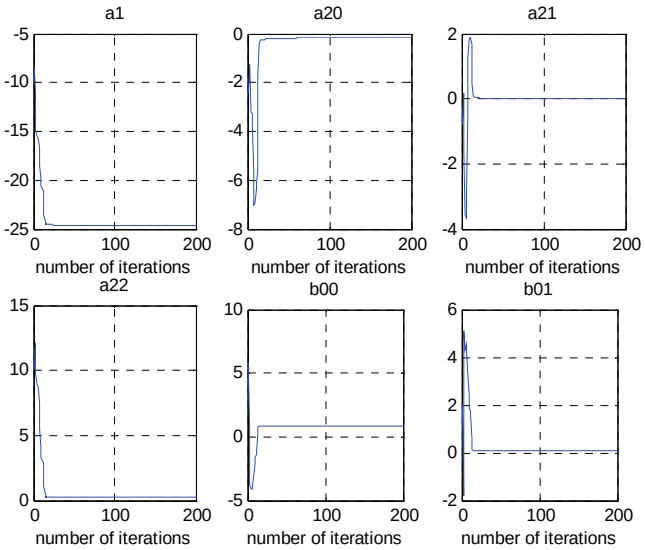


Fig. 2. Values of the unknown parameters

As givev from Fig.2 the unknown parameters after small number of iterations are obtained. The output of the model Ym(t) and the prediction error E are given in Fig.3.

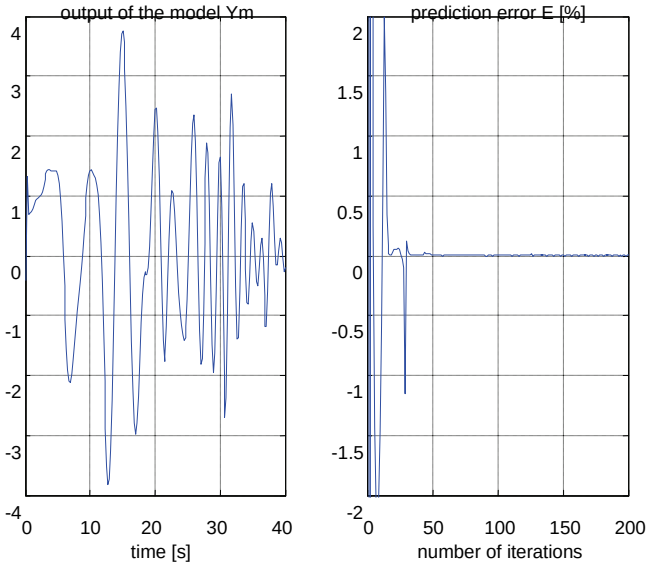


Fig. 3. Model output and prediction error

An example shows that the method is simpler and more efficient in comparison with other methods for solving the same problem.

Then the efficiency of the identification algorithm in case of noise influenceis examined. Therefore the signaly(t) with independent zero – mean white Gaussian noise iscorrupted. First the ratio nose-to-signal q= 1% is used.The values of the unknown parameters $a_1; a_{2,0}; a_{2,1}; a_{2,2}; b_{0,0}; b_{0,1}$ are given in Fig.4.

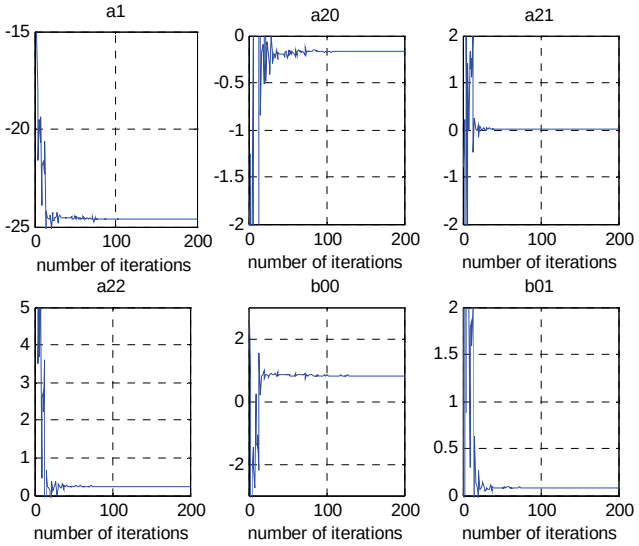


Fig. 4 Values of the unknown parameters
(ratio nose-to-signal q= 1%)

Then the signaly(t) with independent zero – mean white Gaussian noise with q= 10% iscorrupted. The values of the unknown parameters areshown in Fig.5.

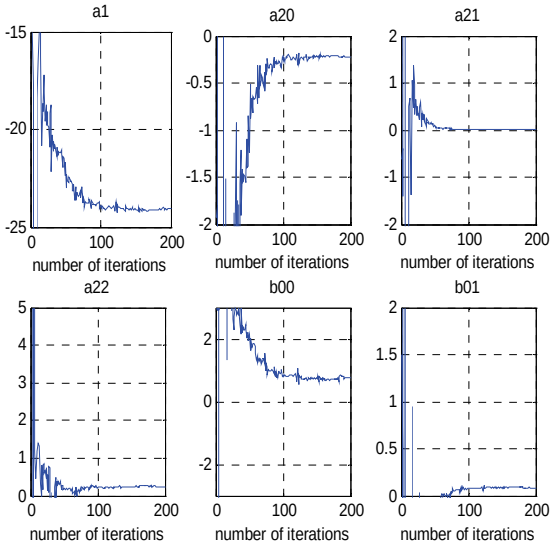


Fig. 5.Values of the unknown parameters
(ratio nose-to-signal q= 10%)

IV. Conclusion

A comparatively new method for on-line identification of continuous linear time-varying systems is applied based on the block-pulse functions. A suitable m-file is proposed. Numerical examples are considered and the noise immunity of considered algorithm is investigated. Comparing with other methods for solving the same identification problem, no analogue devices, no large computations in the data preparation stage and no initial values are involved in this method. Therefore the estimation procedure is much simpler.

Acknowledgement

The results were obtained under a project №7 funded by the grant for research in TU-Varna for 2011.

References

[1] Z.H. Jiang, and W. Schaufelberger, “A recursive identification method for continuous time-varying linear systems”, IEEE 34th MidwestSymposium, Conf. Proceedings, vol. 1, pp. 436-439, Monterey, USA, May 1991.

[2] Nath A. K., T. T. Lee, On the multidimensional extension of block – pulse functions and their completeness”, Int. J. Systems Sci., vol. 14, no. 2, pp. 201 – 208, 1983.

[3] Todorova M., “Research of the possibilities and application of two dimensional orthogonal functions for dynamic distributed parameters systems identification”, PhD Thesis, Varna, 2003.

[4] Wang Shienyu, Jiang Weisun, “Identification of nonlinear distributed parameter systems using block pulse operator”, IFAC Identification and System Parameter Estimation, York, UK, 1985.

[5] Saha D.C., G.P.Rao, “A general algorithm for parameter identification in lumped continuous systems – the Poisson moment functional approach”, IEEE Trans. on Automatic Control, vol.27, pp.223-225, 1982.

[6] Chang R. Y., S. Y. Yang, M. L. Wang, “A new approach for parameter identification of time-varying systems via generalized orthogonal polynomials”, Int. J. Control, vol.44, pp.1747-1755, 1986.

[7] Todorova M., “Identification of one class of distributed parameter systems based on orthogonal functions”, ICEST,vol. 1, pp. 457-460, Macedonia, 2010.

[8] Todorova M., M. Vasileva, “Two-dimensional block-pulse functions application for identification of the air power line parameters under the influence of lightning overvoltages”, ACTA Universitatis pontica euxinus, vol. VI, no. 8, pp. 85-89, 2007.

[9] Todorova M., T. Pencheva, “Haar orthogonal functions based parameter identification of Wastewater Treatment Process with Distributed Parameters”, Chemical and Biochemical Engineering Quarterly, Croatia, Vol.19, No 3, pp. 275 – 282, 2005.

[10] Kolarov T., M. Todorova, D. Genov, “Identification of the connected motions of the immersed towed trolling depressor”, Seventh international conference on marine science and technology Black Sea, Varna,vol. 1, pp. 75 – 80, 2004.

[11] Garnier, H.; Sibille, P.; Richard, A., “Continuous-time canonical state-space model identification via Poisson moment functionals”, Decision and Control, Proceedings of the 34th IEEE Conference, vol.3, pp.3004 – 3009, 1995.

A Model of Remote Control of Railway Traffic Based on PLC Technique

Saša Krstanović[1], Gordan Stojić[1], Dragan Šešlija[1], Ilija Tanackov[1], Laslo Tarjan[1] and Jovan Tepić[1]

Abstract – **Shown in this paper is a model of railway traffic regulation, based on application of Programmable Logic Controller (PLC) technique which represents a pioneering attempt in recent modes of remote control and command of train traffic. The model has been tested on a simple example of a railway station.**

Keywords– **trains, remote control, safety, signalization, PLC.**

I. INTRODUCTION

Railway interlocking systems are apparatuses that prevent conflicting movements of trains through an arrangement of tracks. These systems are finite state machines that take into consideration the position of the switches (of the turnouts) and do not allow trains to be given clear signals if the routes to be used by the trains intersect or run a turnout set against and thus cause train derailment.

Computer applications to railway interlocking systems [1]are safety critical applications and must obey the rules and conditions or informal specifications of the already designed or implemented railway interlocking system [2], [3].

According to this model, any given proposed situation is safe **if and only if** a certain set of criteria (the position and type of trains and the movements allowed - the latter depend on the position of the switches and the color of the semaphores) is met [2], [3]. This model is independent from the topology of the station and the fact that trains could occupy more than one section is also considered.

II. OVERVIEW

In this section we shall discuss some preliminary notions on the matter which can be helpful for unacquainted readers.

Locomotives and railways rolling stocks do not have wheels with plain rims but **single flanged conical wheels** which are guided along the tracks. Therefore this is a guided transportation system.

A train can be guided from one track to another only at certain places, where a**turnout** is installed (US: railroad switch).It sends trains on direct track or diverted trackwhich is displayed via turnout indicator (US: switch indicator). The turnout hastwo directions: the facing direction and the trailing (opposite) direction.

If the train came from the trailing direction, it could find the switch in the wrong position, and the train would run a turnout set against (US: make a trailing point movement). If this would happen,a turnout could be seriously damaged and/or the train could derail.

The turnouts can be designed so that running a turnout set against is allowed at very low speed . They are called ***trailable turnouts***.

However, the design of the turnouts does not usually allow running a turnout set against.

The movements of the rolling stock are controlled by **signals and semaphores**. Formally speaking, semaphores are distinguished from signals (signals are mechanical devices, usually incorporating coloured lights too) and we shall consider possible states only for (main) semaphores.

There are also ***advanced semaphores***, coordinated with the corresponding main semaphore, that can indicate the train to prepare to stop at the next (main) semaphore (i.e., to proceed at restricted speed) and***complex semaphores*** at junctions and they are controlling each possible route.

There are two main safety devices related to railway traffic: automatic block systems (regarding traffic in railway lines) and railway interlocking systems (regarding coordination among turnouts and signals/semaphores within stations, junctions, etc.).

The purpose of an **automatic block system** is to avoid collisions between trains running on a line by dividing it into intervals (denoted blocks or sections). Fixed semaphores indicate whether or not a train may enter a block (this is based on automatic train detection).

On single-tracked railway lines it is necessary to space trains to avoid collisions by maintaining at least one whole section clear behind each train at every moment. It is also necessary to avoid collisions of trains moving in opposite directions. Therefore, before authorising a train to proceed, it has to be checked that the sections ahead (until the next passing loop) are clear too.

Railway interlocking systems are designed to prevent conflicting movements through an arrangement of tracks such as junctions or crossings.

Once remote controls of turnouts and signals were concentrated in signal boxes (in order to spare workers the task of having to walk to turnouts (respectively: signals) whenever the position of their switches (respectively: arms) has to be changed), it was possible to think of global coordination of all the turnouts and signals within a railway station or junction. The apparatus that takes care of this is the railway interlocking system.

Since 1990s, high-tech companies began to install ***computer-controlled railway interlocking systems***[4-7] with topology-independent railway interlocking systems [9].

Railway interlocking systems should comply to two safety requirements, namely that trains following signals):

[1]All authors are with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, Novi Sad, Serbia,e-mail: sasa.krstanovic@gmail.com, gordan@uns.ac.rs

- cannot collide, and
- cannot derail, i.e., trains should not be authorised to run a turnout set against (if the turnout is not a trailable one) and changes of switches under a train should not be authorised.

Within a railway interlocking system **a route** denotes a path along the topology of the station or junction (for instance a path from an entrance of the station to a certain track where the train will stop). Establishing a route implies to adequately set the switches and signals/semaphores along the train route. Once an engine driver has been given a clear signal indicating a route, the route cannot be changed before the train has completely cleared it (then it is said that ***the route is locked***).

The standard approach to railway interlocking systems design is to predefine the admissible train routes and to manually study in advance their compatibility.

In this paper presented is a model of train traffic regulation based on the application of Programmable Logic Controller (PLC) and it presents a pioneering attempt in the recent ways of regulation train traffic in Serbia.

III. A Model of Remote Control of Railway Traffic

In the model presented in this paper the traffic flow on the railway is circular (Fig.1). In the model there is Station 1 and Station which are connected with interstational distances in areas A and B.

Circular connecting of train stations does not exist in practice, but in order to show a more realistic simulation of real situations with only 2 instead of 3 stations in laboratory conditions with 2 PLCs, H0 system model was made. However from the point of view of regulation manner and railway safety) the presented model truthfully presents the realistic situation in railroad traffic.

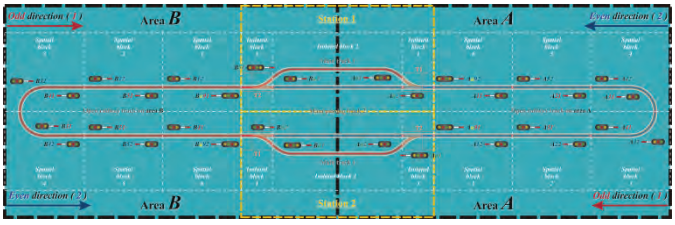


Fig.1. Situational Diagram

The model can work under two regimes: when it is governed by an operator (train dispatcher) and under the automatic regime, when the model forms train routes, based on the previously defined parameters from the timetable, dependency tables and train traffic monitoring

The whole system is meant to be modular and can be later upgraded with new modules or the existing can be changed/expanded.

In order to have safe and smooth train traffic, dependence tables have been done on the model (In total 19: 3 acceptances + 2 exits + 2 entrances + 2 combinations exit and entrance + 1 passing + 2 combinations exit, passing and entrance + 4 overreaches + 1 overtaking + 2 crossings). The dependence table represents the signal and turnout positions for all possible traffic situations on the presented model. In this paper presented is only one characteristic example of a constructed dependency table. (Table I).

TABLE I
Dependency table

Signals ▶	Signals on open railway											Turnouts			Signals in stations							
	Area A, Station 1					Area A, Station 2						Station 1		Station 2	Station 1				Station 2			
Rd. pth. ▼	A11	A21	A31	A41	A51	A12	A22	A32	A42	A52	A62	T1	T3	T2	Ao1	Ao2	Bo1	Bo2	Ao1	Ao2	Bo1	Bo2
ENTER												+	+	+								
ENTER												-	-	+								

Space which is included in the station and open railway is defined by the number of their spatial blocks (SP) and isolated blocks (IB) which spread between isolated structures- places where the track is electrically separated into independently controlled and governed blocks.

In the model there are station tracks and tracks on the open railway (Fig.1).

Tracks in the station accept trains which after the performed manipulation (train entrance/exit, merchandize upload/unload, taking/leaving carts) and/or finished traffic reasons (crossing/ overtaking) continue their ride.

For every train ride (entrance, exit or pass trough the station) tracks and turnouts are positioned in the ride direction. A train ride on the track on the open railroad is possible only if the acceptance exists that is given by the neighboring station. The acceptance given for one path is valid for all consecutive trains until the path changes.

When the ride path is formed, the tracks in the station and on the open railroad get the color of the path in which the train is supposed to move.

Turnouts in the model have been labeled T1 and T2. Each turnout consists of the pair of linked tapering rails, known as *points*, lying between the diverging outer rails.For each turnout in the model, a position is defined based on intended ride. When the ride path is formed, the turnouts are automatically set in required position.After a finished ride over the turnout, the turnout automatically returns to its regular position (direction).

When forming the ride path or changing the direction of the acceptance in the model, all safety procedures used in practice were applied.

In this model three types of signals exist, entrance signals (Au91, Au92, Bu91 and Bu92), exit signals (Ao1, Ao2, Bo1 and Bo2) and spatial signals (A11~A51, A12~A52, B11~B51, B12~B52)(Fig.1). Entrance and exit signals that enable the ride with turning have four lanterns, while the rest have only three.

Entrance signals allow or forbid trains to enter the station and exit signals allow or forbid the exit of trains from the station. Spatial signals allow or forbid the entrance into the next spatial block depending on the fact if the next SB is available or not.

The regular position of the entrance and exist signals is “STOP”. Train dispatcher or the model under the automatic regime, set the signals in the position of the enabled ride by forming entrance/exit ride path.

The signals in the model have been marked in accordance to existing rues on Serbian railroads, and signal signs they emit correspond to the signals signs being emitted on railroads where such signals (for example tracks on Corridor X). Signal signs emitted by the signals in the model have a double meaning. They inform the locomotive driver how to drive in the SB he is in and what to expect in the next SB.

Station panel (SP) represents a command-control panel from which the train dispatcher regulates the train traffic. (Fig. 2). At the same time he has the option to control. In this model there are two SP one for each station.

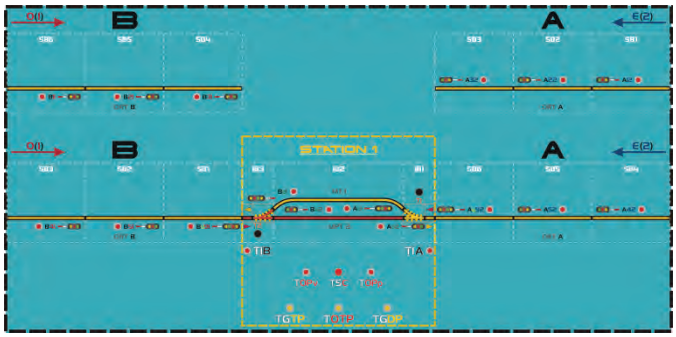


Fig.2. Station panel

From SP it is possible to allow and forbid entrance/exit of trains, to change the direction of the acceptance and to intervene in case of immanent danger. In that case the train dispatcher can put any signal to STOP position and to turn of the charge in the contact network on the electrified rails.

On SP exist the following switches: signal buttons, TGTP (combinatory button for acceptance demand), TOTP (combinatory button for denial of the demanded acceptance), TGDP (combinatory button for allowing acceptance), buttons TIA and TIB (for permitting the exit to areas A and B), TOPv (combinatory button for ride path recall), TSC (combinatory button for putting up the signal sign "STOP") and TOPp (combinatory button for the recall of overrun path) All buttons listed are used in combination with the corresponding signal button. The model executes the command only if the right combination is used.

With SP, based on the indicators, it is possible to control the position of the signal (type of signal sign), availability/ unavailability of the tracks and SB, position and availability of the turnouts and train location.

IV. PLC TECHNIQUE

A process control system is made up of a group of electronic devices that provide stability, accuracy and eliminate harmful transition statuses in production processes. As technology quickly progresses, many complex operational tasks have been solved by connecting programmable logic controllers (PLCs) and a central computer. Beside connections with devices (e.g., operating panels, motors, sensors, switches, valves, etc.) possibilities for communication among instruments are so great that they allow a high level of exploitation and process coordination. In addition, there is greater flexibility in realizing a process control system. Each component of a process control system plays an important role, regardless of its size (e.g. without a sensor, the PLC would not know what is going on during a process).

In an automated system, a PLC controller is usually the central part of a process control system. Unlike general-purpose computers, the PLC is designed for multiple inputs and output arrangements, extended temperature ranges, immunity to electrical noise, and resistance to vibration and impact. With the execution of a program stored in program memory (battery-backed or non-volatile), PLC continuously monitors status of the system through signals from input devices. Based on the logic implemented in the program, PLC determines which actions need to be executed with output instruments. A PLC is an example of a hard real time system since output results must be produced in response to input conditions within a bounded time, otherwise unintended operation will result. [8].

Advantages of control panel that is based on a PLC controller can be presented in few basic points:

- compared to a conventional process control system, number of wires needed for connections is reduced by 80%,
- consumption is greatly reduced because a PLC consumes less than a bunch of relays,
- diagnostic functions of a PLC controller allow for fast and easy error detection,
- change in operating sequence or application of a PLC controller to a different operating process can easily be accomplished by replacing a program through a console or using a PC software (not requiring changes in wiring, unless addition of some input or output device is required),
- needs fewer spare parts,
- it is much cheaper compared to a conventional system, especially in cases where a large number of I/O instruments are needed and when operational functions are complex,
- reliability of a PLC is greater than that of an electro-mechanical relay or a timer.

Steps in systematic approach in designing a process control system:

1. Select an instrument or a system you want to control;
2. Specify all input and output instruments that will be connected to a PLC controller and following an identification of all input and output instruments, corresponding designations are assigned to input and output lines of a PLC controller;
3. Make a ladder diagram for a program by following the sequence of operations that was determined in the first step;
4. Program is entered into the PLC controller memory and when finished with programming, checkups are done and, if possible, an entire operation issimulated.

Before this system is started, you need to check once again whether all input and output instruments are connected tocorrect inputs or outputs. By bringing supply in, system starts working [8].

PLCs are programmed using application software on personal computers. The computer is connected to the PLC through Ethernet, RS-232, RS-485 or RS-422 cabling. The programming software allows entry and editing of the ladder-style logic. Generally the software provides functions for debugging and troubleshooting the PLC software, for example, by highlighting portions of the logic to show current status during operation or via simulation. The software will upload and download the PLC program, for backup and restoration purposes.

For the model described in this paper, PLC has to be capable of having complete control under the automatic work regime or partial under the manual work regime. In Fig.3 presented is the algorithm based on which the PLC program instructions are performed.

Fig.3.*Algorithm based on which the PLC instructions are performed.*

V. CONCLUSION

On railroad infrastructure in Serbia and surrounding countries, the existing system of railroad traffic is based on relay technique which has been integrated near the end of 1960s. This technique of traffic regulation, because of its age and overuse, is exposed to regular system failures. This endangers the safety and causes difficulties in traffic regulation causing significant train delays.

In the Serbian railroad system important prioritized are investments aimed at revitalizing and modernizing the infrastructure and adding more modern transport vehicles. The new techniques of traffic regulations, electronic station panels, are very expensive and are mostly a product of big industrial giants [1].

This paper represents a pioneering attempt of the application of PLC technique in regulating railway traffic in Serbia which is aimed at demonstrating the existence of domestic personnel resources and techniques that can provide contemporary and less costly solutions.

ACKNOWLEDGEMENT

This paper is a contribution to the Ministry of Science and Technological Development of Serbia funded project TR 36012.

REFERENCES

[1] Siemens, Thalis, Bombardier… promotional material
[2] Železnice Srbije,Saobraćajni Pravilnik, 1998.
[3] Železnice Srbije, Signalni Pravilnik, 2000.
[4] Siemens, Proyecto y obra del enclavamiento electronico de la estacion de Madrid-Atocha, Proyecto Tecnico, Madrid, 1988.
[5] Siemens, Microcomputer Interlocking Hilversum, Siemens, Munich, 1986.
[6] Siemens, Microcomputer Interlocking Rotterdam, Siemens, Munich, 1989.
[7] Siemens, Puesto de enclavamiento con microcomputadoras de la estacion de Chiasso de los SBB, Munich, 1989.
[8] Nebojša Matić,Introduction to PLC controllers on-line,Mikroelektronika, 2003.
[9] L. Villamandos, Sistema informatico concebido por Renfe para dise nar los enclavamientos, Vfa Libre 348 (1993) 65.
[10] Eugenio Roanes-Lozano, Antonio Hernando, A logic approach to decision taking in a railway interlocking system using Maple, Mathematics and Computers in Simulation, Article in Press, 2011.

Embedded Control of Pneumatic Muscles

Mladen Milushev[1], Todor Djamiykov [2] and Marin Marinov[3]

***Abstract* – In this report we present a base module for synchronal control of a six-legged robot. The starting point is maiking the same kind of module using a microcontroller (made by RENESAS) for the controlling of a three-joint leg, with a reflex implementation capatability.**
Presented is the research on the PWM muscles' control. The hardware and software components of the base module and algorithms are shown in the present report. The results from the real-time tests of their cooperation are given as well as the according parameters.
The end-goal is to achieve the quickest possible adaptation in the process of creating different set variants of autonomous multilink architectures for mobile robots' control.

***Keywords* – Mechatronics System, Walking Robot, Embedded Control, Modular Control, Fluidic Muscles.**

I. INTRODUCTION

Anyone developing a mobile robot needs to solve the problems related to locomotion in a specific environment. The nature of environment where the robot is supposed to be functioning is the crucial criterion for the driving force type. On an uneven surface the robot can move if its locomotory is very flexible and adaptive to every change in the terrain profile. Such a moveable machine is mostly based on the joints, aka leg. The larger the number of legs used, the more reliable the accomplished movements.

The efforts towards coordination of the legs increase significantly with the implementation of more legs. For a robot carried out with six legs with six joints each eighteen degrees of freedom must be calculated. The coordination and control through a centralized control system of those 18 freedom degrees provides for the anticipation of considerable difficulties. The present study deals with some of the tasks and problems-to-be-solved mostly related to the development of a system for the six legs' control.

The developing of control systems recently employs hierarchical principles based on two antagonistic approaches: *composition* and *decomposition*, (aka Top-Down- and Bottom-Up-principles) applied in an appropriate proportions. The Bottom-Up-principle helps identify the basic Mechatronic functions and add relevant Mechatronic structures. Thus a unified structural model is established and can further be applied with the Top-Down principle. The structural model describes the basic functions used as targets by the latter principle. In the defined Mechatronic circumstances the Bottom-Up-principle allows to fix the related joint's bending angle. Through the three joints forming a leg the position of that leg can be also set. If a gait and a respective posture of the body have to be realized by using the Top-Down principle, first the gait has to be distributed to that leg's locomotive functions and subsequently disintegrated to individual speeds for the separate joints. In case of some cross basic mechatronic functions, aka abstract functions stemming from the Bottom-Up principle, it's presumable for the Top-Down principle to initiate a preparation for them.

[1]Mladen Milushev is with the Department of Automation of Discrete Production Engineering, Faculty of Mechanical Engineering, Technical University - Sofia, 8 Kliment Ohridski blvd., 1000 Sofia, Bulgaria, e-mail: mcm@tu-sofia.bg

[2]Todor Djamiykov is with the Department of Electronics and Electronics Technologies, Faculty of Electronic Engineering and Technologies, Technical University - Sofia, 8 Kliment Ohridski blvd., 1000 Sofia, Bulgaria, e-mail: tsd@tu-sofia.bg

[3]Marin Marinov is with the Department of Electronics and Electronics Technologies, Faculty of Electronic Engineering and Technologies, Technical University - Sofia, 8 Kliment Ohridski blvd., 1000 Sofia, Bulgaria, e-mail: mbm@tu-sofia.bg

II. SYSTEM DESCRIPTION

In this paper we consider a walking robot with six identical legs equally distributed along both sides of the robot's body in three opposite pairs. The leg's joints are driven by pneumatic muscles (FESTO). So far the six-legged walking robot (Fig. 1.) has been developed using Solid Woks.

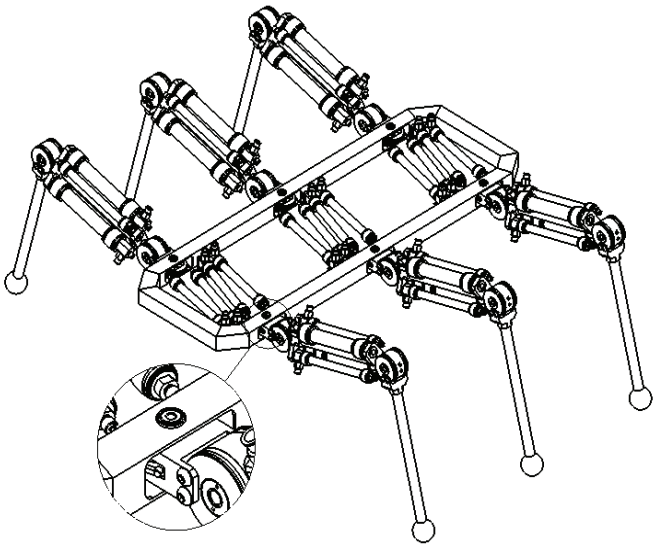


Fig.1. Leg-to-body attachment design.

For the walking robot ***BiMoR*** (**B**iologically **M**otivated **R**obot) a hierarchical and distributed computing architecture has been selected (Fig. 2.). By distributing the possibility of concurrent control functions is implemented on various micro-

controllers. Through the concept of distribution the need for communication is generated. The communication based on the master-slave principle with provides a suitable option for the control system for keeping the protocol economical and within the determined time limits for securing safety to the critical functions. It is important that the used sensors provide information about the absolute coordinates.

The hardware of the slave control system must fulfill the following tasks:

- collection and analysis of the measured variables;
- calculation of tax information;
- output control signals to the actuators.

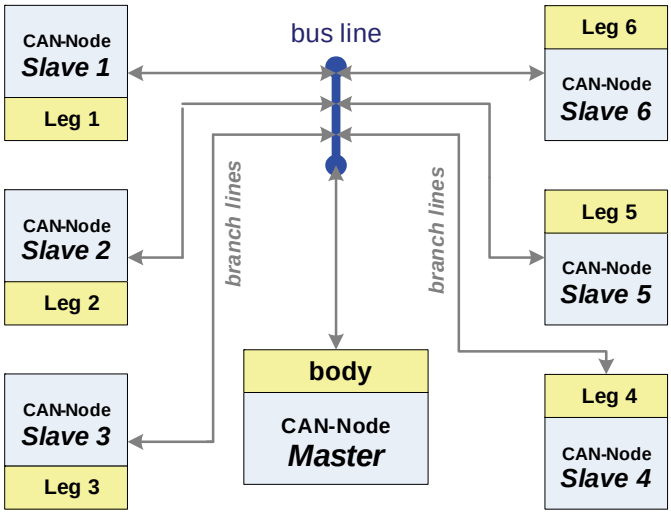


Fig.2. The Master in connection with all six slaves

For executing the basic legs functions like the closed-loop joint control (valve control, recording signals from the joint encoders) six R8C/23-microcontrollers are installed. On a basic level each sensor and actuator are connected with the interface board to the micro-controller board. The R8C/23 microcontroller is installed on industrial controller boards and contains one Full CAN module, which can transmit and receive messages in both standard (11-bit) ID and extended (29-bit) ID formats.

III. Synthesis and Structure of a Control Algorhithm

In the six-legged walking robot, subject of the present study, for the purpose of creating antagonistic actors pneumatic muscles are put into use. The task to control a muscle-joint comprises several partial loops aimed at controlling the strength and pressure of the antagonistic actors as well as the angle's position of the joint. The only variables that can be affected by the close loops are the air flows going in and out of the muscle in three stages: enter-close-exit. The possibilities for controlling the speed of the in and out airflows through the pressure within every muscle of the joint as well as the position of the latter are described by the "equality of forces" of the muscles and the outside forces affecting the joint.

$$F_a + F_S \equiv F_p \tag{1}$$

Equation (1) shows that such a system needs to point out not only the joint's position but also the resultant force acting in the entire system. In the applied control system the input quantity is not the force but the max pressure $\boldsymbol{p_s}$. Subsequently, the joint can be controlled with the next sequence:

For certain values of $\boldsymbol{\theta}$ and $\boldsymbol{p_s}$ the required pressure for both muscles is calculated with equations (2) and (3):

$$\theta \geq \theta_{med} : p_a = p_s \cdot (\theta_p - \theta / \theta - \theta_a) \, , \, p_p = p_s \tag{2}$$

and for

$$\theta < \theta_{med} : p_p = p_s \cdot (\theta - \theta_a / \theta_p - \theta), \, p_a = p_s \tag{3}$$

Where $\boldsymbol{\theta_p}$ and $\boldsymbol{\theta_a}$ the minimal values of the joints angles in the opposing positions, while $\boldsymbol{p_p}$ and $\boldsymbol{p_a}$ are the pressure values in both muscles.

The actual value for $\boldsymbol{F_S}$ is calculated according to equation (4) with $\boldsymbol{\theta_n}$ being the true value of the joint's angle:

$$F_S = p_p \cdot \{(\theta_p - \theta_n) - p_a \cdot (\theta_n - \theta_a)\}/(\theta_p - \theta) \tag{4}$$

Recalculating $\boldsymbol{p_p}$ and $\boldsymbol{p_a}$ according to equations (2), (3) and (4): the presented relationship provides for carrying out the algorithm controlling airflows that enter and exit the muscle according to the joint's position $\boldsymbol{\theta}$ and the muscle's pressure $\boldsymbol{p}$.

Each leg's control can be devised in two stages. On the first stage the leg, without coming in contact with the surface, from the far back position, aka the Posterior Extrem-Position (PEP) moves forward to a front position, aka the Anterior Extrem-Position (AEP). This stage is known as passive phase, aka Protraction. In the next stage defined as active phase, aka Retraction, the leg is moved from a front position Anterior Extrem-Position (AEP) backwards to the Posterior Extrem-Position (PEP). During the active phase the leg touches the ground; the body is supported and pushed forward by the leg. The sequence of both phases is cyclically performed by each of the six legs.

For a successful walk the passive (Protraction) leg must rapidly move to the next constant position. On an even surface a precise positioning of the legs is not necessary. For the executing of the outlined stages following strategies are applicable:

• The active-leg-control is done as a time-function or as a function of another active leg.

• The passive leg is freely forwarded to ensure contact with the ground

The leg's movement in the passive phase must and in a way allowing the speed and body angle to steady. In case of unevenness the control height and perpendicularity axe of the body are compensated by the joints of the leg closing the passive phase. The switching from one supporting leg to the next must follow at a strictly defined moment.

Fig. 3 shows the control structure of a leg and respective joint.

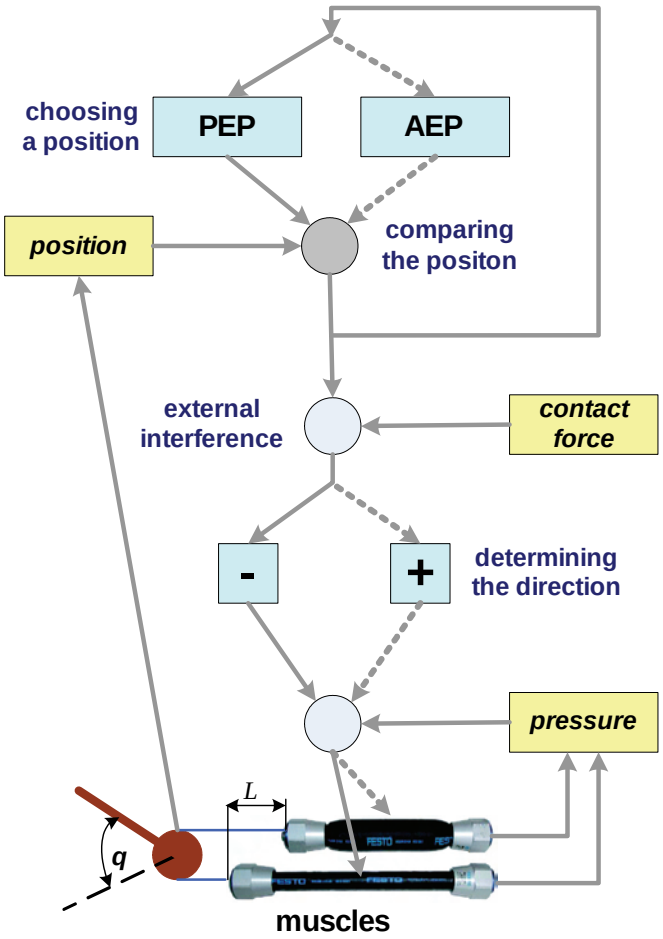


Fig.3. Model for control and regulation of the leg

The model comprises several control loops using current information about the joint's position, the pressure in the two muscles and the effort related to the force of contact with the surface.

IV. Experiments and Results

As main module of the structure a EVBR8C/23 microcontroller produced by the RENESAS company is applied. The module performs the basic functions of the close loop control, which depend on the information about the joint position and pressure within the muscles. On this level every sensor and actor (a pair distributed to every joint) are connected over an interface module, which contents scaling precise amplifiers for the signals coming from the joint's sensors; six electronic keys for the joints power distributors and three input PWM (Pulse Width Modulation) channels for each of the joints. For obtaining the PWM signals needed to control the muscles pressure altering speed, for every one of the three joints a separate programmable pulse with modulator is foreseen. The parameters of the PWM signals for every joint are given separately by the output ports of the microcontroller and the pressure values within the applied muscles are taken into account. The programmable pulse with modulator includes three programmable timers based on the free programmable logic FPGA.

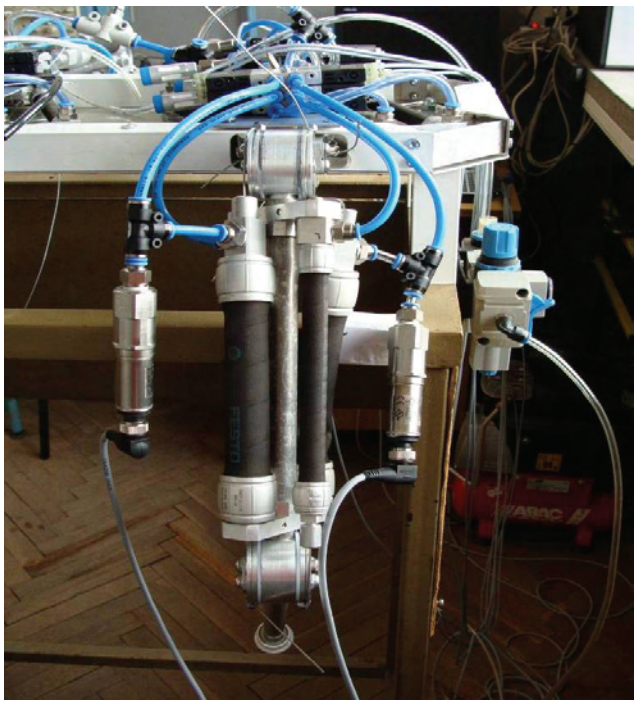


Fig.4. Experimental set

The experiment shown in Fig. 4 uses following components:

FESTO pneumatics:

- Valves type CPE10-M1BH-5/3G-GS6-B
- Inductors type GRU 1/8B with linear features
- Sensors for pressure type SDET-22T-G14-U-M12

Velleman Measuring equipment:

- Functional generator and transient recorder type PCSGU250
- Measuring and processing software Pc-Lab2000LT

The experiments carried out included:

Processing the valve under constant powering and inducting of the airflows: Fig. 5 and Fig. 6 show the change in pressure under that type of valves control.

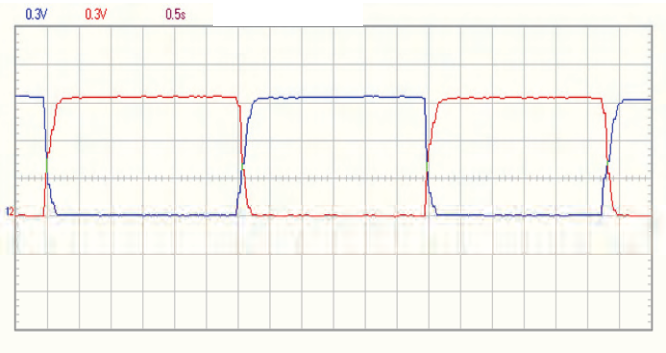


Fig.5. Without throttling

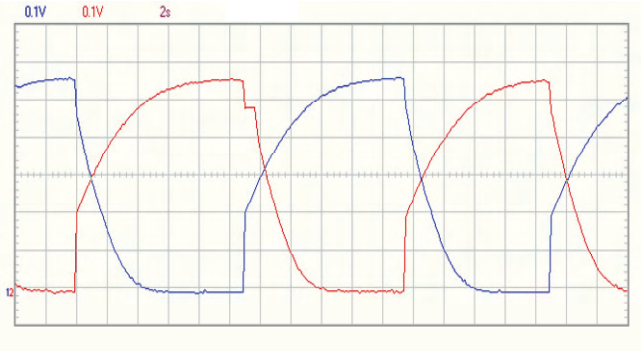


Fig.6. With throttling

The inducting shown in Fig. 6 goes in opposite directions. The shortage is in the non-linear features at the start of both processes and the uneven change of the in and out airflow over a certain time period.

Valve functioning with an altering power supply – PWM and using the data from the sensors for pressure and position: Figure 7, 8 and 9 display the change in pressure while controlling the valves over a width modulated signal.

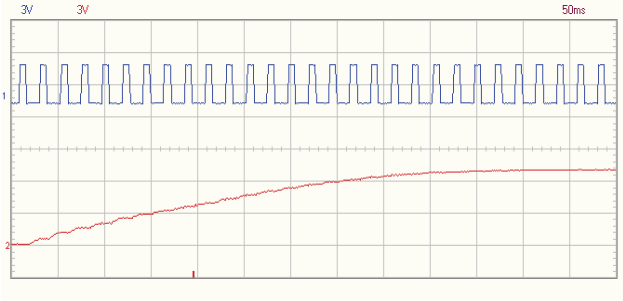


Fig.7. Entering flow, pulse width of 1 / 3T

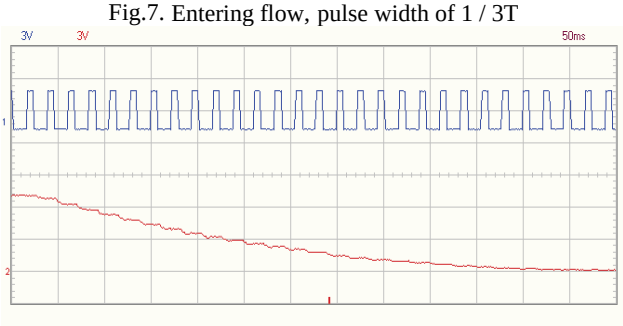


Fig.8. return flows, pulse width of 1 / 3T

In that case the pressure change in both directions is close to linear and with the same shape and duration. With changing the duration of the impulse the possibility to change the speed and reach the pre-given position arises. In Fig. 7 and Fig. 8it is 500 mS and in Fig. 9 – 200 mS.

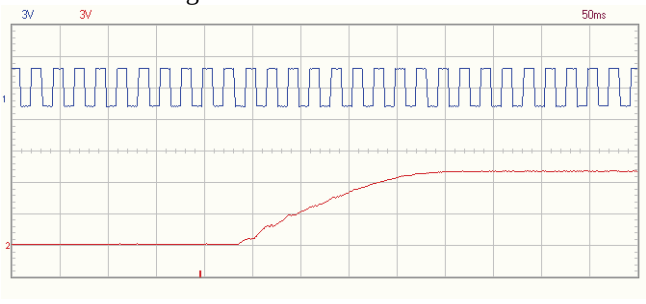


Fig.9. Enter flow, pulse width of 1 / 2T

V. Conclusion

In consequence of the experiments and measurements carried out as well as the analyses of the obtained parameters from practical tests, the accomplishing of a functioning valves control over a width modulated signal can be reported. The results obtained with the valves of the available type proves beyond doubt that the maximal frequency should not exceed 47Hz and the optimal value for the impulse lies between 18% - 34% of the duration, while the duration itself should not fall bellow 27mS.

Created is an algorithm for accounting of the pressure within both muscles – the protagonist and the antagonist respectively. Hence, we can correct their similar behavior in regard of the entering and exiting airflows, which is a condition for testing new control algorithms.

The further development and testing of the programmable width modulator (PWM) will provide for the realization of various speeds over PWM signals for the separate joints. That in turn would allow a better coordination of the leg as a whole.

Acknowledgement

Every research this paper is accounting for has been done under the assignment of Contract ВУ – ТН – 201/2006 entitled “Research of a Modular Architecture for the Control of Mechatronic Elastic Multi-Link Devices”

References

[1] O. Matsumoto, W. Ilg, K. Berns and R. Dillmann. Dynamically stable control of the fourleggedwalking machine BISAM in trott motion using foot force sensors. In prodeedings of the International Conference of Intelligent Autonomous Systems (IAS 6), p. 301-306. Venice, July 2000

[2] Cruse H., Dean J., Kindermann T., Schmitz J., Schumm M., (1999)”Walknet - a decentralized architecture for the control of walking behavior based on insect studies”, in: Hybrid Information Processing in Adaptive Autonomous Vehicles.(ed) G. Palm. Springer

[3] Frik M., Guddat M., Losch D.C., Karatas M. ”Terrain Adaptive Control of the Walking Machine Tarry II”. Proc. European Mechanics Colloquium, Euromech 375 - Biology and Technology of Walking, Munich, 1998, pp. 108-115.

[4] Ayers J., ”A Conservative Biomimetic Control Architecture for Autonomous Underwater Robots”, p 241- 260, Neurotechnology for Biomimetic Robots, Ed. Ayers, Davis and Rudolph, MIT Press, 2002

[5] H.-J.Weidemann, F. Pfeiffer, J. Eltze: The six-legged TUM walking robot., Intelligent Robots and Systems (IROS), 2004, Volume 2, 1026 -1033.

[6] T. Kerscher, J. Albiez, K. Berns, Joint control of the six-legged robot AirBug driven by fluidic muscles. Proceedings of the Third International Workshop on Robot Motion and Control, Poland, 2002.

[7] K. Berns, J. Albiez, V. Keppelin, C. Hillenbrand: Airbug - Insect-like Machine Actuated by Fluidic Muscle. 4th International Conference on Climbing and Walking Robots, 2001, 237-244

[8] A. Hildebrandt, O. Sawodny, R. Neumann, A. Hartmann, A Flatness Based Design for Tracking Control of Pneumatic Muscle Actuators, Seventh international Conference on Control, Automation, Robotics And Vision (ICARCV’O2), 2002.

Algorithms for Control of a Line Robot

M. Todorova[1], N. Nikolov[2], I. Penev[3]

***Abstract* – The paper introduces authors' research on movement algorithms of a line robot on a preset trajectory. A scale model of the robot based on a single-board microcontroller was created. The paper outlines the apparatus and devices as well as the implemented control algorithms.**

***Keywords* – robot, control algorithms, sensor, single circuit board microcontroller, motor.**

I. INTRODUCTION

The last few years saw a number of new highly technological products with an optimal life cycle developed in the area of mechatronics. Robots incorporate:

- precise mechanics and mechatronics;
- highest level of information technologies designed for processing digital and analogue data [1];

Robots feature intelligent action control depending on the kind of tasks that have to be accomplished. This can be achieved by receiving feedback from sensors.

The tasks of controlling a mobile object can be divided into several groups according to the methods and resources used for collecting information.

This paper studies algorithms for controlling a robot on a trajectory predetermined by means of a line.

II. TECHNICAL IMPLEMENTATION

The modular scheme of the technical implementation of the system is shown on fig. 1. The system uses the following modules:

- single-board microcontroller;
- data processing module from analogue sensors;
- data processing module from digital sensors;
- communication module (interface module);
- operator interface module;
- motor command module;
- real-time clock module;

There are two alternate ways of powering up the system – either by an external 9V source or by an accumulator battery. The external dimensions of the robot are 20x12 cm.

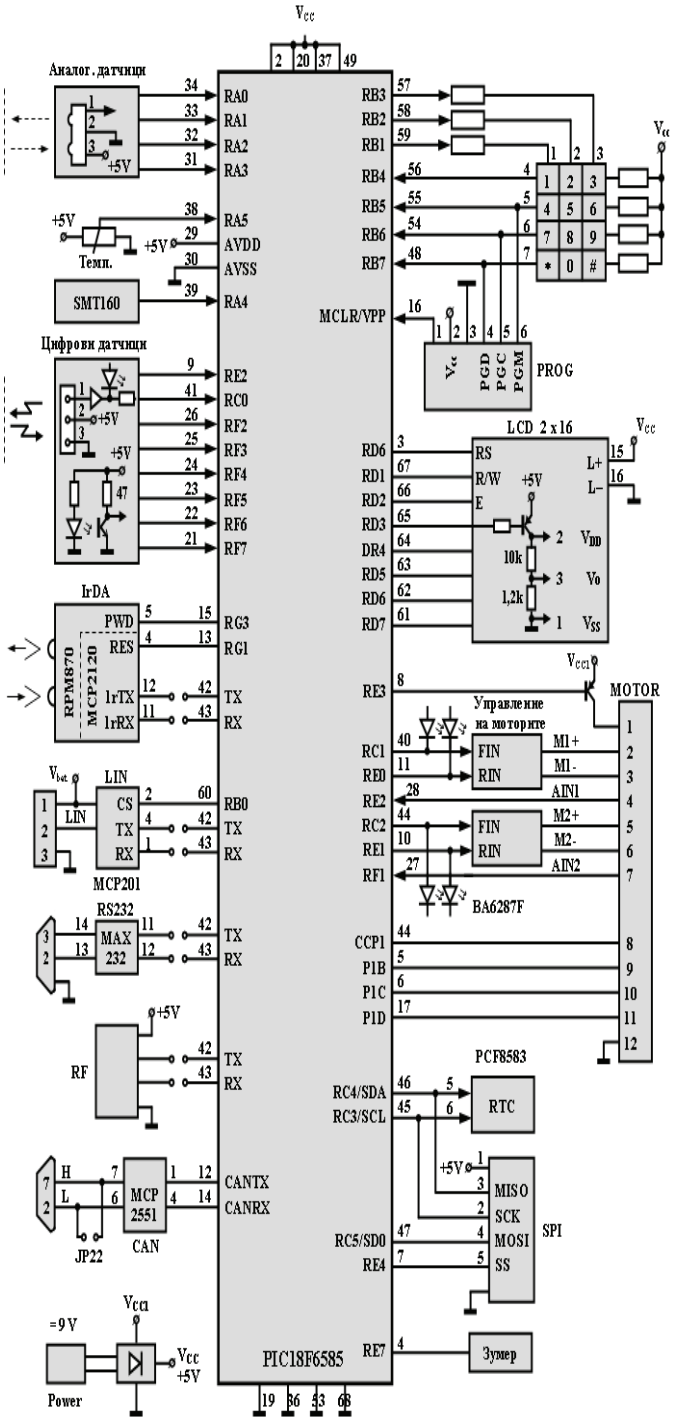


Fig. 1 Technical implementation

[1] Maya P. Todorova, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: mayasvilen@abv.bg

[2] Nedyalko N. Nikolov, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: ned.nikolov@tu-varna.bg.

[3] Ivaylo Penev, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: ivailopenev@yahoo.com

III. Control Algorithms of a Line Robot on a Preset Trajectory

The type of the robot control algorithm depends mainly on:
- the data processing modules from the analogue and digital sensors;
- the DC motor command module.

The sensors used in the system are analogue and ON/OFF sensors. The type and the number of the sensors as well as their location in the system determine the algorithm used for controlling the object as they provide information about the position of the object in relation to obstacles or the preset track. The platform utilizes two analogue sensors – one measuring the distance in front of the robot and another measuring the distance to the left. The sensors work on the principle of measuring reflected infrared light. The distance to the object is determined on the basis of reflected light readings.

When the track is marked by a line (black or white) the ON/OFF sensors come into play. The number of the ON/OFF sensors is five. They are placed in a straight line at 2 cm intervals 5 mm above the surface of movement. They work by measuring reflected infrared light recognizing the surface colour beneath the sensor. If it is white the reading of the sensor is 1 and respectively, the reading is 0 when the colour beneath the sensor is black.

The DC motor command module consists of two similar circuits based on the BA6287F chip. The chip allows the motors to function in one of the following modes:
- forward movement;
- stop motion mode;
- reverse movement;
- stand-by mode.

The single-board microcontroller controls the direction and speed of rotation of the motor. Alterations in the speed are achieved by means of signal with wideband impulse modulation generated respectively by PWM1 modules for the first motor and PWM2 modules for the second. The saturation coefficient of the impulse is programmable and proportionate to the rotation speed of the motor and therefore to the speed of movement of the robot.

A. Contour following algorithm based on special points

A sample contour of robot's movement is shown on Fig. 2.

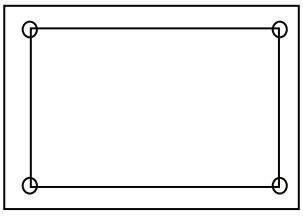


Fig. 2 Movement contour with special points

This particular movement contour is closed. It has a rectangular shape. The main difficulty in this algorithm is maintaining continuity of velocity and acceleration in the special points. The trajectory set by each side of the rectangle is reduced to a Bezier function with six supporting points. Starting and pulling off is achieved at zero velocity and acceleration. The problem is solved by doubling the supporting points. Each section is described by a 5^{th}-degree polynomial. Nevertheless, the movement follows a straight line. [2]

B. Straight line movement algorithm

The robot is set in motion by two DC motors with reducers. The classic design motor is controlled by changing the voltage. The momentum of the high speed revolution motor cannot be applied directly to the wheels. The use of reducers is indispensible. The control is accomplished by changing voltage. The need for corrections in velocity is induced by various disturbances as:
- unevenness of the track;
- wheels do no spin with the same speed, etc.;

These disturbances lead to diversions from the perfect trajectory of movement. The possible diversions are shown on Fig. 3.

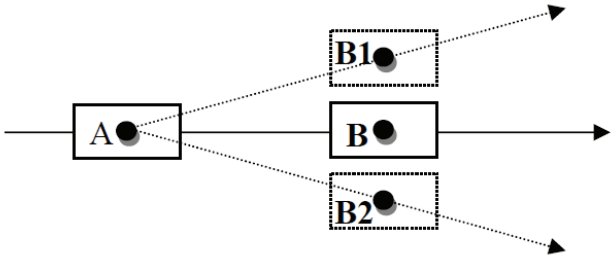


Fig. 3 Types of deviations

A smooth-surface track should be used in order to test the straight line movement of the robot. In the straight line movement the velocity back and forth is fixed. If diversions from the movement line occur, the algorithm is adjusted by increasing or decreasing the speed of one of the motors.
If the robot has to move with maximum speed, 5V current is supplied. If lower speed is desired, the voltage is decreased accordingly. For achieving average speed of movement the robot is supplied with 5V current at first and then the voltage is decreased to 0V. The voltage values [0, 5] corresponding to the robot's movement has to be represented using a table.

C. Movement along a line set trajectory

Labyrinth movement algorithms can be used for achieving movement along a preset trajectory. Information about object's orientation and position is received by reflective sensors placed on the bottom side of the robot. They are placed at 5 – 6 mm intervals. There are two possible scenarios:
- following a white line on a dark surface;
- following a dark line on a light surface.

The difference between the two scenarios is the choice of active value.

The arrangement of sensors is shown in Table 1.

TABLE 1
THE ARRANGEMENT OF SENSORS

L2	L1	C	R1	R2
1	1	0	1	1
Left		Centre		Right

The represented system comprises five information components. Their position in relation to the line is as follows:
- 1 1 1 1 0 → the line is in rightmost position
- 1 1 1 0 1 → the line is slightly to the right
- 1 1 0 1 1 → central position
- 1 0 1 1 1 → the line is slightly to the left
- 0 1 1 1 1 → the line is in leftmost position

The graph is shown in Fig.4

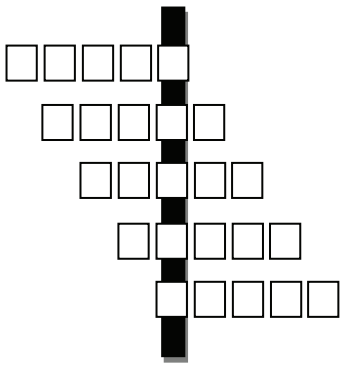


Fig.4 Position on five sensors

The sensor reads 1 when it is on a light surface and 0 when it is on a dark surface.

The received information is used for a pattern to define:
- the current position of the robot;
- subsequent actions (follow-up);

The presence of five sensors determines 32 possible combinations. Some of them are less likely to happen for the purposes of the current task. With inputs such as 0 0 1 1 0 ,1 0 1 0 1, etc. robot's behaviour is hardly determinable.

When the robot reaches a barrier the readings of the sensors change from „11011" to „11111". In this situation the robot has to stop.

IV. CONCLUSIONS

Practical testing of the proposed algorithms is carried out with two options available to reflective sensors - straight line and triangle. In both versions are achieved the objectives, as 32 commands in both cases are different. It is envisaged to situate markers in the form of white spots inside the wheel with directed at them additional reflective sensors. This will allow for feedback on speed and apply the algorithms for speed control of each of drive wheels, which will improve the management of the robot. A significant moment in the control algorithms is driving the robot to the final goal overcoming the existing obstacles in the environment.

REFERENCES

[1] Prof. Veselin Pavlov, TU Sofia, Design of robots - the integration of creative engineering and advanced training of "mechatronic engineers type" ,2009.

[2] V. Balavesov, V. Vecigyan, "Some examples on planning paths with Bezier functions "

[3] R. Dofr, R. Bishop, Modern Control System, Peason Education, 2008.

[4] N. Kenarov „PIC microcontrollers" Part 2- Varna 2006.

[5] www.microchip.com – PIC18F6585 Data Sheet – 68 pin High performance 64 KB Enhanced Flash Microcontroller with ECAN module – 2004

[6] www.pololu.com.

[7] Harris, Tom. "*How Robots Work*". http://science.howstuffworks.com/robot.htm.

[8] "Robot (technology)". *Encyclopaedia Britannica Online.* http://www.britannica.com/EBchecked/topic/505818/robot.

Analysis of Opportunities for Increasing Energy Efficiency of Electric Vehicle with Hydrogen Cell

Blagoy Burdin[1], Borislav Traykov[2], Dimitar Bojchev[3]and Zhelyazko Kartunov[4]

Abstract - **One of the main problems of electric vehicle is the great value of start power. In current paper is made an analysis of opportunities for increasing energy efficiency by using PID regulator. It is made an expression of estimation start power energy efficiency, based on practical measurements.The results are confirmed with practical experiments and represented with graphics.**

Keywords– **Fuel cells, PID regulator.**

I. INTRODUCTION

Important technical limitation a fuel cell system is a current (or power) slope of fuel cell must be limited in order to prevent fuel starvation phenomenon [1] [2]. In addition, for vehicle applications power drive train consumed high power in a short time in vehicle acceleration process (around two times of average power during drive cycle [3].

This paper present method to decrease energy consumption in acceleration vehicle process. The experimental results are obtained with measurements of fuel cell car made for Shell Eco-Marathon 2011 competition.

II. ESTIMATING START POWER (ENERGY)CONSUMPTION

The energy, required for electric vehicle starting can be estimating by consummated power vs. time The energy in starting process is given by Eq.1:

$$E_s = \int_{t_0}^{t_1} P(t)dt \tag{1}$$

E_s - Energy of starting process;

$P(t)$-Consumed power.

In current paper consumedpoweris represented by power series expansionEq.2:

$$\int_{t_0}^{t_1} P(t) = a_0 + a_1 p + a_2 p^2 + \ldots + a_n p^n \tag{2}$$

In considering case it is made a computer simulation of dependence between start power in function of time. It is used a numerical method with inverse matrix division. Using practical measurement results, shown in table 1 it is made a matrix Eq.3

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ .. & ... & ... & ... \\ 1 & \sqrt{M} & M & M^{\frac{3}{2}} \end{bmatrix} . [a_k]^{\cdot} = [P'_{di} + (i-1)]^{\cdot} \tag{3}$$

Aftersolvingthe Eq.(3) becomes:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ .. & ... & ... & ... \\ 1 & \sqrt{M} & M & M^{\frac{3}{2}} \end{bmatrix} \backslash [K'_{di} + (i-1)]^{\cdot} \tag{4}$$

The results are shown in Fig.1.It can be seen that for practical measurements in the purpose of current case the most appropriate polynomial order is 4-th.

TABLE 1: POWER CONSUMPTION, CURRENT OF VEHICLE ELECTRIC MOTOR AND BATTERY VOLTAGE VALUE.

t,s	0	0,5	1	1,5	2
I,A	80	45	36	34	32
U,V	22	23	24	24	24
P,W	1760	1035	864	816	768

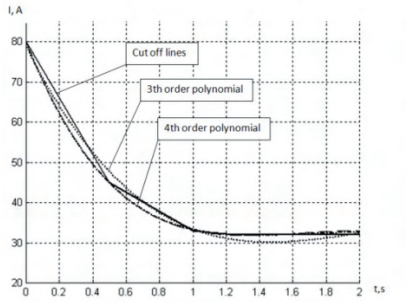


Fig.1 Computer simulation of dependence between start power in function of time

[1]Blagoy Burdin is PhD student with the Faculty of Transport at Technical university – Sofia, E-mail: b.burdin@abv.bg

[2]Borislav Traykov is associate professor with the Faculty of Transport at Technical university – Sofia, E-mail: btrykov@tu-sofia.bg

[3]Dimitar Bojchev – PhD at Technical university –Sofia

[4]Zhelyazko Kartunovis PhD student with the Faculty of Transport at Technical university – Sofia, E-mail: g_kartounov@tu-sofia.bg

Bearing in mind the above the start energy can be calculated by substitution of (2) in (1):

$$E_S = \int_{t_0}^{t_1} \left(a_0 + a_1 P\ (t) + a_2 P^2(t) + a_3 P^3(t) + a_4 P^4(t)\right) dt \quad (5)$$

Where $t_0 = 0$.

After calculation the expression becomes:

$$E_S = a_0 + \frac{a_1}{2} t_1^2 + \frac{a_2}{3} t_1^3 + \frac{a_3}{4} t_1^4 + \frac{a_4}{5} t_1^5 \quad (6)$$

A. Theoretical analysis of energy efficiency.

The relatively estimation achieves energy efficiency can be computing by the next Eq:

$$\rho = \frac{E_s - E_s'}{E_s}.100,\% \quad (7)$$

For increasing energy efficiency is applied Proportional integralderivative control regulator (PID). The regulation law can be express according Eq.8.

$$U(t) = k_p.\Delta(t) + k_I.\int \Delta(t) dt + k_d \frac{d\Delta(t)}{dt} \quad (8)$$

Where Kp- proportional coefficient;

Ki – international component;

Kd- derivative component.

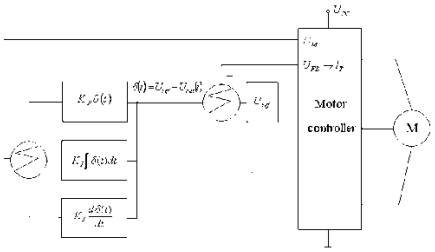


Fig.2 Functional scheme of used PID regulator

TABLE 2: THE REGULATOR COMPONENTS, ADJUSTED USING ZIEGLER- NICHOLS CLOSED LOOP TUNING METHOD.

Controller	K	T_i	T_d
P	$0,5.K_u$	-	-
PI	$0,4.K_u$	$0,8.T_u$	-
PID	$0,6.K_u$	$0,5.T_u$	$0,125.T_u$
20% over adjusting	$0,2.K_u$	$0,5.T_u$	$0,125.T_u$

where Tu – continues oscillations period;

Ku–ultimategain.

The functional scheme of regulator is shown in fig.2. The regulator components are adjusted using Ziegler- Nichols closed loop tuning method (table2).

B. Practical results.

It is made a practical experiment before and after applying the PID regulator. Using the numerical results it is made a 4-th order polynomial interpolation Fig.3.

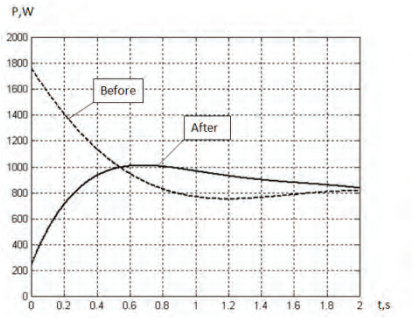


Fig. 3 Consumed power vs. time before and after using adjusted PID regulator

The real energy efficiency is calculated using (7)and represents anenergy savingsbetweenthe twointegrated areasofpower shown in Fig. 3.

The results of experiment shows that using the PID controller the start energy efficiency can be increased at least 8%.

Applying of PID controller is an real approach to increase energy efficiency of the fuel cell vehicle during the start mode. The same approach can be applied to increase energy efficiency in other driving modes of the vehicle.

III. CONCLUSION

Applying of PID controller is an real approach to increase energy efficiency of the fuel cell vehicle during the start mode. The same approach can be applied to increase energy efficiency in other driving modes of the vehicle.

ACKNOWLEDGEMENT

The results are obtained under a project funded by the grant for scientific research in TU - Sofia.

REFERENCES

[1] P.Thounthong, Stph. Rael andB. Davat “Fuel Cell and Super Capacitors for Automotive Hybrid Electrical System”, ECTI Transactions on electrical eng., electronics and communication Vol.3, No1 Feb. 2005.

[2] J. Larminie and Andrew Dicks, "Fuel cell systems explained", second edition, 2003.

[3] K.J. Kelly and A. Rajagopalan, “Benchmarking of OEM Hybrid Electric Vehicle at NREL: Milestone Report”, U.S. Department of Energy (DOE), Contact No. DE-AC36-99-GO10337. Aug. 2001.

Increasing the Efficiency of Warehouse Operations Applying the RFID Technology

Siniša Sremac[1], Ilija Tanackov[2], Jovan Tepić[3], Gordan Stojić[4] and Saša Krstanović[5]

***Abstract* – Many companies use sophisticated systems to manage goods in the warehouse, integrated with supply chain systems, enterprise systems and electronic data interchange. The main objective of Radio Frequency Identification (RFID) is to enable automatic identification of products regardless of their origin and purpose, and the free circulation and free movement in all links in the supply chain. The use of this technology for automatic identification of goods in the warehouse and updating status has the positive influence on the parameters of the logistic processes such as: speed performance, the validity of recorded data, the efficiency of the flow of goods, increasing of reliability, operational flexibility, etc.**

***Keywords* – Automatic Identification, RFID, Supply Chain, Warehouse.**

I. INTRODUCTION

Supply chain can be defined as a synchronized and isochronous implementation of the transport, handling and storage operations to ensure the flow of goods from the sender to the receiver [1]. A storage operation is an essential link in the supply chain. This paper will demonstrate one of the ways to increase the storage efficiency, and therefore the supply chain performance.

RFID technology is a technology for automatic identification that has a very wide application in management of flows of goods in the warehouses. It represents the Automated Data Collection (ADC) system that enables businesses to wirelessly capture and move data using radio waves [2]. The main objective of application of this technology is to enable automatic identification of products regardless of their origin and purpose, and free circulation and free movement in all links in the supply chain. According to previous researches, cost savings of 8 to 30% out of the total logistics costs can be realized in the supply chain through the use of RFID technology [3].

There are some individual examples of using this technology with the aim of increasing the efficiency of storage. HP logistic centre in Memphis has been monitored. Applying RFID technology eliminates the need to work with bar codes [4]. Time for planning logistics processes has been significantly reduced from several minutes to a few seconds, which reduced storage costs. Study of Auto-ID Center [5] showed that the implementation of RFID technology can reduce inventory loss by 10%.

[1]Siniša Sremac is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: sremacs@uns.ac.rs.

[2]Ilija Tanackov is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: ilijat@uns.ac.rs.

[3]Jovan Tepić is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: jovan.tepic@uns.ac.rs.

[4]Gordan Stojić is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: gordan@uns.ac.rs.

[5]Saša Krstanović is with the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, E-mail: sasa.krstanovic@gmail.com.

II. RFID SYSTEM

Radio frequency identification is an automatic, wireless identification technology capable of gathering data without human intervention and line of sight requirement [6].

The basic elements of the RFID system are: antenna, tag, reader and data processing system (Fig. 1) [7]. Communication between the tag and reader in the product is done wirelessly, at exactly the right frequency. Reading speed and data transfer speed are associated with the frequency. As the frequency is higher, the data transfer is faster. A very important factor in planning the RFID system is the amount of data transferred, especially when tag passes quickly through the readings zone. Other important factors in planning the system are the reader power and interference generated by the facilities and equipment around (especially metal).

Fig. 1. RFID system

III. ASSUMED SYSTEM

Each store has its unique layout, organization and structure. RFID technology meets these challenges with its modular and flexible architecture, on the basis of which it can be adapted to any storage structure [8]. Application of RFID technology in warehouse operations include the following:

1. *Reception* - It is possible to set a stationary RFID reader in a stock to control the entry of goods. Each passage of goods through the entrance activates the reader that reads the goods (Fig. 2). This allows the automatic reading of the products movement and updating the database of the warehouse.

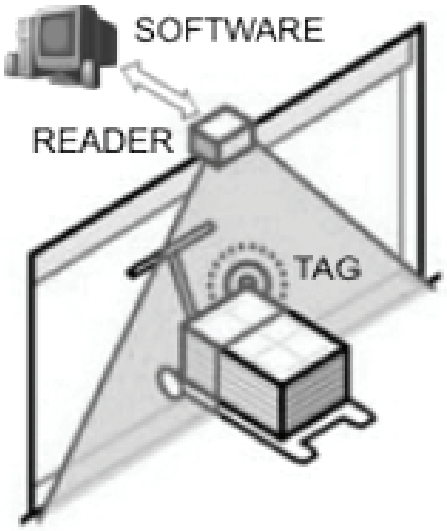


Fig. 2. Reading of goods at the entrance

2. *Disposal* - The correct identification of the stored goods is essential for efficient storage management. If the user has not selected a location for storing, the software selects the best possible location based on the criteria.
3. *Picking* - Takes place according to the orders issued directly from the system or the accounts taken from the system for ordering. Different principles are used: First In - First Out (FIFO), Last In - First Out (LIFO), First Ended - First Out (FEFO), according to shelf life, the quantity and others.
4. *Wave picking* - Efficient determination of the location of collection of goods is caused by a high degree of optimization process based on: the selection of routes, the planning time required for the process of collecting, different types of orders and the realization order priorities.
5. *Packing / dispatch* - The system proposes the most effective form of packaging for any item, calculates the optimal weight and pack size for the selected transmission method. Places for packing / dispatching equipped with RFID readers increase security and speed up the verification of orders. When the loading is finished, the system may exchange information with courier services, transport operators and shipping companies, to make supply chain more efficient (Fig. 3).
6. *Control interface* - Graphical management interface based on the "web" technology makes all relevant information be available in real time. There is a possibility of finding a bottleneck before it becomes a critical problem in the warehouse. All key logistic parameters for measuring logistics performance and parameters defined by users can be tracked and compared with the planned values.
7. *Task & resource menagement* - Is especially important in combination with a networked computer system. Task & resource management ensure appropriate action to be executed and the adequate resources to be allocated at the right time.

Using RFID technology at the entrance and exit, the process of collecting data on input and output flows in the warehouse is completely automated. All data on the quantity and type of a product as well as the time of entry and exit from the warehouse is automatically read from the tag. Also, any movement of goods within the warehouse is automatically registered and loaded into the system for data processing.

A. Advantages

The benefits of using RFID technology in relation to other information technologies in the identification of products in stock are [9]:

- contactless data transmission,
- relatively large distance between the tag and reader,
- high speed and accuracy in data reading,
- storing more information on the tag,
- the possibility of subsequent writing data,
- simultaneously reading of multiple tags and
- fewer errors in identification and data entry.

Packages are read regardless of orientation and without direct visibility between the product and a reader. Tags can be exposed to dirt, heat, humidity and pollution that other technologies (e.g. bar code) make useless.

RFID technology is very fast. Tag can be read and feedback is available in milliseconds [10]. Readers read multiple tags at once, which is much faster and more efficient than other technologies.

One of the great advantages is the reading distance and low energy consumption. Reading distance is affected by many growth factors, such as: type of tag, electromagnetic noise, the orientation of tag, type of antenna and legislation. The current reading distance ranges (for some types of tags) from a few tens of centimeters to several hundred meters.

There are tags with different memory capacity. They are selected according to the desired application and the necessary amount of information that is stored in them.

The participants in the supply chain can use RFID systems for applications of different purposes and carefully planned systems can use the same tags in order to reduce costs of implementation.

RFID technology is designed as a simple replacement of bar codes [11]. However, replacement of bar codes is not going at a satisfactory pace and that process is not simple. There are still few companies that use RFID technology and those are mainly the biggest ones. The main reason for the relatively slow proliferation of RFID technology in modern storage systems is that the bar code technology is in mass use in product identification since the late 60s of the last century. During this time, it is widely implemented in all business information systems, particularly systems for production management, warehousing, distribution, transportation and trade. Therefore, the entire philosophy of storage and other systems designing, which was based on the possibilities and advantages of bar code technology for years, needs to be changed now.

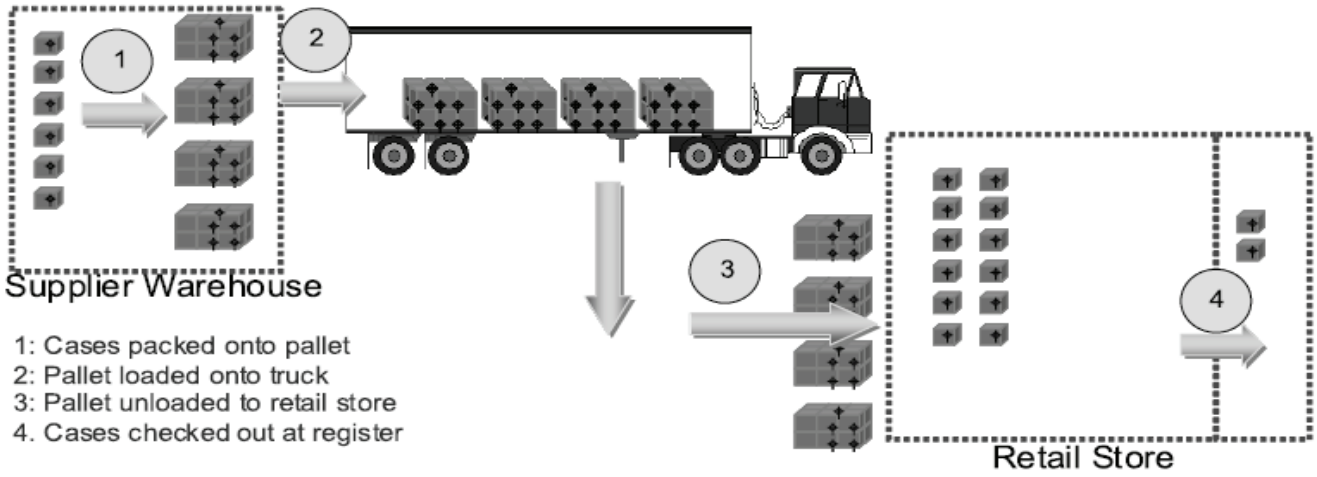


Fig. 3. Application of RFID technology in the warehouse as part of the supply chain [12]

B. Disadvantages

Currently the biggest disadvantages of RFID technology in relation to other information technologies for automatic identification of products are:

- high price,
- lack of desired tags,
- radio interference,
- undeveloped awareness of users,
- lack of privacy and
- data security.

The cost of tags, the cost of integrating RFID technology into existing storage systems, readers and software maintenance costs, and the costs of employees training make the introduction of this technology a significant investment. Prices of tags that are embedded in products range from a few tens of cents to several hundreds of euros, depending on the type of the tag you wish to use [2]. Also, the cost of supporting infrastructure is still very big.

The principle of reading tags is not based on optics, as with bar codes, but it requires a clean environment in order to make contact and read the tag. Radio signals from the aspect of the environment in which they work and the applied frequency, can be disturbed by generating or absorbing in the obstacles [13].

It is clear that a higher degree of automation and efficiency of storage is achieved applying RFID technology, but there are dangers related to data security [14]. Great possibilities open a number of potential abuses, too, especially if more information is stored in the tags.

In addition to these shortcomings, the expansion of RFID technology is still expected in the automatic identification market in the future. There is an evident increase in its application not only to the storage systems and supply chains, but in other industries as well.

C. The Expected Effects of the Warehouses Operations

The advantages of using RFID technology in warehouse include: total visibility and complete history of the stock, reducing the amount of inventory in the warehouse, facilitating the delivery of "just in time", the complete process of products control, reduced preparation and loading time, improved sorting and selecting products, reducing storage space, increased security, reduced number of errors and lower overall operating costs. [15]

Applying technology for the automatic identification and updating the situation in the warehouse, has the positive influence on parameters of processes such as the speed of implementation, the validity of recorded data, the efficiency of cargo flows, increased reliability, operational flexibility, etc.

High reliability of storage of goods, excellent control and tracking system increase the quality of any logistics operations and contribute to the delivery of products from the warehouse without any errors. Time and cost counting on entering the store are reduced, by extracting the data automatically which eliminates the need for manual intervention [16].

Readers, covering storage shelves and other storage places, can automatically record the removal of the unit and update the database. If a unit was placed in the wrong place or is it is necessary to complete the order urgently, fixed readers or workers with mobile readers can automatically find the product [17]. Products are ensured against theft by readers which have an alarm or they send a notice if the unit is moved without permission.

In the warehouse operations the following effects are achieved:

- more efficient flow control in the warehouse according to the principle of storage: FIFO, LIFO, FEFO, shelf life and quantity,
- formation of a unified system of coding, marking and identification products,
- business automation,
- processing of orders and dispatch,

- efficient communication between different business partners in all links of the supply chain,
- increased productivity,
- complete products control and monitoring from the entrance to the exit of the warehouse,
- costs reducing,
- saving time,
- improving production,
- increased reliability and
- improving services and meeting customer requirements.

IV. CONCLUSION

Application of RFID technology can be imagined in any area of human activity where the results are present. Currently it is the most widely used in transportation and logistics, manufacturing and control. More and more companies integrate RFID technology into the warehouse management system. These systems are equipped with specially designed equipment. It's a big initial investment, but in the long period of time the efficiency of storage is greatly increased and costs are reduced.

Application of the proposed RFID system a high level of automation in the selected points is achieved. Entering and updating information on admission and shipping from the warehouse is fully automated. In all these processes in the warehouse, the level of human interaction is kept to a minimum. This method enables significant acceleration of procedures of collecting and processing data related to the storage process.

RFID technology provides completely new possibilities and benefits. In order to use them properly and to get the expected effects, it is necessary to change the way of the information systems design in logistics-distributive and commercial centers. The above activities require the adequate expertise and available resources.

The next step towards improving the automation of all supply chains would be connecting storage systems for data processing with the other participants systems in the supply chain by a corresponding information system. Such integrated compact system would enable fast and qualitative data and information exchange necessary for achieving a high level of service quality.

ACKNOWLEDGEMENT

This paper is a contribution to the Ministry of Science and Technological Development of Serbia funded by project TR 36030.

REFERENCES

[1] I. Tanackov, G. Stojić, "Logistika", Fakultet za poslovni menadžment, Bar, 2008.
[2] M. Lazarević, "Razvoj modela za upravljanje proizvodima u toku životnog veka primenom RFID tehnologije", Doktorska disertacija, Fakultet tehničkih nauka, Novi Sad, 2009.
[3] A. Angerer, "The Impact of Automatic Store Replenishment Systems on Retail", PhDthesis, The University of St. Gallen, Graduate School of Business Administration, Dissertation no. 3123, 2005.
[4] S. d'Hont, "The Cutting Edge of RFID Technology and Applications for Manufacturing and Distribution", Texas, 2001.
[5] Auto-ID Center, EPC Information Service, White Paper, 2004.
[6] V. Simić, B. Dimitrijević, "Innovation in Information Management Using RFID Technology Throughout Product Life Cycle", XLV International Scientific Conference on Information – ICEST 2010, Ohrid, Macedonia. pp. 303-306, 2010.
[7] www.rfidjournal.com
[8] S. Chawathe, V. Krishnamurthy, S. Ramachandran, S. Sarma, "Managing RFID Data", Conference Proceedings 30st International Conference on Very Large Data Bases (VLDB), pp. 1189-1195, Toronto, Canada, 2004.
[9] S. Sremac, I. Tanackov, G. Stojić, J. Tepić, "Informacione tehnologije za automatsku identifikaciju robe u transportu", 18. Telekomunikacioni forum TELFOR 2010, Beograd, 2010.
[10] I. Šenk, "Kontrola pristupa i RFID tehnologija", Infoteh-Jahorina, vol. 8, ref. E-III-14, pp. 580-584, 2009.
[11] F. Zoëga, "Review of the current state of Radio Frequency Identification (RFID) Technology, its use and potential future use in Construction: RFID in Construction", Danish Technology Institute, 2006.
[12] F. Wang, P. Liu, "Temporal management of RFID data", Conference Proceedings 31st International Conference on Very Large Data Bases (VLDB), pp. 1128-1139, 2005.
[13] P. Jones, C. Clarke-Hill, P. Shears, D. Comfort, "Radio frequency identification in the UK: opportunities and challenges", International Journal of Retail&Distribution Management, vol. 32, no. 3, pp. 164-171, 2004.
[14] S. Sarma, S. Weis and D. Engels, "RFID systems, security and privacy implications", Technical Report MIT-AUTOID-WH-014, AutoID Center, MIT, 2002.
[15] RFID for asset tracking and inventory management for warehousing and retail, www.gaorfidassettracking.com
[16] M. Rakić-Skoković, G. Ostojić, M. Lazarević, "Primena RFID tehnologije u automatizaciji skladišta", Infoteh-Jahorina, vol. 8, ref E-IV-2, pp. 533-537, 2008.
[17] G. Ostojić, M. Lazarević, V. Jovanović, "Primena RFID tehnologije u industrijskim sistemima", Infoteh-Jahorina, vol. 5, ref. C-7, pp. 189-192, 2006.

Analytical and Simulation Performance Result Analysis for Parallel M/M/1 Queuing System

Ruben Nuredini[1], Zoran Gacovski[2] and Jasmin Ramadani[3]

Abstract – **In this paper we display and compare the performance results, in terms of average system time, for a queuing system consisted of multiple M/M/1 subsystems. The analytical results are obtained by calculating the system time according to the commonly used formula for system time and the simulation results are obtained by simulating this system using MATLAB toolbox - SimEvents. The basic idea for this system is to simulate the behavior and performance of toll plaza. Therefore, the basic characteristics of the system are that the servers are identical and the switching of the entities through the queues is equiprobable. The point of this paper is to identify and calculate the deviation of the obtained results.**

Keywords – **Simulation, parallel M/M/1 queuing system, SimEvents, system time.**

I. Introduction

The inspiration for this paper came from real-life toll plazas. The area around the toll-plaza can be considered as a system of several M/M/1 queuing subsystems each one consisting of the traffic lane as a queue and the tollbooth as the server. Assuming that all drivers are rational, they always queue up to the shortest queue which results in queues of equal length before every tollbooth. The total incoming traffic flow (Φ) is evenly distributed and therefore the arrival rate in each subsystem depends on the number of tollbooths (N) which implies $\lambda = \Phi/N$. Another assumption is that all of the tollbooths apply identical manner of toll collection which means that they all have same service rate (μ). The formula for average system time in M/M/1 queuing system is [1]:

$$T = \frac{1}{\mu - \lambda} \quad (1)$$

II. Constants

Our analytical results areobtained for constant values for the service rate and total traffic flow. The service rate μvaries depending on the technology that is used for toll collection. According to [2], the service capacity of several most used toll collection technologies are displayed in Table 1.

TABLE I
SERVICE CAPACITY OF TOLL COLLECTION TECHNOLOGIES

Person distributes change, issues receipt etc.	350 vph
Person distributes commuter tickets	500 vph
Automatic coin machine, regular coin only	500 vph
Automatic coin machine: some coins, tokens	650 vph
Mixed mode, any of the above plus ETC	700 vph

The arrival rate λdepends on the total traffic flow of the highway and the number of tollbooths. According to [3], the maximum traffic flow per lane is 2000 vph while the usual number of lanes can range from 1 to 6. The number of tollbooths is usually double from the number of lanes.

III. Calculations

We consider a system with μ=350 vph and Φ=1200 vph to be our sample system. By changing the number of tollbooths (N) and applying (Eq.1.) we obtained the analytical results displayed in Table 2. The system times are displayed in seconds.

TABLE II
ANALYTICAL RESULTS

N	4	5	6	7	8	9	10
T	72	32,73	24	20,16	18	16,62	15,65

It is obvious that by increasing the number of tollbooths, the arrival rate decreases, which causes lower system times.

IV. Simulations

The simulation results are gained from a simulation model built in SimEvents. SimEvents extends MATLAB toolbox Simulink with tools for discrete-event simulation of the transactions between components in a system architecture [4]. Our model is consists of several interconnected blocks from the SimEvents Block Library (Fig.1.). The simulation of the events starts with a Time-Based Entity Generator which

[1]Ruben Nuredini is master candidate on the Faculty of Information-Communication Technologies at FON University,Vojvodina blvd., 1000, Skopje, Republic of Macedonia E-mail: ruben.nuredini@fon.edu.mk

[2]Zoran Gacovski is with the Faculty of Information-Communication Technologies at FON University, Vojvodina blvd., 1000, Skopje, Republic of Macedonia E-mail: zoran.gacovski@fon.edu.mk

[3]Jasmin Ramadani is master candidate on the Faculty of Information-Communication Technologies at FON University, Vojvodina blvd., 1000, Skopje, Republic of Macedonia E-mail: jasmin.ramadani@fon.edu.mk

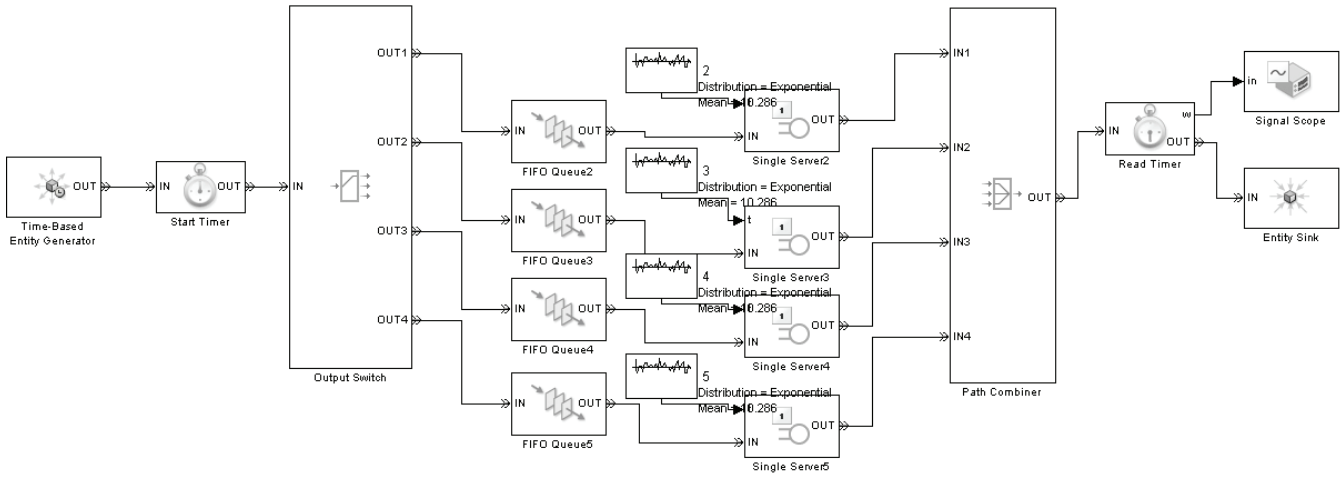


Fig. 1. The simulation model

generates entities every 3 time units (seconds) based on 1200 vph traffic load. The entities are then equiprobable distributed using an Output switch block through the FIFO Queue blocks and then conducted to the associating Single server block. Each of the single servers blocks has service time that corresponds with the 350 vph service rate of the tollbooth. The number of the existing Queue/Server pairs in the model depends on the number of tollbooths the simulation is taken for. After the completion of the service, the entities are combined with a Path Combiner block and then destroyed into an Entity sink block. The system time is measured by using Start timer block positioned before the Output switch and Read timer block positioned after the Path combiner block. The visual results of the plot acquired with the signal scope blockfor simulation with duration of 30000 time units are displayed in (Fig.2.) to (Fig.5.).

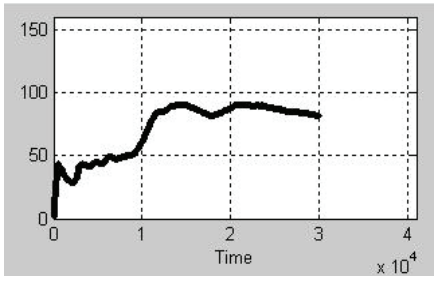


Fig. 2. Simulation results for N=4

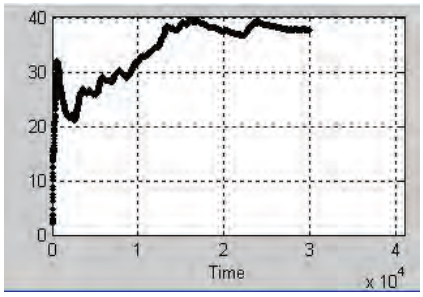


Fig. 3. Simulation results for N=5

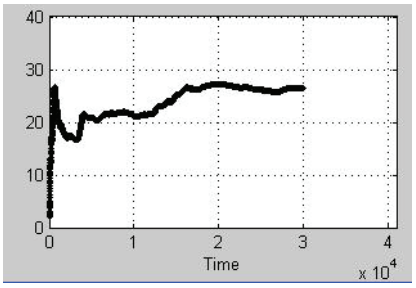


Fig. 4. Simulation results for N=6

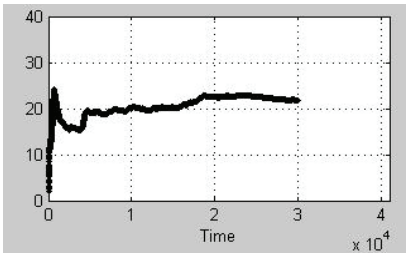


Fig. 5. Simulation results for N=7

V. Conclusion

It is obvious that the simulation results tend to get close to the values obtained by analytical calculations. That means that this simulation model is authentic because it verifies the analytical results.

References

[1] N.Chee-Hock, S.Boon-Hee, *"Queuing Modelling Fundamentals With Application in Communication Networks"* 2nd Edition,John Wiley & Sons Ltd., pp.111, 2008.

[2] P.Prevendouros, *"Automated tolls for Greece: System review and performance with AVC"*,Department of civil engineering, University of Hawaii at Manoa

[3] E.James L., *"Traffic Engineering Handbook"*, Pline, Prentice Hall, Englewood Cliffs, NJ, 1992.

[4] The MathWorks Inc., *"R2010b Documentation → SimEvents"*, The MathWorks Official Website, 2011

A Parametric Identification Approach Based on the Final Value Theorem of the Laplace Transform

Milica B. Naumović[1] and Lazar Popović[2]

Abstract –**The paper describes the application of the finite value theorem of the Laplace transformin the identification procedure of a continuous-time parametric model. The presented identification approach involves only a number of integrating processes and is suitable for easy automation. Two examples are used to demonstrate the merits of the proposed identification algorithm.**

Keywords– **Identification, Step response, System modelling, Parameter estimation, Numerical simulation.**

I. INTRODUCTION

An important step in process control is the identification of a "suitable" model of a continuous-time system from real observations [1]. What suitable means depends primarily on the concrete application one has in mind.Thus, it is necessary to select an appropriate level of model complexity depending on the purpose of system identification.Moreover, the acceptance of models should be guided by "usefulness" rather than by "truth".

Many identification methods discuss the parameter estimation problems both of continuous-time and discrete-time system models.A detailed overview of such methods is given in Ljung [2]. Least-squares, step response, and frequency response methods are representative as deterministic off-line identification approaches.

The system identification techniques based on the continuous-time model were initiated in the middle of the last century [3], but, for some time, were overshadowed by the overwhelming developments in discrete-time methods. This was mainly due to the "go completely digitally" trend that was the result of the paralleldevelopment in digital computers.

This paper studies a deterministic off-line identification method ofthe rational transfer function which can be performed by using the data of a constant steady-state output step response. Such identification methodology, known as transient response analysis, is simple to apply and understand, and often providesonly information good enough for the estimates of the input-output gain, the dominant time constants, as well as the time delays. These properties make the methods suitable for the first-stage of the analysis to prepare for the other experiments in the system identification.

The paper is organized as follows.Section 2 introduces some preliminary facts before a simple identification method is presented. The properties of the identification procedure are summarized in Section III and illustrated by several simulation results.Finally some conclusions of the work are presented in Section IV.

[1]Milica B. Naumović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milica.naumovic@elfak.ni.ac.rs

[2]Lazar Popović is with the Tagor Electronic, Tihomira Brankovića 21, 18000 Niš, Serbia, E-mail: lazar985@gmail.com

II. THE PARAMETRIC IDENTIFICATION METHOD

In this section, some assumptions are made for the plant to be identified, and an identification approach is considered.

A. The Plant to be Identified

A single-input single-output linear dynamical system can be described by a time-invariant linear differential equation

$$a_n c^{(n)}(t) + \cdots + a_1 c^{(1)}(t) + c(t) = b_m u^{(m)}(t) + \cdots + b_0 u(t) \qquad (1)$$

where $u(t)$ and $c(t)$ denote the input and output of the system, t represents time variable, and $c^{(i)}(t) \overset{\text{def}}{=} \mathrm{d}^i c(t)/\mathrm{d}t^i$, $u^{(i)}(t) \overset{\text{def}}{=} \mathrm{d}^i c(t)/\mathrm{d}t^i$. Coefficients a_i and b_i are the parameters of the system. Under the zero initial values, taking the Laplace transform to both sides of (1) and using the differential property,we can obtain

$$\left(a_n s^n + \cdots + a_1 s + 1\right) C(s) = \left(b_m s^m + \cdots + b_1 s + b_0\right) U(s) \qquad (2)$$

where $C(s)$ and $U(s)$ are the Laplace transforms of $c(t)$ and $u(t)$. Hence, we have the transfer function of the system

$$H(s) = \frac{C(s)}{U(s)} = \frac{b_0 + b_1 s + \cdots + b_m s^m}{1 + a_1 s + \cdots + a_n s^n}. \qquad (3)$$

Assume that $u(t)$ is a step function with the amplitude U_0.

Starting from the unit step response

$$\eta(t) = \frac{c(t)}{U_0}, \qquad (4)$$

by using the final value theorem of the Laplace transform

$$c(\infty) = \lim_{t\to\infty} c(t) = \lim_{s\to 0} sC(s), \qquad (5)$$

the coefficients of the model transfer function

$$\hat{H}(s) = \frac{\hat{b}_0 + \hat{b}_1 s + \cdots + \hat{b}_m s^m}{1 + \hat{a}_1 s + \cdots + \hat{a}_n s^n} \qquad (6)$$

can be determined.

B. The Algorithm of the Proposed Identification Method

Note that the method is based on the area determination and the graphical interpretation of the first step of the algorithm is shown in Fig. 1. According to the final value theoremofthe Laplace theorem, we have

$$\lim_{t\to\infty}\eta(t)=K_0=\lim_{s\to 0}\hat{H}(s)=\hat{b}_0\,, \tag{7}$$

or

$$\hat{b}_0=K_0\,, \tag{8}$$

where K_0 is indicated in Fig. 1.

It is suitable to define the integral $\eta_1(t)$ as follows [4]

$$\eta_1(t)=\int_0^t\left[K_0-\eta(\tau)\right]\mathrm{d}\tau\,, \tag{9}$$

and correspondingly the function

$$\hat{H}_1(s)=\frac{1}{s}\left[K_0-\hat{H}(s)\right]=\frac{1}{s}\left[K_0-\frac{\hat{b}_0+\hat{b}_1 s+\cdots+\hat{b}_m s^m}{1+\hat{a}_1 s+\cdots+\hat{a}_n s^n}\right]$$
$$=\frac{1}{s}\cdot\frac{\left(K_0-\hat{b}_0\right)+\left(K_0\hat{a}_1-\hat{b}_1\right)s+\left(K_0\hat{a}_2-\hat{b}_2\right)s^2\cdots+K_0\hat{a}_n s^n}{1+\hat{a}_1 s+\cdots+\hat{a}_n s^n}. \tag{10}$$

The value

$$K_1=\int_0^\infty\left[K_0-\eta(\tau)\right]\mathrm{d}\tau \tag{11}$$

can be calculated as it is indicated in Fig. 1. Than, the final value theorem of theLAPLACE transform gives

$$\lim_{t\to\infty}\eta_1(t)=K_1=\lim_{s\to 0}\hat{H}_1(s)=K_0\hat{a}_1-\hat{b}_1\quad. \tag{12}$$

The next step of integrating leads to the function $\hat{H}_2(s)$, the integral $\eta_2(t)$ and its limit value K_2, as follows:

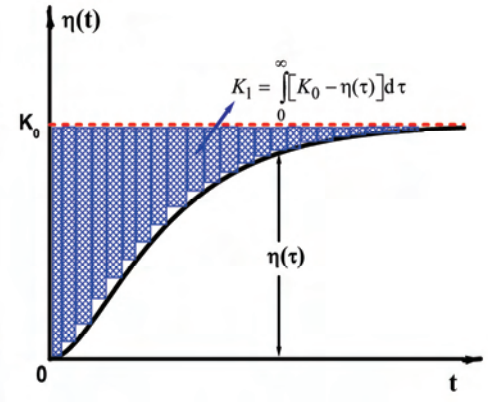


Fig. 1. Evaluation of the unit step response

$$\hat{H}_2(s)=\frac{1}{s}\left[K_1-\hat{H}_1(s)\right]$$
$$=\frac{1}{s}\cdot\frac{\left(K_1-K_0\hat{a}_1+\hat{b}_1\right)+\left(K_1\hat{a}_1-K_0\hat{a}_2+\hat{b}_2\right)s+\cdots}{1+\hat{a}_1 s+\cdots+\hat{a}_n s^n}, \tag{13}$$

$$\eta_2(t)=\int_0^t\left[K_1-\eta_1(\tau)\right]\mathrm{d}\tau\,, \tag{14}$$

and

$$\lim_{t\to\infty}\eta_2(t)=K_2=\lim_{s\to 0}\hat{H}_2(s)=K_1\hat{a}_1-K_0\hat{a}_2+\hat{b}_2\quad. \tag{15}$$

Continuing with the integration, a system of linear equations is obtained with a general formula as

$$(-1)^r\hat{b}_r+K_{r-1}\hat{a}_1-K_{r-2}\hat{a}_2+\cdots+(-1)^{r-1}K_0\hat{a}_r=K_r$$
$$r=0,1,\ldots,n\quad. \tag{16}$$

Table I contains a summary of the identification procedure in the case when all b_j parameters except b_0 in (3) are equal to zero.

TABLE I
ALGORITHM FOR COMPUTING THE PARAMETERS OF THE TRANSFER FUNCTION IN (3)

$b_i=0\,,\ i=1,2,\ldots,m\,;\ b_0\neq 0$
$\hat{b}_0=K_0$
$\hat{a}_1=\frac{K_1}{K_0}$
$\hat{a}_2=\frac{K_1}{K_0}\hat{a}_1-\frac{K_2}{K_0}$
$\hat{a}_3=\frac{K_1}{K_0}\hat{a}_2-\frac{K_2}{K_0}\hat{a}_1+\frac{K_3}{K_0}$
$\vdots$
$\hat{a}_n=\frac{K_1}{K_0}\hat{a}_{n-1}-\frac{K_2}{K_0}\hat{a}_{n-2}+\cdots+(-1)^{n+1}\frac{K_n}{K_0}$

Fig. 2 presents a simple block-diagram for performing all calculations in the described identification procedure. The measured data, obtained experimmentally from a real-time set-up, can be used inthe MATLAB®Simulink environment to identify unknown system parameters.It is obvious that the considered identification approach is very simple and can be realized with the minimalcomputational effort.

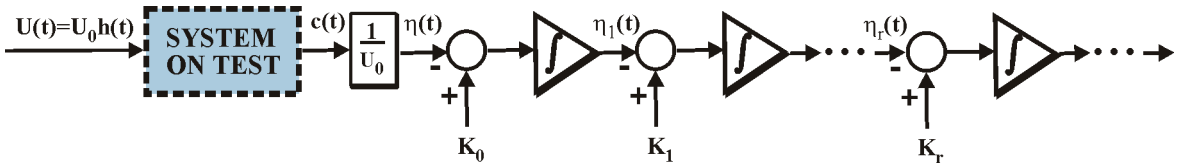


Fig.2. The schematic diagram of the considered identification procedure

III. Simulation Examples

In this paper, the described method will be illustrated by two examples related to the fourth-order linear objects without finite zeros.

A. Example 1.

Consider the following fourth-order linear system given by

$$H(s)=\frac{0.2}{0.2s^4+0.8s^3+2s^2+1.2s+1}. \qquad (17)$$

In simulation, the input is taken as a step signal with the amplitude equal to 10. The important point is that the duration of the simulation should be sufficient to ensure that the input signal be able to excite the slowest system mode. The estimated parameters for a fourth-order model are shown in Fig. 3. It can be seen that after 25 seconds the estimated parameters converge to the true values.

B. Example 2.

Fig. 4 visualizes one of the variety of configurations to be obtaind with the ECP Model 210 Rectilinear Plant by using springs of varying stiffness [5]. A drive motor provides actuation to the system via the first mass, and position measurements $x_i(t)$, $i=1,2$ are taken by quadrature encoders.

The equations for the considered mass-spring system may be found using Newton's laws to write force balance equations in matrix notation as

$$\mathbf{m}\ddot{\mathbf{x}}(t)+\mathbf{c}\dot{\mathbf{x}}(t)+\mathbf{k}\mathbf{x}(t)=\mathbf{F}(t), \qquad (18)$$

where

$$\mathbf{x}(t)=\begin{bmatrix}x_1(t)\\x_2(t)\end{bmatrix},\ \mathbf{m}=\begin{bmatrix}m_1&0\\0&m_2\end{bmatrix},\ \mathbf{c}=\begin{bmatrix}0\\c_2\end{bmatrix},$$
$$\mathbf{k}=\begin{bmatrix}k_1+k_2&-k_2\\-k_2&k_2+k_3\end{bmatrix},\ \text{and}\ \ \mathbf{F}(t)=\begin{bmatrix}f(t)\\0\end{bmatrix}. \qquad (19)$$

The parameters of this set-up can be found in the literature[5], as follows: $m_1=1.7\,\text{kg}$, $m_2=1.2\,\text{kg}$, $k_1=k_2=800\,\text{N/m}$, $k_3=450\,\text{N/m}$, $c_2=9\,\text{Ns/m}$. The above motion equation results in the transfer function

$$H(s)=\frac{X_2(s)}{F(s)}=\frac{392.2}{s^4+7.5s^3+1983s^2+7059s+666700}, \qquad (20)$$

which can be equivalently rewritten in such a way as in (1) with:

$$b_0=5.8827\cdot10^{-4},\ a_1=0.0106,\ a_2=0.003,$$
$$a_3=1.1249\cdot10^{-5},\ a_4=1.4999\cdot10^{-6}. \qquad (21)$$

It should be noted that all a_i - parameters have low values, and moreover are of different order of magnitude, which is the known characteristic of the electro-mechanical systems.

For the purpose of the fourth-order model parameter estimation, the algorithm described in the previous section was applied. Table II presents the estimated values of the parameters obtained at some different conditions of experimentation. The results illustrate some of the fundamental problems of system identificationrelated to the experiment duration and the accuracy of data presentation. Thus, the electro-mechanical plant given in Fig. 4 can be

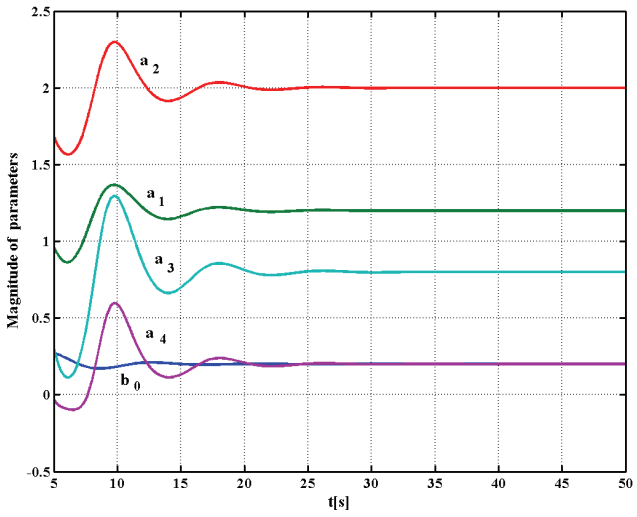


Fig.3. Estimated model parameters in Example 1

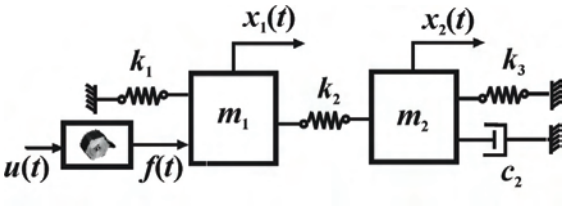


Fig.4. The scheme of the electro-mechanical plant

TABLE II
ESTIMATED PARAMETERS IN EXAMPLE 2

Estimated parameters	Working formats in MATLAB® Simulink environment		
	format long		format short
	Measured time $t = 20\text{s}$	Measured time $t = 10\text{s}$ $t = 20\text{s}$	
$\hat{b}_0$	0.00058827058647	0.00058827056152	0.0006
$\hat{a}_1$	0.01058797060125	0.01058754730445	0.0106
$\hat{a}_2$	0.00297435128462	0.00297645908868	0.0032
$\hat{a}_3$	0.00001124942294	0.00000424347456	-0.0013
$\hat{a}_4$	0.00000149999767	0.00001896499007	0.0065
Efficiency	very good	poor	very poor

adequately represented by the model obtained after the identification procedure lasting 20 seconds and retaining the larger number of decimal places corresponding to the MATLAB® data presentation in long format.

IV. CONCLUSION

System identification is a well-established field. However, the search for the simple procedures of identification is still a special scientific challenge. This paper presents an identification algorithm implemented in the MATLAB®-Simulink environment based on the well-known final theorem of the Laplace transform. The properties of the described identification method are illustrated by the simulation results.

At the present stage, some conclusions can be drawn from the above study. The method is more difficult to implement if the model to be identified is given in the form of the transfer function with the finite zeros. The quality of the estimation has not be analyzed in the case of the noise corrupted system step responses.

ACKNOWLEDGEMENT

This paper is supported in part by project Grant III44004 (2011-2014) financed by Ministry of Education and Science, Republic of Serbia.

REFERENCES

[1] R. Pintelon, and J. Schoukens, *SystemIdentification: A Frequency Response Approach*, New York: IEEE Press, 2001.
[2] L. Ljung, *System Identification: Theory for the User*,Englewood Cliffs,NJ: Prentice-Hall, 1999.
[3] G.P. Rao, and H. Unbehauen, "Identification of continuous-time systems", Control Theory and Applications, IEE Proc, vol. 153, no. 2, pp. 185-220, 2006.
[4] H. Rake, "Step response and frequency response methods", Automatica, vol. 16, no. 5, pp. 519–526, 1980.
[5] T. R. Parks, "ECP Instructor's Manual for Model 210/210a Rectilinear Control System", 1999. Available: http://maeweb.ucsd.edu/~ugcl/download/manuals/linear.pdf (accessed on June 3rd 2011).

Session PO5:

PO5 – RENEWABLE ENERGIES

Wind Generators

HristoI. Toshev[1], Chavdar D. Korsemov[2]

Abstract – **The paper presented discusses the wind generators, which implement the real possibilities for rational wind energy use. The theoretic formulation of wind energy transformation into rotational mechanic energy and the features of the two types – wind turbine generators (WTG) and wind vane generators (WVG) are considered. Comparison is done between 5 different types of WTG and the perspectives of WVG in the production of electrical energy at the lowest price are presented.**

Keywords – **wind turbine generators, wind vane generators, electro-energy**

I. INTRODUCTION

The possibilities for rational use of the wind energy are indeed remarkable. About 1÷2% of the sunny energy, falling on earth is converted into wind. For comparison, the energy, absorbed by plants in photosynthesis, and its conversion into bio mass is only 0.02÷0.03%, or about hundred times smaller. Modern civilization, often regarded as lavish, consumes totally about 0.005÷0.006% of this energy[1].According to the estimates in [2,3,4],the actual energy consumption can be satisfied byobtaining wind energy at a height of80mover 20% of the regions with average annual wind speed 6.9m/sec.

The world stores of easy obtained wind energy are estimated at1 500 GW, with annual production of $3 \cdot 10^{12}$ kWh, that makes 500 kWh per every inhabitant of the world population of 6 milliards[5].

The main technology of wind energy generation is based on the wind wheel, known for centuries. Its improvement in the last decades puts this technology close to its limit capacities -– energy deriving up to 50 % of wind ground level at height up to 180m. The energy generation is limited by Betz law, and the diameter of the wheel (propeller) – by the mechanic strength of the known materials.

There is also another technology [4], realized by a fan (kite). It enables energy obtaining at height of 200÷800m, where the wind speed is higher, less dependent on the covered surface, the fan itself has a simple construction, small weight, but requires more complex control.

II. THEORETIC FORMULATION

2.1. Conversion of the wind energy into rotational mechanic energy

Kinetic energy, carried by the wind through a section *Q*for1 sec, is:

$$E_R = \frac{1}{2} m v^2 = \frac{1}{2} P Q v^3,$$

or expressed by the surface power density (specific power)

$$G = \frac{1}{2} \rho v^3 [W / m^2], \qquad (1)$$

where $\rho[kg / m^3]$ is air density in the region considered, and $V[m / s]$ – the wind speed. The density ρ is a function of its temperature, the atmosphere pressure and is given by the relation [6]

$$\rho = \frac{\rho_0}{R_d (T + \beta.z)} \left(1 + \frac{\beta.z}{T}\right)^{1 - \frac{9}{R_d \beta}}$$

where $\rho_0 = 1{,}225[kg / m^3]$ is the density of the dry air at sea level pressure and temperature 25^0 C, R_d - a gas constant, T – local temperature in ^0K , z – height above the sea level, $g = 9{,}81[m / s^2]$ - earth acceleration, $\beta = 6{,}5^0 C / km$ - vertical temperature gradient for standard atmosphere.

The power density *G*depends on the third power of the wind speed. The speed itself depends on the height above the earth, up to 450 mit increases, after that – decreases. In order to use the wind energy, it is of interest to know the interval 40 - 150 m, in which the wind turbine generators (WTG) are located. For this area the relation can be presented by an empirical equation [6,7]

$$v_h = v_{h_0} \left(\frac{h}{h_0}\right)^{\alpha},$$

where v_h is the speed at height*h*, v_{h_0} is the known speed at height h_0, α is an experimentally determined coefficient. The value of α may vary in a wide range, depending on the height, the day time, the season, the temperature, the site features. Table 1 gives typical values of α , depending on the terrain surface (covering surface) [7].

Regardless of the fact that the value of α is not big, it influences significantly the specific power, due to its strong dependence on the wind speed. This dependence can be used for selection of the optimal height of WTG tower. It enables the obtaining of a more realistic preliminary estimate of the wind energy resources of the country by the available data about the speed, collected by meteorology.

They refer to terrain height above 10 m. For a pile, high 80mand for the smallest value of α = 0,1, the specific power increases $\left[v_{10}\left(\frac{80}{10}\right)^{0.1}\right]^3 = 1.87$ times.

[1] Hristo I. Toshev and [2] Chavdar D. Korsemov are with the Institute of Informationand Communication Technologies, BulgarianAcademy of Sciences, Acad. G. Bonchev str., bl. 2, 1113 Sofia, Bulgaria, E-mail: toshev@iinf.bas.bg, chkorsemov@iinf.bas.bg

TABLE 1

Terrain type	α
Sea, lake, smooth earth surface	0.1
High grass on an even terrain	0.15
High crops, bushes	0.20
Forest site with many trees	0.25
Small settlements with trees, bushes	0.30
Towns with high buildings	0.40

When a fluid flow meets a barrier, a dynamic force appears on it, equal to the speed of the fluid and the barrier. This force is used in the wind turbine to extract wind energy. Meeting the turbine fans, the wind forms a force, tangential to the fans and it drives them in a rotating movement, which directly or with the help of a speed box (for increase of the rotating speed), is transferred to the electric generator. In the current constructions the turbine, the speed box and the generator are made as a monolithic block and together with the propeller they form the wind-turbine generator (WTG). The extracting of the complete kinetic wind energy is not possible. The limit theoretic generation at completely smooth fans surfaces and non-viscosity fluid is given by Betz bound and it is $C_p = \frac{16}{27} = 59{,}7\%$ [8]. It is assumed that the value of 0.8 is a technically achievable approximation to this bound, hence the real limit extraction, often quoted, is C_p=0,8.59,3=47,3%. With the development of technologies, materials, and knowledge of the aerodynamic processes, the real extraction is increased and it comes close to the limit one. Only in the last years it has increased from 22% up to 38% [8].

III. GENERATORS

3.1. Wind turbine generators (WTG)

The main type of generators, used until 1990 for power up to 1.5 MW, are the asynchronous ones, with a cage rotor, directly connected to the power line, from which they obtain the excitation - a Danish concept for WTG. They differ by their simplicity, reliability and a century developed technology of mass production. They operate at constant revolutions which does not allow maximal extraction of the wind energy. It is obtained for one single wind speed. The connection to the network is done by controlled valves restricting the big current overloading. After the transition process of switching $(0{,}2 \div 2s)$,the valves are shunted by contactors. In order to compensate for the magnetizing (excitation) current, some adjusted capacitor batteries are joined. The excitation current depends on the degree of loading and for its constant compensation, the battery must be adjustable.

A serious disadvantage of this type of generators is its firm characteristic resistance moment/revolutions. The typical for the wind speed changeability of frequent stormy nature causes stress loading on the propeller shaft and especially on the speed box.

The introduction of a winded rotor is an improved modification of the asynchronous motor. The slippage is controlled by the degree of loading on the rotor windings, and thus on the revolutions, and the characteristic moment/revolutions becomes soft. The wind blow and the impulse loading of the generator are smoothed. The operation revolutions can be co-coordinated with the wind speed for maximal energy extraction. They are available in brush and non-brush variant. The heat mode of the rotor is lightened in the brush variant. Usually the revolutions differ from the synchronous ones by 1-10% [9].

The most widely used type of a generator in WTG is the Doubly Fed Induction asynchronous Generator (Doubly-FedInductionGenerator – DFIG), shown on Fig.1. Regulating the energy, obtained by the rotor windings, the speed of propeller's rotation is controlled.

The double inversion enables efficient control of the slippage power (which reaches up to 30% of the nominal power), as well as its transfer together with this of the stator to the network. The new valves of IGBT type allow inversion with a high yield.

In order to design generators with acceptable dimensions, and hence – weights, their revolutions are most often chosen within the limits 1000-2000 rev./min and the poles number - 4 or 6. To achieve these revolutions, three-degree speed boxes are used. It is considered, that the losses in every transmission unit, if not higher than 6 is about 1% of the transmitted power in case of good support.

Synchronous and asynchronous generators with variable rotational speed and with double conversion of the whole power – both the main one and that of slippage, are used. The entire disconnection between the propeller rotational speed and the network frequency enable the operation in four quadrants – in a mode of a generator, of a motor, of a consumer, and a generator or reactive power (capacitive, induction). This is an important feature, the use of which can improve the reliability of the network, to which it is attached. For small active power due to the wind low speed, such a generator can provide reactive power to the network up to its nominal apparent power.

The present-day WTG are constructed for nominal wind speed within the range 12÷16m/sec.It is selected according to the function of distribution for the site, where WTG is installed. For speeds above the nominal ones, the nominal output power is obtained with the help of the angle of vanes attack.

For the widely spread two vanes propellers, the maximal extraction is achieved at relation peripheral speed/wind speed λ=7÷8. At nominal wind speed of 12 m/sec, the peripheral one for the propeller is 320km/h. The peripheral speed influences to the greatest extent the acoustic interference and the amortization speed of the propeller.

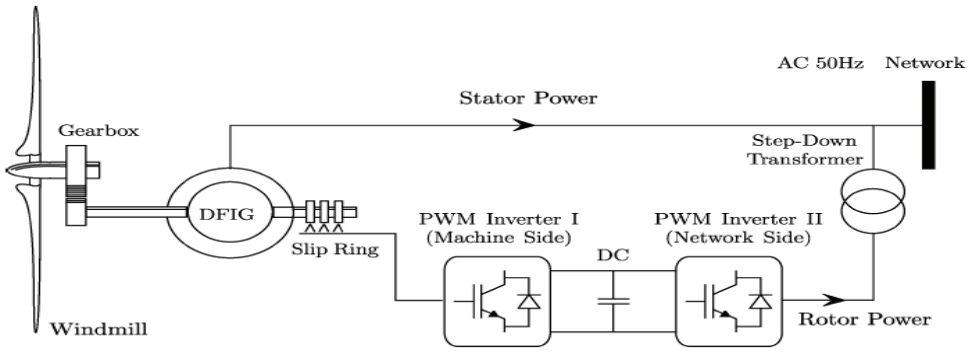


Fig. 1

At nominal speed of the wind 12m/sec, extractionC_p=38%, dry air, at sea level, temperatures of 25°C, the useful power density is 381W/m^2.

For such density, power on the propeller's shaft of 3MW, the propeller diameter must be 100.2m and it will extract energy from 7885m^2. The averaged power density is 87,3W/m^{2}atvery good effectiveness.

When building wind farms, in order to support the aerodynamic interference in acceptable bounds, the distances between the separate WTG must not be smaller than 4-5 multiple of the propeller diameter. With this constraint, the terrain density of a farm, built of WTG with nominal power of 2 MW,at nominal wind speed of 12m/sec (diameter of 90m),is 9MW/km^2 [4].

Both densities –the wind and the terrain are very low. Though the primary energy is free of charge, its extraction requires large investments and big areas. This concentrates the design of wind farms on deserted land only.

The achievements in generators construction, computer technologies and power electronics have enabled the design of new generations of generators with complete conversion and variable speed that allows maximal extraction of the energy for any wind speed. These are mainly the synchronous generators with constant magnets made of rare metals and the asynchronous doubly fed generators. The generators with complete conversion enable operation in a three-quadrant mode – active power, reactive induction and reactive capacitive in arbitrary portions. Their main advantage is that even at weak wind they can generate reactive power equal to the nominal apparent one.

Interesting comparison is presented in[10] between 5 different types of WTG with nominal power of 3MW:

— doubly fed asynchronous generators with complete conversion and a three-degree transmission box (transmission relation 1:80) – DFIG;

—directly connected to the propeller (without a speed box) 80-poles synchronous generators with electric excitation and complete conversion – DDSG;

—directly connected to the propeller 160-poles synchronous generators with constant magnets and complete conversion – DDPMG;

—synchronous 112-poles generators with one-degree transmission box with constant magnets and complete conversion – PMG1G;

—doubly fed asynchronous generators 80-poles generators with one-degree transmission boxes with complete conversion – DFIG1G.

The values of the main parameters are given – geometric dimensions, weights, losses in the separate units, efficiency coefficient,energy produced annually, including the construction and mounting activities for towers building. Despite of the essential differences, the margin in the significant indicators is not big – 10% in the cost of the energy produced and 2.2% - in the annually produced energy.

3.2. Wind vane generators (WVG)

The vane technology is not developed to the extent of its real application, but thanks to its potential possibilities, it must be accounted in the development of wind energetic strategy [4,11].

The extraction of wind energy from the wind vane generator (WVG) is accomplished with the same aerodynamic force, as in WTG. The barrier in them is made as a vane, (kite)of folio, similar to a ship sail, as shown in Fig.2. With the help of two cables the vane is connected to the ground equipment, which transforms the vane movement into electric energy. The cables are winded/unwinded on separate drums, each one connected to an electric machine that can work in a generator and in a motor mode. With the help of the cables the information about the vane and the wind speed is used to regulate the attack angle, so that cyclic movement, normal for the wind speed, is obtained.

The most significant advantages of this technology, are:

— the energy extraction is realized at big height with the help of simple and light equipments;

— the heavy equipment is found on the earth;

— possibility to work in a wide range of the wind speed of 2-50m/sec;

—higher usability in comparison with WTG;

— higher terrain density of the energy extracted.

Many simulation investigations have been carried out. The results from them are illustrated by an example WVG with the following basic parameters:

— vane ofpolyhedron with usable surface of500m^3, weight of300kg;

—cables of a composite material, with length of 900m, diameter - 30mm and linear weight of 68.6gr/m;

—area, swept by the vane axaxΔr=300mX300mX50m.

—losses for vane control - 30% and efficiency coefficient of the ground equipment - 70%.

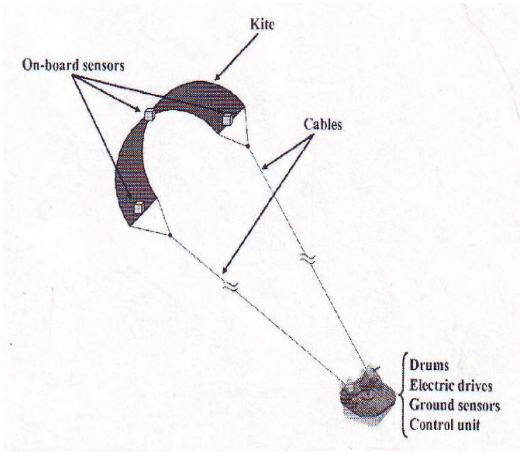


Fig. 2

The power, generated by such a WVG, for wind speed of 9m/sec, is 2MW, and for speed of 12m/sec, the lowest nominal speed for a present-day WVG is already 5kW.

Taking into consideration the aerodynamic interference, the degree of vane regulation, protection against cables intertwining with neighbouring WWG, at nominal speed of the wind 12m/sec, the terrain density of the nominal generated power is 80MW/km^2. This density is one order higher than the density of WTG with nominal 2MW.

IV. Conclusions

The storing of wind electric energy is solved nowadays in a new way. There are constructed powerful electric networks with built in systems for voltage and frequency support, which enables their assuming as infinitely powerful. The connection of WTG to such networks does not practically alter their mode, independently on the changeability of the power injected. The introduction and use of considerable power does not lead to surpassing of the feasible limit values of the network basic parameters. At considerably small share of the wind electric energy, for example, up to 10 %, the built-in system for network stability support may remain without any changes. At larger participation, with the connection of WTG farms in the operative control of the network and the use of the prognostic values for wind speed, the relative value may reach big values, 30% or more. A network is presented in [12], in which the wind energy may cover the whole consumption and may be injected in other networks.

Considerable increase of the energy, generated by the wind, is noted since 1980.This is connected to a great extent with the improvement of the shape of turbine propellers, which leads to raising the share of the energy extracted by wind, improvement of the electric generators, especially those with variable speed, of the electronic converters, that realize partial or complete conversion and improving the technology of connection to the power line.

In[4] a probable cost of the energy, obtained by vane technology in 2030, is given. In Table 2 this cost is compared to IEA prognosis for the same time, by different resources, taken from [13].

TABLE2
ENERGY COST IN 2030

Resource	Maximal c$/kWh	Minimal c$/kWh	Average c$/kWh
Coal	5	2,5	3,4
Gas	6	3,7	4,7
Sun	50	18	32,5
Nuclear	3,1	2,1	2,9
WindWTG	9,5	3,5	5,7
Wind WVG	4,8	1	2

The promises of WVG are very attractive, which requires the technology putting into operation. The main problem is the necessity of a very complex control, which covers the most actively developed areas of human knowledge – aerodynamics, control, microelectronics.

Acknowledgment

This work is a part of IICT-BAS research project “Modeling, Optimization and Multiple-Criteria Decision Making”

References

[1] Danish Wind Energy Association – (http://www.windpower.org/en/core.htm) /fr/tour/wres/index
[2] Archer C. L., M.Z. Jacobson Evaluation of global wind power, *J. Geophis. Res.* Vol. 110, p.D 12110, 2005.
[3] http://www.electron-economy.org/article-27628373.htmlJ.
[4] L.Fagiano, M. Milanese, D. Piga, High-Altitude Wind Power Generation, *IEEE Trans. Energy Conversion*, vol. 25, No.1, March 2010, pp. 168-180, 2010.
[5] Wind Force 12, 2005 (http://ewea.org/fileadmin/ewea_documents/documents/publications/WF12/wf12-2005.pdf), 2005.
[6] M.Fripp, R.Wiser. Effects of temporal Wind Patterns on the Value of Wind-Generated Electricity in California and Northwest. – *IEEE Trans. Power Systems*, vol.23, No 2, May 2008, pp.477-485, 2008.
[7] Tai-Her Yeh, Li Wang. A study on Generator Capacity for Wind Turbines Under Various Tower Heights and Rated Wind Speeds Using Weibull Distribution. -*IEEE TransEnergy Conversion*, vol.23, No 2, June 2008, pp.592-602, 2008.
[8] R.Thresher, M.Robinson, P.Veers. To Capture the Wind, *IEEE Power & Energy Mag.*, No 6, Nov./Dec. 2007, pp.34-46, 2007.
[9] L.Gertmar, L.Liljestrand, H.Lendenmann. Wind Energy Powers-That-Be Successor Generation in Globalization, *IEEE Trans. Energy Convers.*, vol.22, No 1, March 2007, pp.13-28, 2007.
[10] H.Polinder, F.Pijil, G.Vilder, P.Tavner. Comparison of Direct-Drive and Geared Generator Concepts for Wind Turbines. -*IEEE Trans. Energy Convers.*, vol.21, No 3, Sept.2006, pp.725-733, 2006.
[11] M. L.Loyd, , Crosswind kite power, *J. Energy*, vol. 4, No.3, pp. 106-111,1980.
[12] L.Söder, L.Hofmann, A.Orths, H.Holttinen, Y.Wan. A Tuohy Experience From Wind Integration in Some High Penetration Areas. -*IEEE Trans. E.C.*, vol.22, No 1, March 2007, pp.4-12, 2007.
[13] IEA, 2008 Projected Cost of Generating Energy-2005 update http://www.ica.org/Textbase/publications/free_new_Desk.asp?PUBS_ID=1472, 2008.

Wind Farms and Their Connection to a Power Line

HristoI. Toshev[1], Chavdar D. Korsemov[2]

Abstract – **The paper shows, that with the rapid use of wind energy, significant power resources can be obtained by the connection of a set of wind generators in the so called wind farms. The paper discusses the technical features of these farms and the conditions for their operation. Their connection to the power line is of considerable importance for their correct functioning, which requires the solution of different specific problems – the random natute of wind, the selection of the farm site, the compensation of the short-time changes of wind power, the prognosis of wind intensity etc.**

Keywords – **wind generators, wind farms, electroenergy**

I. INTRODUCTION

The wind is a very convenient energy source. It is found everywhere and free of charge. However, its use is connected with large investments due to its small energy density. The achievements in the area of material knowledge, aerodynamics, processing industry technologies and the increased requirements of ecology made the cost of wind energy generationreasonable. The last two decades show considerable achievements in wind energetics. This refers mainly to the improved aerodynamics of the propellers, thus increasing the extraction $C_p$up to 37% (at a limit value of 47.3%) and especiallytheir productionby new strong and light materials, enabling constructions with large dimensions. The typical propeller for a wind generator (WTG) with power of 1.5 MW has a diameter of77 mand it extracts the energy from a wind flow with a section of 4650 m^2. Thus, power of 20 MW is obtained from a surface of 1 km^2 [1].

The connection of a separate generator to the power line is an isolated case, connected with big expenses for its joining. That is why in practice the connection of a set ofWTGis realized in the so called wind farms.

The paper presented discusses some basic features of the WTGfarms.

II. WTG FARMS AND THEIR EXPLOITATION

When combining a set of WTG, located in farms, considerable energy power is obtained. This presumes the construction of a common infrastructure, building fund, service, devices intended for connection to the network, a data acquisition system, continuous monitoring control, a system for remote measurement and control done by the operating control of the network, which includes the farm, a system for distribution of the active and reactive power among the separate WTG, lightnings protection, overvoltage protection in the general network, prognosis of the wind speed.

The singular power of a WTG, compared to the conventional thermal or nuclear generators is small—within the limits of 1-5 MW. No considerable increase of this power can be expected. It is connected with increase of the surface, covered by the propeller and increases linearly with the second degree of its diameter, that leads to many constraints of another nature – weight, dimensions, possibility for transporting, mounting, available strong materials, lightning resistance, etc. For example, a WTG with power of 2.2 MW,has a propeller with a diameter of112 m, weight of 20 tonsand it obtains the energy from a section of 9800 m^2. [2].

The connection of a separate generator to the power line is an isolated case, connected with big expenses for the joining itself. Its influence on the network can be regarded as an equivalent consumer.

For the main network, to which the farm is attached, it is regarded as a singular equivalent generator. Its dynamics is very different from conventional synchronous generators, supplying the network. This is particularly well expressed for a WTG with complete conversion. For big farms, above 150 MW, the WTG may form several separate groups, each one with a separate collecting feeder. The farm with power of 640kW, presented in [3]is comprised by four separate groups, included intwo different networks.

The farms are situated on large sites and the joining cables have big length, connecting dozens of generators. Every connection and disconnection of a generator is a pulse disturbance introduced in a homogeneous line with many reflection points. The set of the falling and reflected waves causes overvoltages that load both the cable and the devices connected to it. The sources of overvoltages in the collecting cable are the generators switchings, and in the high voltage air lines, to which the farm is connected – the lightnings occurrence.

The matching transformer farm/network is subjected to overvoltages in the two windings. The overvoltage front of raising is multiple, at least 10 times steeper on the primary winding, due to the low wave resistance of the cable (below 40 Ω),in comparison with the one of the secondary winding, connected to the air line, for which the wave resistance is 300-400 Ω.

The correct exploitation requires qualified and constant support. This is particularly important for the speed boxes. It is recommended the wind energy supplies to be located near to the energy users. Anyway, this is satisfied only in some rare cases, having in mind that these power sources are situated outside inhabited areas. The connection is usually to a network of middle or high voltage. The wind energy generation is organized in farms, with a total power of dozens orhundreds MW, included in the industrial network and its operative control.

[1] Hristo I. Toshev and [2] Chavdar D. Korsemov are with the Institute of Informationand Communication Technologies, BulgarianAcademy of Sciences, Acad. G. Bonchev str., bl. 2, 1113 Sofia, Bulgaria, E-mail: toshev@iinf.bas.bg, chkorsemov@iinf.bas.bg

In [4] some data are presented for two WTGfarms, the first one being onshore, built inNässudden – Sweden, and the other one – offshore, inKentishFlats, Northern sea - England, property of Elsam (Denmark).

The farm inNässunddenincludes 100 WTG of different power, different owners and made by different manufacturers, a significant part of them belonging to VestasandOlsvennez. Thirty of the WTG are property of Vattenfall, and the data, presented below, are relevant for them.

The WTG inKentishFlatsare of one type, 30 in number, each one with nominal power of 3 MW. The whole farm is put into operation in September, 2005.

The lifecycle of the generators in the two farms is 20 years. Theavailability is about 97.5%, determined as a relation between the expected time for operation readiness and the calendar time

$$D = \frac{T_{M_P}}{T_{M_P} + T_P},$$

where T_{M_P} is the average interval between repairs, and T_P – the time for planned or extra repair (service) activities.

The investments according to the data attached are 3M€/BΓ forNässundenand 10M€/BΓ -forKentishFlats. These sums cover all the expenses for the equipment purchase, mounting, the construction of a farmnetwork, and the common connection to the industrial network, the guaranteed support by the suppliers, which is two years for Nässundenand 5 years forKentishFlats.

In order to achieve the high level of operational readiness, constant support is provided, that is of two types–prophylactic and in failure cases. The prophylactic can also be planned or extra, caused by the alterations of some parameters registered during WTG monitoring. The planned support includes screws tightening, oil and filters replacement, check of the security system.

Vattenfallprovides the support in accordance with annual contracts with companies, manufacturers of WTG– Vestas, SiemensandEnercum.

ForKentishFlats,Vestasensures complete support for 5 years, its cost being included in the investment expenses.

The two farms rely mainly on the planned support. ForVattenfallthe planned service is realized twice a year, basic and auxiliary. The basic support is accomplished by two specialists for 7 hours, and the auxiliary one – also by two persons, but for 4 hours, and the hour payment for every specialist is54 €.

In Elsamthe planned service is realized every 3-6 monthsfor the smaller and older WTG, and every 6-12 months - for the big WTG. There is a team of 10-15 workers, which supports the old WTG. For more radical repairs, some external specialized services are employed. InKentishFlatsthe planned support is realized by two workers for two days for the WTG, with daily payment for each worker of 750 €. When a more thorough support is needed, the payment is raised to 850 Euro. These expenses include all the expenses for the working staff, like training, equipment for safe work, pensions, transport, offices holding.

The replacement of some large dimension units in WTG takes a considerable part in farms life cycle. The prices of the units, without the disassembly and assembly works for the sites discussed, are as follows:speed box –300 к€, generator - 150 к€, transformer -100 к€, vanes -200 к€

For the farm inNässudden, the speed box is replaced twice, the generators - once and for 1/10 of the WTG the vanes and transformers are replaced. Some expenses for demounting and recycling are foreseen for the last year.

For KentishFlatsfarmthe expenses for large dimension units include: 11 speed boxes, 20 generators, 3 transformers and 3 vanes. Besides this, some additional checks and repairs are intended in the years 4, 6, 9, 13 and 14. One day is needed for the replacement of a large dimension unit.

III. Connection to the Power Line

The connection of WTGfarms toa power line is related with the solution of many problems. The generated power has random nature with a large range of alteration –from zero up to the nominal value and limited predictability. The null power is often with a probability above 70%. The power control is possible only in direction of random maximum decrease. With this feature the network gets one more stochastic process together with the loads. The two processes are not correlated, compensation is possible in certain periods, in others–depositing, but in general, the dispersion increases. Due to the high probability for farm inactiveness, the nominal power of every farm must be reserved by conventional sources power. This reservation is especially heavy at relatively big share of the wind energy in the network. The reservation could be decreased with the help of the construction of several farms, located in sites with non-correlated wind energy. This requires good knowledge of the wind picture of the network region and appropriate planning of wind energetic development. Meteorological services data can be used for initial evaluation of a more continuous period of one up to three years.

The selection of farms sites is a complex optimization task. Besides the wind power, “statistic” independence, the location of the consumers, the distance to the electric network, the allocation of the conventional generators, their dynamics for compensating the changes of the wind energy must also be taken into consideration. The connection to the farm end (the generators remote points) may lead to decrease in network losses.

The compensating of short time alterations of the wind power is a separate task. It is done by rotating reserve powers and connected with additional dynamic losses. According to UCTE requirements (15),each change of the power consumed (for the wind power –positive or negative) must be compensated in an interval up to 15 min. The WTG with a cage rotor and especially those with complete conversion, have high dynamics and can participate in the compensation process, caused by load changes.

The predicting of the generated power has great significance in restricting the reserve and particularly the reserve rotating powers for every farm. It is usually assumed that the wind farms introduce strong disturbances in the energy network operation. The use of short-time (15 min) and daily (up to 38 h) prediction enables considerable decrease of

the necessary rotating reserve powers. There are some well known methods, which, using the prognostic information, allow the short-time planning of the power, supplied by the conventional and wind generators for optimizing the value of the generated energy and the network reliability [2,5,6].

The stochastic feature, introduced in the network by the wind farmsin case of relatively large participation in the general power, is a potential source for instability. The operative staff must often distribute the generators loading, as well as the rotating reserve powers, including the wind, for minimizing the cost of the energy produced, accounting the constraints for:

- consumed electric energy,
- consumed thermal power by thermal and nuclear stations,
- possibilities of the generators for impulse (short-time) alteration of the generated power,
- actual prognosis for possible wind energy,
- special operation mode at small consumption–night hours, off days, holidays,
- water consumption in hydro electrical stations,
- electrical energy accumulation at PAWEC.

The rules for electric network operation, as well as the including of new energy sources is regulated by gridcodes, specific for different countries and regions, but not differing much. Due to the principal difference of the wind energy generation in comparison with the conventional sources, there are introduced specific requirements in the countries with advanced wind energetics.

In USA several organizations encode the wind energy share in the network - FERG, AWEA, WECC, NERC [29]. The including of wind powers above 20 MW is treated by Order No 2003 from July 2003 and Appendix G, Order No 661 from 5 July 2005.

Some more specific requirements of AWEA grid code for the wind farmsare:

- supervisor control, data acquisition, equipment for exchange of telemetric data and remote control, with the main purpose -exchange of forecast, setting of the power generated by the farm, when the network operative control implies restriction on the maximal power;
- exchange of information about the actual status of farm equipments. Used to renovate the network model in optimizing the operation mode;
- it is recommended that the WP would work with a power factor of +/- 0,95;

The WP participates in network voltage maintenance, as the farm remains connected, at time-voltage dependence in the connection point above the curve, shown on Fig. 1.

This is connected with the possibility to cover short-time deep voltage drops [7]. The WTGfarm (or a separate WTG) must remain connected to the network in case of voltage drop up to 15 % from the nominal one for 625 ms (75 network semi periods at 60 Hz) and constantly - in case of decreased voltage up to 90 %. It must also support other degrees of falling with a duration according to the figure. The relation accepted is important for network restoring after short-time overloading, including short circuits. The accepted norm is a strict requirement towards the generators under stress loading.

The Swedish code of operation at dropped voltage permits also work after short circuit at the WTGfarm output, but only for 240 ms and lower feasible durations of operation with decreased voltage, as shown on Fig. 2 [8]

The energetic system must be stable with respect to the damages, including the short circuits, which are a natural and quite frequent event in its sections. A damage in one section must not lead to cascade falling off of other sections. This refers both to the whole system, and also to the WTGfarm, composed of several independent units. The short circuits are connected with the consumption of big reactive power, the supply of which is not a problem for the WTG.

The influence of the changeability of the wind energy on the whole network decreases considerably for wind farms, which are not closely located. Measurements of the power network in Western Denmark and North Germany show that even at mighty windy fronts the gradient of the wind power does not exceed 10 MW/min (installed power of 2.4 GW) and 16 MW/min (installed power of7.05 GW), respectively for the first and the second network [9]. The smoothing of the power, supplied by WTGfarms, is investigated in [10]. It is demonstrated, that at considerable share of the wind energy, it is impossible all the WTG to operate in a mode of maximal energy extraction.

The influence of the wind speed dynamics on the non-harmonic spectral components (flicker) is studied in [10]. The optimization in the network is realized by optimization of smoothing. Some approaches are proposed for different number of farms, different degree of correlation of the wind energy in them, different averaging interval.

Modern WTG enable the efficient and sufficiently quick control of the generated power, allowing significant participation in the complete power of the network. In Schleswig-Holstein the wind energy covers 33% of the annual consumption, and the maximal share of its power is 44% [9]. In Western Denmark the complete consumption in the network could be covered for hours by the wind farms.

IV. CONCLUSIONS

The connection of WTG farms to the network causes losses distribution. The total losses diminish, but in separate sections of the complex network they might increase and surpass the feasible limits. This implies network re-sizing.

At small share of the wind energy in the network total power, no particular control is necessary, it can be regarded as commutation of more powerful consumers and the WTG may work in a maximal power mode. At significant participation - 10% and more, the wind farm is included in the operative control of the industrial network.

When the wind is absent (below the lower bound for WTG), the conventional generators must take up the whole power of the network.When the wind is strong, both the network and the farm must enable quick control.

It is probable that at certain time moments of severe winds,

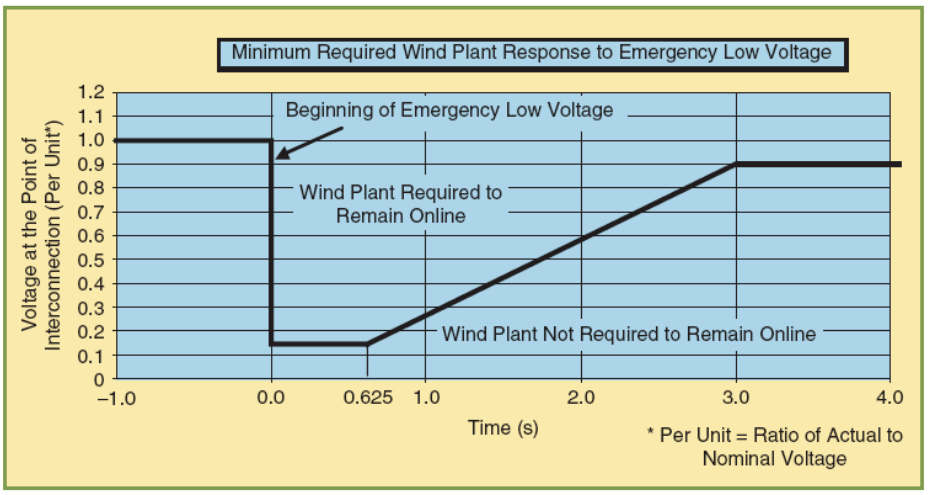


Fig. 1

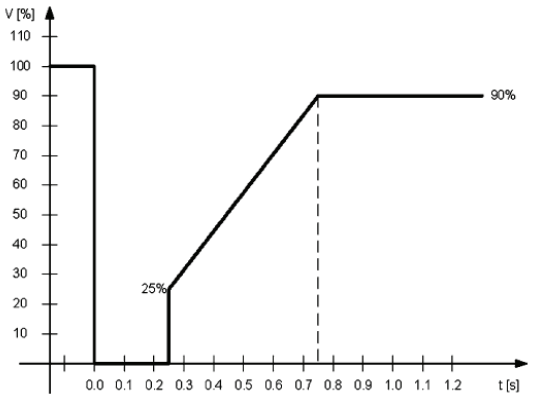


Fig.2

the relative share of the wind farmsis increased multiple –in days off the total consumption decreases twice and more. The other generators must be able to ensure quick reaction to the alterations in the supplied wind energy. This requires a constant reserve of sufficient rotating power in the system. The farms may be given the task to provide the necessary reactive energy in the system for constant voltage support. The WTG with full conversion may provide reactive energy, equal to their nominal power.

The wind farms are a new branch with their advantages and shortcomings. Their implementation must be realized after thorough analysis and evaluation of the perspectives for decreasing their cost and the development of other energy sources, especially nuclear stations.

ACKNOWLEDGMENT

This work is a part of IICT-BAS research project “Modeling, Optimization and Multiple-Criteria Decision Making”

REFERENCES

[1] RegulationN14/15.06.2005 of the technical requirements and norms of design, construction and use. 2005.

[2] R.Thresher, M.Robinson, P.Veers. To Capture the Wind. – In:*IEEE Power & Energy Mag*., No 6, Nov./Dec. 2007, pp.34-46, 2007.

[3] D.Anderson, A.Petersson, E.Agneholm, D.Kalsson. Kriegers Flak 640 MW off-Shor Wind Power Grid Connection – a Real Project Case Study.- *IEEE, Trans. E.C.*, vol.22, No 1, March 2007, pp.79-85, 2007.

[4] J.Nilsson, L.Bertling. Maintenance Management of Wind Power Systems Using Condition Monitoring Systems-Life Cycle Cost Analysis for Two Case Studies. – In:*IEEE, Trans E.C.*, vol.22, No 1, March 2007, pp.223-229, 2007.

[5] A.Hansen, P.Sorensen, F.Iov, F.Blaabjerb. Power Control of a Wind Farm with Active Stall Wind Turbines and AC Grid Connection Riso National Laboratory Wind Energy Department.P.O.Box 49, DK-4000, Roskilde, Dennmark.

[6] B.Ummels, M.Gibescu, E.Pelgrum, W.Kling, A.Brand. Impact of Wind Power on Thermal Generation Unit Commitment and Dispatch. – In:*IEEE, Trans. E. C.* No 1, March 2007, pp.44-51, 2007.

[7] R.Zavadil, N.Miller, E. Muljadi. Making Connections. – In:*IEEE Power&Energy Mag.* Nov./Dec. 2005, pp. 26-37, 2007.

[8] D.Anderson, A.Petersson, E.Agneholm, D.Kalsson. Kriegers Flak 640 MW off-Shor Wind Power Grid Connection – A Real Project Case Study, *IEEE, Trans. E.C.*, vol.22, No 1, March 2007, pp.79-85, 2007.

[9] L.Söder, L.Hofmann, A.Orths, H.Holttinen, Y.Wan. A Tuohy Experience from Wind Integration in Some High Penetration Areas. – In:*IEEE Trans. E.C.*, vol.22, No 1, March 2007, pp.4-12, 2007.

[10] P.Li, H.Banacar, P.K.Kenng, H.G.Far, B.T.Ooi Macromodel of Spatial Smoothing in Ooi Wind Farms.– In: *IEEE Trans. On Energy Conversion*, vol.22, No 1, March 2007, pp.119-128, 2007.

Choosing the Best Approach to Wind Energy Utilities

Aleksandar Malecic

Abstract – **In this paper are discussed different options for wind turbines (design, control strategies and location) and choices among them. Also, development from the point of view of business is mentioned.**

Keywords – **wind turbine, wind farm, control strategy, design,business.**

I. Introduction

Wind energy can be transformed into electrical energywith different approaches to design, control and allocation.Once a choice is made, it should be close to optimum. Windenergy is renewable and clean and such is its role – toprovide renewable and clean energy with an eco-friendlyapproach. It means that the chosen approach to business andtechnology must take into account society and theenvironment as a whole. Once a direction is chosen andglobalized, it is difficult to change it. There is no time for toomany mistakes. This challenge is discussed in this paper.

II. Current Situation

Wind energy technology and its implementation aredeveloping relatively fast lately, but, because its role is notjust to produce electrical energy but also to replace existingnon-renewable and harmful energy sources, thisimplementation should be as fast as possible.

Wind energy (wind mills) was actually used before theappearance of fossil fuel fired engines. The reason for thistechnology being replaced in the first place was itsinconsistency as a power source [1]. There is a hugedifference between steam engines and electricity generation,but the problem of inconsistency of wind energy and the challenge how tooverride it is similar.

"Classical" energy, compared to wind energy technology, is relatively simple for modeling and controllersare standardized. Dynamical aspects of wind turbines havevery non-linear characteristics. Wind cannot be controlledand turbines produce energy from noise (air fluctuations with veryunpredictable behaviour). Electrical grid systems demand relatively stable energy production and wind as an energy source is variable on different time scales (variable speed, seasons of the year) [1].

Hence, this technology faces withtwo big challenges: how to produce a significant amount ofelectrical energy (compared to non-renewable energysources) and how to integrate wind power into a grid systemand keep it stable.

Aleksandar Malecic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: malecic@open.telekom.rs

III. Performances and Location

The cost of generation of electrical energy from windturbines is falling because they are increasingly morepowerful and cheaper. A wind turbine converts mechanicalenergy of the wind into mechanical energy and then convertsit into electrical energy. The degree of power produced bydifferent turbines varies from a few hundred kilowatts tomegawatts.

Onshore wind farms are constructed inland in regionsrich with wind energy in order to maximize their efficiency. Offshorewind farms are located in the sea relatively far from thenearest coast. There are generally higher wind speeds. Also,if one wanted to achieve a significant penetration of windenergy percentage-wise compared to non-renewable andnon-ecological energy sources, a significant number of wind farmsshould be planned to be constructed offshore as soon as possible [2]. Their costis higher in comparison to onshore wind farms.

It is worthy to mention here the possibility of oil priceshock as a way to provoke the urge for wind energy usage. Agood historical example is the crisis during the 1970s. Thiswas a period when the era of modern wind turbines began. Itis highly uncertain when exactly will the problems related topeak oil happen (it depends on national economies and the data about oil reserves are secret), but they will happen soon.

A significant number of offshore wind farms shouldalready be working before it happens because in that casenew wind farms would be a continuation of somethingalready existing and expected social and economicalpost-peak-oil complications (or even chaos) would affectbuilding new wind turbines to a lesser extent. Also, the needfor offshore wind farms in deep waters would be a lesserchallenge because of experience, already existing companiesand hopefully lower prices.

IV. Design

The majority of commercially available wind turbines arewith the horizontal axis of rotation (compared to vertical-axisturbines) [1]. One may assume that vertical-axis windturbines are more practical because they are independent ofwind direction (horizontal-axis turbines need the yawmechanism to face the wind) and gearboxes and generatorscan be placed on the ground. This is not the case becausethey have problems during the rotation into and away fromthe wind and with high speed of the wind.

Horizontal-axis wind turbines are made of bladesadjoined at a hub connected with a nacelle. The nacellecontains components for energy transmission (mechanical(drive-train) and electrical) and control and it is placed on atower. They usually have three blades. Concerning energy capture efficiency, stability and material used, it is generally accepted that the optimal number of blades for a wind turbine is three. There are turbines with other numbers of blades, but the

majority is made with three blades. A turbine with one or two blades has problems with too fast rotation and those with more than three blades need to be thinner, use extra material (cost more without significant improvements in energy capture) and cause more turbulence (disrupt the incoming wind in a non-productive way).

V. Wind Power

The equation for the power of the wind (Eq. (1))can be written from the fact that it depends on kinetic energy ($mv^2/2$) and that the mass flow rate of air flowing through an area A isρAv (ρ is the density of air and v is the wind speed). It is:

$$P = \frac{1}{2} \cdot (\rho \cdot A \cdot \nu)^2 = \frac{1}{2} \cdot \rho \cdot A^3 \quad (1)$$

The power extracted from the wind is C_p P. C_p is the power coefficient. It is a function of λ (= ω_m R/ν – ω_m is the angular velocity and R the radius of the rotor) and β (the blade pitch angle).

Because of the fact that this is a cube function of the wind speed, it is important for wind turbines to be designed the way they support higher wind loads (protection from damage) [1]. Above wind speeds that provide the highest efficiency of a wind turbine (higher wind speeds are rare and as such have a lower power density), the power output must be controlled to reduce the load.

VI. Modelling and Control

Two different approaches to identification of a wind energy conversion system exist: the procedure based on collected data (totally independent on dynamics and structure, “black box”) and the identification that relies on its structure and dynamics (modeled drive-train and components) [3]. Deterministic components of wind fluctuations on the turbine are wind shear (when blades rotating change their heights) and tower shadow (effects of the tower on the airflow). There is also a stochastic component related to turbulence.

Control is focused on four subsystems [3]: the aerodynamic subsystem (wind energy into mechanical energy), the mechanical subsystem (the drive-train transferring the torque from the rotor to the electric generator and the support of the rotor and other devices in height), the electrical subsystem (conversion of mechanical power into electricity) and the pitch servo subsystem (the device that rotates the blades and modifies the pitch angle). Modeling and control is based on collective behaviour of these interconnected components.

A wind turbine behaves nonlinearly. Still, experiences with linear modeling with deviations from the nominal optimal point are relatively satisfying.

Passive stall control doesn’t depend on moving of adjustable parts. It is based only on the design (aerodynamics) of the blades which provide losing of power when the wind speed is too high. This method has low efficiency at low wind speeds and, as a type of control with a constant speed turbine, has lower energy efficiency than turbines with the variable speed. Active control, besides of its complexity and need for design of control mechanisms, is preferred, especially for large wind turbines.

There are three regions of operation:

1) The wind speed is too low for energy production.

2) The wind speed is high enough for energyproduction and lower than the critical value. Energycapture is important in this region.

3) The wind speed is beyond the critical value. Theobjective of the controller is protection from fatigueload.

Control strategies can theoretically be with fixed and variable speed and with fixed and variable pitch. The variable speed strategies are more common lately. Variable speed wind turbines, compared to those previously used with constant speed, adapt the blades velocity to fluctuations of the wind speed and this way operate at optimum energyefficiency (maintain C_p at the maximum value-the parameterλ mentioned earlier depends on the generator torque). In thecase of variable pitch control, the pitch angle of a bladechanges according to the wind speed when it is beyond itscritical value. The advanced control approach is dividedbetween optimal energy capture in the region 2 and loadmitigation in the region 3.

The majority of control algorithms depend on measurements from turbine structure and drive train [4]. This approach has a delayed response to the current situation and the wind speed can be very variable. This can especially be a problem at high wind speeds when loads need to be mitigated at the moment they appear. Samples of the wind speed measured upwind can be used for a preview controller which expects changes from upwind measurements. These expectations differ from the real wind speed at the location of the turbine, so they can be used only as an approximation.The challenge for this and new wind turbine technologies in general is how to develop sensors and actuators that provide local control of aerodynamic effects and that are effective, reliable and don’t add extra costs.

Different approaches todesign of controllers have been proposed, such as for example gain scheduling [3],parametric controller [1], fuzzy logic [5], and neural networks [6]. Also, pitch angles of the blades can be controlled collectively or individually. In order to produce and distribute enough electrical energy, there must be taken into account which approach is the best. Once a certain approach to control and design (and business) is globally applied, it must be very close to the best possible. There is a paradox that wind turbines are supposed to be a more sustainable option to energy production, but they still use energy and natural resources. Technological improvements of wind energy utilities and replacements of outdated ones should be observed in this context. If wind farms really significantly penetrate into electrical energy production, it will be achieved through the collaboration of people, companies and technological solutions.

Because wind farms are often connected to the powergrid without anything in between, it is preferred for wind turbines in farms to have, besides of control strategies mentioned above, power electronic technology [1]. There are different configurations of wind farms: centralized and decentralized. An obvious disadvantage of the centralized approach is that individual wind turbines don’t always operate at their optimum, but an advantage is its robustness to failures of the grid. In the decentralized control case individual wind turbines

operate at their optimum. A highvoltageDC link can also be used for wind farms. In this case there are two connected power converters, one on the side of the wind farm and another one on the side of the grid. On the side of power distribution, there are also specific requirements for keeping a balance between produced and needed power. The active power control approach means that the power demand is predicted and wind farms are adjusted to this prediction. The imbalance between production and consumption is visible if the frequency in the system moves farther from 50 Hz (or 60 Hz used in some countries). The frequency control is used to keep the balance. Voltage regulators are used to keep the voltage within the required limits. Also, tap-changing transformers are used to maintain voltage levels. Wind farms should be protected from faults in the network and external control (from people) is mostly provided.

Researches on smart grids [7], focused on improved efficiency, stability, and flexibility will also be important for implementation of wind farms and renewable energy in general. The intent is to improve the redistribution of produced electrical energy and to maximize local usage of electrical energy from renewable sources.

With the growth of share of the wind power in the system, the frequency fluctuations will increase [8]. There is a need for an additional frequency control of such an interconnected system.

VII. BUSINESS

Strong competition between wind turbine manufacturers results with the lack of widely available data. Science and business ("business as usual" historically inherited from previous industrial revolutions – the one that has caused problems in the first place) can observe the same problem with different intents. Companies try to keep their technology secret and remain competitive and scientists will maybe keep their results and ideas secret only by the moment of publishing of their scientific paper. What will at the end really matter for society and the environment as a whole is how will wind energy technology with other renewable energy sources deal with environmental problems.Motivations and expected results are not aligned.

For a company the best way to remain competitive is to protect its knowledge and ideas from "leaking". Employers in a company are not all the time with the best ideas and opportunities to find solutions. The market and competition are strong motivators for further achievements, but the results and their reflections on society, energy production and nature will be the only relevant thing when our current challenges will be observed by future generations.

New technologies and products develop like they are isolated from companies' environment. The way a new product is being made, distributed and sold does not take into account (or it does very superficially) its influence on the environment (production, existence and its further fate after it's not being used). It is difficult (if not impossible) to observe all environmental, societal and economic aspects of a technology from the point of view of competition.

Here is worthy to mention the complexity (interactions between competing ever-changing and as such highly unpredictable forces) of the world in which wind energy technologies are being developed and implemented [9]. Local communities should be independent as much as possible from the national and international energy supply. On the other hand, offshore wind turbines should be built as soon as possible because of available areas and strength and behaviour of offshore winds. Tighter local communities make sustainability or a lack of it more visible and planning of wiser spending of energy would have more chances in this case. Economic growth has its obvious advantages, but this planet is limited by its size and natural resources. This would be more obvious within relatively independent local communities. There is a paradox that at the moment the "best" way to start thinking differently about collectivity is to let the old approach fail working properly. Instead of working on real strategies, people still focus on short-term decisions and politics insisting on economic growth (and growth of energy and resources demand) and competition.

VIII. CONCLUSION

In this paper different options for wind turbines and utilities are observed, such as location, design, control strategy and connection to the grid. There are some that are more widely accepted and implemented than others. Companies and the way they work on new types of wind turbines are also considered. The urgency of the situation (environmental problems, peak oil) demands for the best technological and business solutions from the point of view of all of us and not only companies. The ways communities and long-term strategies fit in the big picture are also important.

REFERENCES

[1] J. Martinez, "Modelling and Control of Wind Turbines", master thesis, Imperial College, London, UK, 2007.

[2] D. MacKay, "Sustainable Energy- Without the hot air", UIT Cambridge, UK, 2009.

[3] F. Bianchi, H. De Battista and R. Mantz, "Wind Turbine Control Systems", Springer-Verlag, London, UK, 2007.

[4] J. Laks, L. Pao and A. Wright, "Control of Wind Turbines: Past, Present, and Future", University of Colorado, Boulder, USA, 2009.

[5] R. Adzic et al, "Maximum Power Search in Wind Turbine Based on Fuzzy Logic Control," Acta Polytechnica Hungarica, Vol. 6, No. 1, pp. 131–149, Hungary, 2009.

[6] M. Sedighizadeh, A. Rezazadeh, "Adaptive PID Controller based on Reinforcement Learning for Wind Turbine Control," World Academy of Science, Engineering and Technology, 37, pp. 257–262, 2008.

[7] M. Chertkov, "Optimization and Control Theory for Smart Grids", research report, 2009.

[8] F. W. Koch, I. Erlich and F. Shewarega, "Dynamic Simulation of Large Wind Farms Integrated in a Multi Machine Network", IEEE, 2003.

[9] M. Hawasly, D. Corne and S. Roaf, "Social Networks Save Energy: Optimising Energy Consumption in an Eco-Village via Agent-Based Simulation", Heriot-Watt University, UK, 2009.

Study of Supply Installation for Ozonation System of Wind Generator

Borislav Dimitrov[1], Emilian Bekov[2], Angel Marinov [3]

Abstract – **The applications of ozone are widely used in many industries, agriculture, medicine, etc. it being obtained by ozone generator plants. The article investigates the operation of the wind generator system – a step-up transformer corona system which enables the ozone production using renewable energy sources. The analysis focuses on the coordination characteristics of individual elements providing for maximum efficiency. There are criteria for determining the parameters of Corona System and transformers to the installed Power of wind turbines and wind resource**

Keywords– **Ozone generator, wind generator, transformer.**

I. INTRODUCTION

Industrial ozone generatingplants [1,2,4] are used to produce ozone for different applications: tank and rural water treatment, grain silos cleaning and so on. The main component of the plant is the CoronaSystem [1,3,4] with power supply voltage range of 9-15kV which is provided by means of a step-up transformer.

In the system under consideration, the corona system power supply is provided by a wind turbine, thereby using a renewable energy resource for ozone production. The wind generator operates in an autonomous-source mode, directly powering a step-up transformer and an ozone generator. A direct connection between the wind turbine, the transformer and the corona system is employed. Moreover, the conventionally used frequency converter [2] has been excluded in order to simplify the design leading to a financial solution, which is a crucial criterion for the use of autonomous renewable sources. It is necessary to study the possibility of coordinating the parameters of the individual components ensuring their optimal operation. The problem in this respect is associated with the voltage provided by the wind turbine, which depends on wind speed and has no constant value. The parameter harmonisation must exclude the possibility of emergency situations such as step-up transformer overloading, breakthrough, etc.

In the present task, the direct use of renewable resource systems without conversion (inverter) and energy storage (UPS) is justified by several considerations:

- The use of an ozone generator for tank water treatment, for example, typically requires the generation of a specified quantity of ozone in a definite time interval, i.e. a definite number of hours per day. In addition, the temporary electric supply cut-off is acceptable.
- Any losses resulting from electrical energy conversion are avoided
- The initial cost of the Water Plant and the maintenance and operational cost are reduced, which contributes to the fast investment turnover

II. ANALYSES AND EXPERIMENT

The experimental studies was based on a wind turbine with nominal parameters P=500W, U=24V, 50Hz. Fig.1 shows the record of the electrical parameters for one day, using a resistive load directly connected to the generator. For wind speed between 3 and 9 m/s, the parameters vary in the following respective ranges U = 15 ÷ 35V, I = 3 ÷12A, f = 30 ÷100Hz, with the following peaks: I = 14A, U = 43V, f = 193Hz . The preliminary measurements indicate that a standard measuring transformer 20000V/100V may be used as a step-up transformer. In the particular case investigated, the high-voltage supply in the range of 8-15kV requires the wind turbines output voltage to be doubled by means of an additional step-up transformer.

The study of the corona system is dealt with in a number of sources [1,3,4]. The corona system employed is made up of horizontal electrodes separated by glass plates. The experimental model is shown in Figure 2, and Figure 3 illustrates a wiring diagram. Both figures use the following abbreviations: WG – a Wind Generator, Tr1 – a Step-Up Transformer, doubling the voltage of the wind generator, Tr2 – a High-Voltage Measuring Transformer 20kV/100V, OG – an ozone Generator. S1, S2 – Automatic Control and Safety Switches, F - Fan.

The data obtained and indicated in tabular form (Table 1) and graphically (Figure 4) show the relation between the number of electrodes, current and voltage on the primary side of Tr2. The characteristic illustrated in Fig.4. corresponds to the wind turbines nominal voltage. An additional factor influencing the electrical parameters are the geometric dimensions of the electrodes, which should comply with the system sizing [1,4]. For the systeminvestigated, two electrodes with dimensions 0.4 x 0,4 m are used, their number being limited to 5 in order to prevent transformer overload at the peak input voltage. The one-day experimental study shows that the voltages illustrated in Fig.3 rangewithin the allowable limits U1 = 10÷40V, U2 = 20÷80V (Fig. 5), U3 = 4000÷16000V.

According to the results obtained, the wind-transformer-corona system operates properly, a transformer of power 2kW,much higher than that of the wind generator, has been deliberately selected for experimental purposes. In real operation,several small transformers connected to a common wind generator may be used (Fig. 6). This approach allows an adequate load selection depending on the available resources.

[1]Borislav Dimitrov, TU Varna, Faculty of Energetic, Studentska 1, 9010, Varna, Bulgaria, E-mail: bdimitrov@processmodeling.com
[2]Emilian Bekov, TU Varna, Faculty of Electronic, Studentska 1, 9010, Varna, Bulgaria, E-mail:emo_bekov@hotmail.com
[3]Angel Marinov, TU Varna, Faculty of Electronic, Studentska 1, 9010, Varna, Bulgaria, E-mail: igdraz@abv.bg

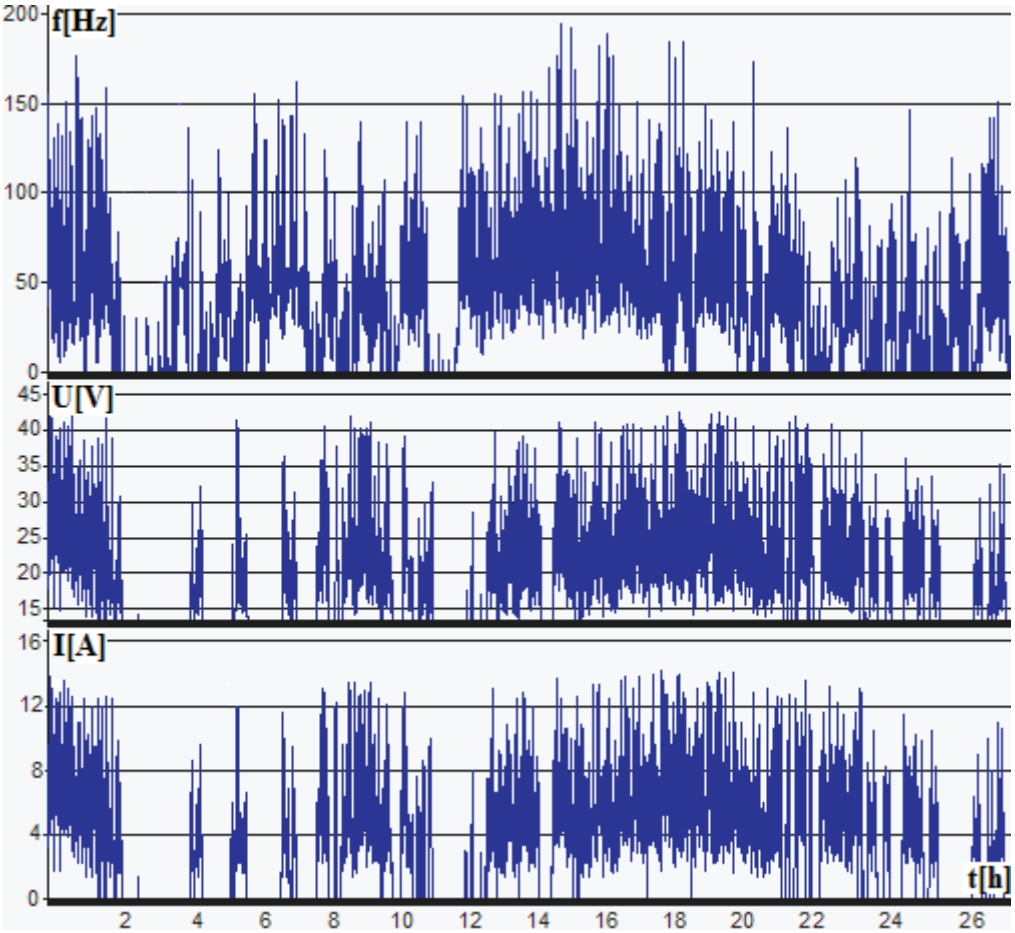


Fig. 1. Data obtained by a Wind Generator Study.

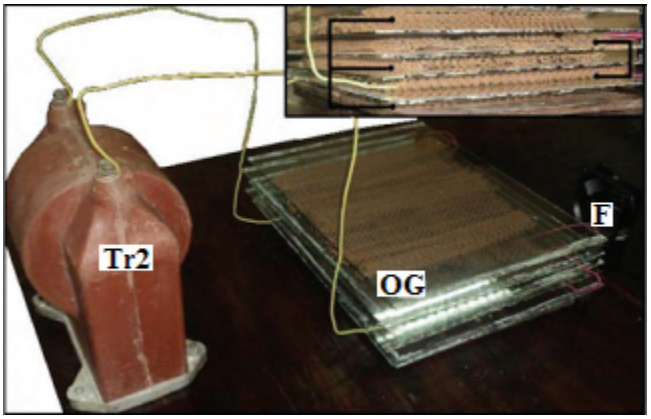


Fig.2. An Experimental Corona System Model powered by a measuring high-voltage transformer.

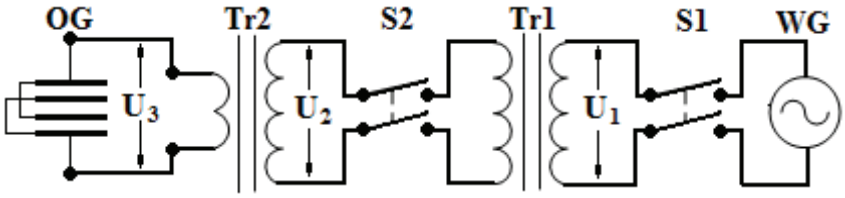


Fig.3 A Corona System Power Supply incl. a transformer and a wind generator.

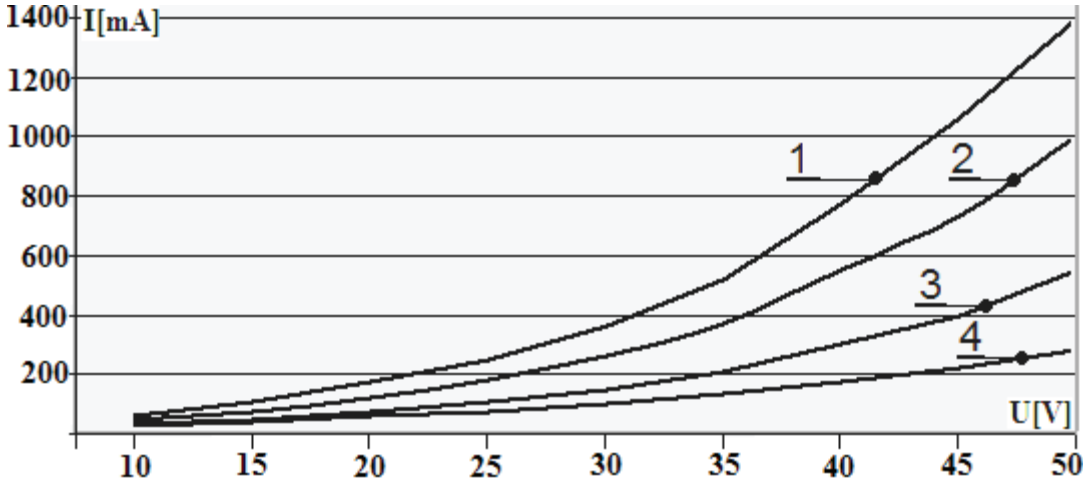


Fig.4 Data obtained according to the different number corona system electrodes used.
Graphs 1÷4 resp. 5÷2 electrodes.

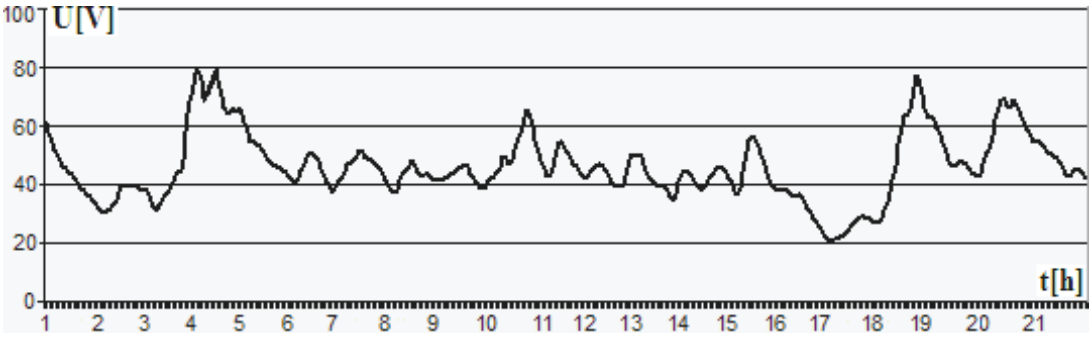


Fig.5. Step-up Transformer Low-Voltage side (Tr2, fig.3).

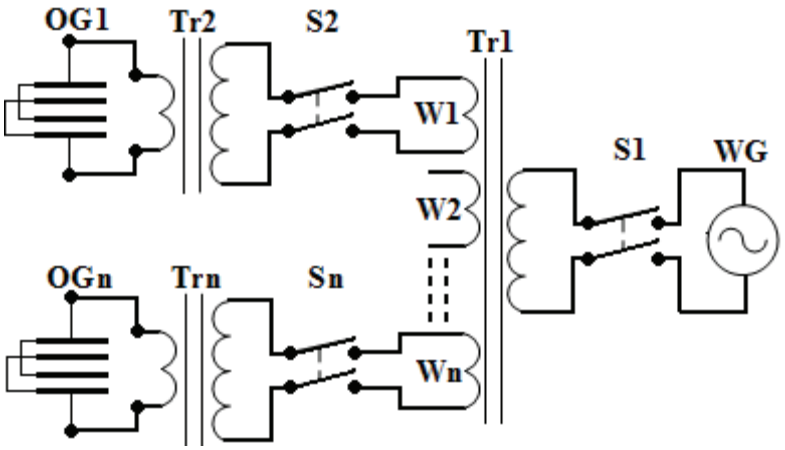


Fig.6 Supply Multiple Corona System.

TABLE 1.

U [V]	Number of Electrodes [mA]			
	2	3	4	5
10	28,59	31,03	44,31	61
15	42,61	49,31	75,42	105
20	59,41	73,64	119,33	172
25	76,56	107,91	181,12	250
30	104,23	147,25	260,51	360
35	134,34	210,04	370,17	520
40	174,77	301,06	550,91	770
45	220,12	405,43	732,45	1060
50	280,03	555,32	1000,34	1400

III. CONCLUSION

From the studies conducted, it can be concluded that the autonomous use of renewable energy sources is appropriate for generating ozone by means of an ozone generating plant. The parameters of the wind generator – step-up transformer – corona system can be harmonised so as to enable direct connection between them (Fig. 3). It is necessary to ensure acceptable mode of operation of the installation, excluding emergency situations of overloading and insulation breakthrough. This can be achieved by following the sequence proposed in the present paper:

- A precondition for proper parameter harmonisation is the use of available data from a prior research (Figure 1) on the operation of the wind turbine. It is due to the variable output voltage depending on wind speed.
- It is necessary to ensure the input voltage allowable range for the high-voltage step-up transformer (Tr2). In the particular research, this is achieved by means of an intermediate step-up transformer (Tr1), but in other cases it may be a step-down one.
- Finally, the overall experiment (Fig.4, Fig.5) should confirm acceptable modes of operation of the system components.

An additional option is to use multiple corona systems fed by a common wind generator (Fig. 6).

REFERENCES

[1] Блинов И.В., Ваняев С.В., Кузнецов К.Ю. Анализ энергетических процессов в высокочастотном озонаторе. Межвузовский сб. научн. трудов Электрооборудование промышленных установок. НГТУ, Н. Новгород, 2006г.

[2] Кузнецов К.Ю., Ларионов Н.П., Блинов И.В. Источники питания промышленных озонаторов. XXII научн.-техн.конференции «Актуальные проблемы электроэнергетики» Тезисы докладов/ НГТУ, 2003 г – с. 67-69.

[3] Панайотов М, Б. Димитров. Изследване на конструктивните и технологични отклонения в озон генераторните станции Международна научно-техническа конференция Електроенергетика 2010. стр. 301-306

[4] Пузиков Н.А., Блинов И.В., Кузнецов К.Ю., Махин Ю.И. Снижение энергозатрат озонных технологий. Тезисы докладов межд. Конгресса «Великие реки 2005», том 1. Н.Новгород. -ННГАСУ, 2005. т.1 – с. 196-198.

Analyses of Characteristics and Efficiency of Fuel Cell

Emilian Bekov[1], Borislav Dimitrov[2] , Angel Marinov[3]

Abstract – **The basic principles of operation of a proton exchange membrane fuel cell (PEMFC) are presented. Practical experiments with PEMFC are made. Current – voltage characteristicin electrolyses, voltage, current and power characteristics of PEMFC are given. The efficiencies at different operating modes are determined and presented. Final conclusion about characteristics and efficiency of PEMFC are derived.**

Keywords– **Fuel Cell, Proton Exchange Membrane.**

I. INTRODUCTION

Renewable energy is one of the most important and prominentsourcesfor energy generation. Hydrogen fuel cells (FC) are a new and interesting concept of an alternative energy source [1]. Different types fuel cells are used for energy generation. In the last few years the Proton Exchange Membrane (PEM) fuel cells are one of the leadersin the field of fuel cells. In this paper the analysis of characteristics of the PEM fuel cell ispresented. The main characteristics are [2]:

- The electrolyze characteristic;
- The current, voltage and power characteristics on FC;
- The characteristics in different load modes.

PEM fuel cells are efficient in transportation, communication and general power supplies. Fuel cellsrun on hydrogen and oxygen and produce water as a byproduct of the energy conversion. They are simpleand have no moving parts; can be used in stack installation to increase power load of the wholesystem.

II. ANALYSES AND CHARACTERISTICS OF PEM FUEL CELL

Theanalyses and characteristics are made usingProton Exchange Membrane Fuel Cell of Horizon Company. The active area of fuel cell is 10 cm^2 [3].

A. The electrolyser

The first analysis is to determine the current and voltage relationship in electrolyses mode as well as the minimum voltage of the electrolyses of water [4]. The experimental data is given in Table 1. The current - voltage characteristic is shown on fig 1.

TABLE I
EXPERIMENTAL DATA DURING ELECTROLYSIS

V [V]	I [A]
1,4	0,025
1,5	0,03
1,6	0,075
1,7	0,12
1,8	0,15
1,9	0,19
2	0,22
2,1	0,26
2,2	0,29
2,3	0,33
2,4	0,38
2,5	0,42
2,6	0,47
2,7	0,52
2,8	0,57
2,9	0,61
3	0,65

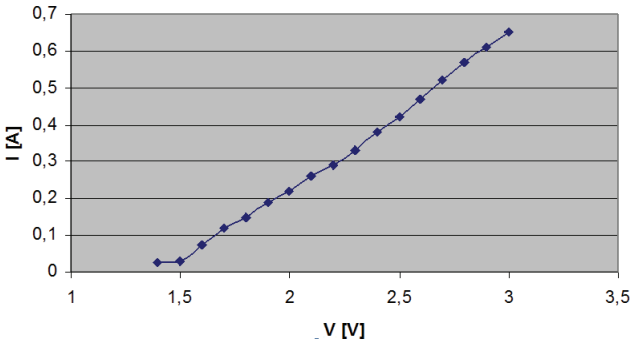


Fig.1. Current - voltage characteristicin electrolyzes

The minimum voltage of the electrolyses of water is 1,4 V.

B. The characteristics curves of the hydrogen PEM fuel cell

The secondanalysis is to determine the current, voltage and powercharacteristics of the hydrogen fuel cell. The experimental data are given in Table 2.

TABLE II
EXPERIMENTAL DATA OF VOLTAGE AND POWER

R [ohm]	V [V]	I [A]	P [W]
200	1,2	0,02	0,024
150	1,05	0,03	0,0315
100	1,0	0,04	0,04
50	0,95	0,05	0,0475
20	0,9	0,06	0,054
10	0,85	0,07	0,0595
5	0,8	0,1	0,08
2	0,75	0,25	0,175
1	0,7	0,35	0,245
0,5	0,65	0,40	0,280

[1]Emilian Bekov, TU Varna, Faculty of Electronic, Studentska 1, 9010, Varna, Bulgaria, E-mail: emo_bekov@hotmail.com
[2]Borislav Dimitrov, TU Varna, Faculty of Energetic, Studentska 1, 9010, Varna, Bulgaria, E-mail: bdimitrov@processmodeling.com
[3]Angel Marinov, TU Varna, Faculty of Electronic, Studentska 1, 9010, Varna, Bulgaria, E-mail: igdraz@abv.bg

The voltage - current characteristic is shown on fig 2.

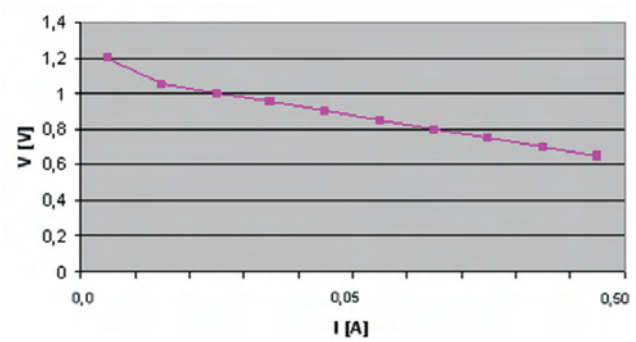


Fig.2. Voltage - current characteristic

The working voltage is around 0.9 V. The working current is up to 400 mA.

The power characteristic is shown on fig 3.

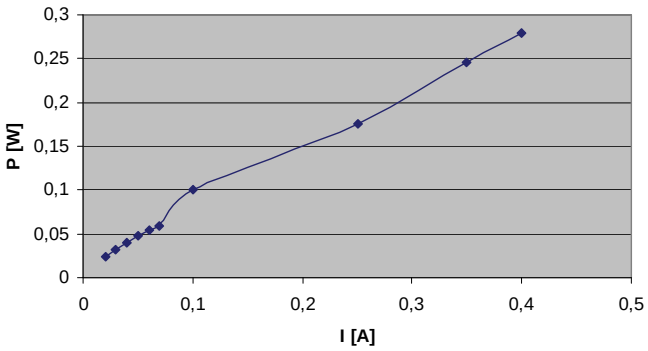


Fig.3. Power– current characteristic

The working power is around 0.25 W.

C. The load characteristics

The loadanalysis determines the operation time of the Horizon hydrogen fuel cell car [3] in different loads. The hydrogen fuel cell is electrolyzing for 30 seconds. The experiments are made in four different load modes on the car. The experimental data is given in Table 3.

TABLE III
EXPERIMENTAL DATA OF VOLTAGE AND POWER

Load [kg]	Time [s]	Curve
0,220	125	1
0,440	65	2
0,880	40	3

The load characteristics are shown on fig 4.

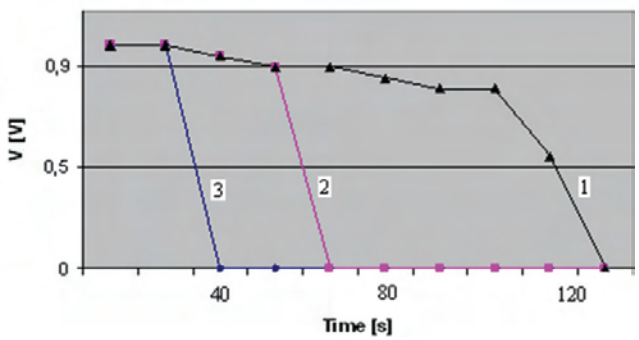


Fig.4.The load characteristics

Increasing onthe load causesthe operation time to decreasein progression.

III. CONCLUSION

The paper shows the main features and important characteristics of a Proton Exchange Membrane Fuel Cell. This study illustrates a big potential of one small fuel cell for energy generation.The minimum voltage of the electrolyses of water is obtained. The working voltages, current and power are shown. The load analyses of hydrogen fuel cell car are presented. The working time in different load modes is given. We concluded that the fuel cell system is effective for driving car and can be alternative of battery power supply. In the future work, we purposed,using stack fuel cells to increased power factor of the system.

ACKNOWLEDGEMENT

The authors express their gratitude to project DO-02-48/10.12.2008, National Scientific Foundation, Bulgaria.

REFERENCES

[1] C. Rayment, S. Sherwin, “Introduction to Fuel Cell Technology”, University of Notre Dame, USA,2003.
[2] P. Christ, D. Biedermann,"Fuel Cell", University of Minchin, Germany, 1999.
[3] B. Holland, J. Zhm, L. Jamef, “Fuel Cell Technology and Application”, France, 2007.
[4] M. Isa,E. Ismail, I. Daut, “Characteristics and the Efficiently of Fuel Cell”, American Journal of Applied Sciences, 2006.

Modeling and Analysis of µCHP System for Domestic Use

Angel Marinov[1], Vencislav Valchev[2] , Georgi Nikolov [3]

Abstract – **Combine heat and power (CHP) is a well known technique that maximizes the utilization of primary fuel sources, by recovering losses and rejected energy. The modern technology allows instantaneous generation of heat and power to be scaled into so called micro CHP systems. These units are applicable to single or blocks of households. Most µCHP systems run on fuels with low CO2 emissions – mostly biomass and thus have a status of renewable energy generation. This paper presents a study of µCHP systems concentrating on their usefulness both on social and consumer level. The study provides environmental, economical and parametric analysis on the µCHP systems. A dedicated model on which the analysis is conducted is developed and presented in the paper.**

Keywords– **Biomass, Combined Heat and Power, Efficiency, Micro-generation, Renewable energy.**

I. INTRODUCTION

Domestic energy consumption takes more than 26% of the total energy consumption in the European Union. More than 60% of that power is used to generate space heating. Domestic heating can be either supplied by a centralized source (heating power plant, gas, electricity) or generated locally by various decentralized means (solar thermal, burning of fossil fuels or solid biomass). Both types of power generation have their advantages and disadvantages. [1]

Decentralized heating attracts users by: breaking their dependence from monopolized central energy structures; allowing them to choose and fine tune power generation based on their specific energy demand; provide economical benefits over years. Decentralized heating systems have also significant social advantages such as: reducing fuel poverty; allowing the use of fast growing solid biomass with low CO_2 emissions; increasing the power generation without the need of improving the distribution system.

Technological development has reduced the size of decentralized heating units, increased their efficiency and allowed the use of various fuels. It is only reasonable to improve and further develop them by introducing large scale generation techniques to the small scale decentralized units. One such solution is the development of micro systems for combined heat and power (µCHP). Thattype of systems utilize the simultaneous generation of heat and power, where heat is the main energy product and power is a byproduct. µCHP allows better fuel utilization, greater economical benefits to users as well generation based on low carbon solid biomass such as: pallets derived from wood and agricultural waste; energy crops. In this wayµCHP systems can be consider renewable energy generation, which allows the use of subsidies as well as better feed-in tariffs where grid injection is available.[3]

The paper presents a study on the use and efficiency of µCHP. A major point in the study is the usefulness of the system to the end consumer. As various parameters are involved, the study is conducted using a MATLAB – Simulink model.

[1]Angel Marinov is with the Technical University of Varna, Studentska str. N1, 9000Varna, Bulgaria, E-mail: igdrazil@abv.bg

[2]Vencislav Valchev is with the Technical University of Varna, Studentska str. N1, 9000 Varna, Bulgaria, E-mail: vencivalchev@hotmail.bg

[3]Georgi Nikolov is with the Technical University of Varna, Studentska str. N1, 9000 Varna, Bulgaria, E-mail: gtn@gbg.bg

II. MODELLING A MICRO CHP SYSTEM

A. Modelling parameters

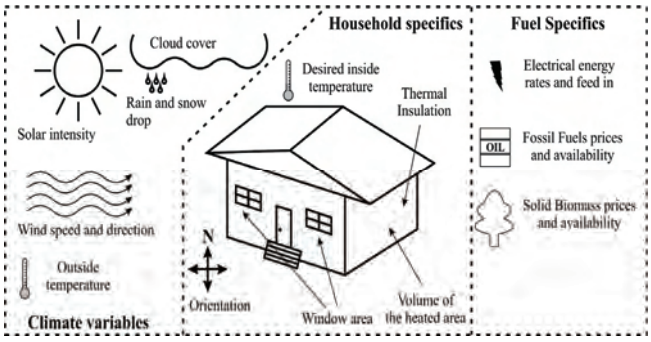


Fig. 1. Variables affecting heat energy generation

In order to evaluate the properties of a µCHP a correct determination of the required energy generation for a household and respectively its price is needed. This involves a wide range of parameters. Those parameters are depicted on figure 1. They can be briefly summarized into three groups, as follows:

Climate variables–this set of parameters involves all climate changes and properties specific to the region that can affect the thermal generation. Such as: solar intensity; cloud cover; rain and snow drop; wind speed and direction; outside temperature. Those parameters have major effect on the level at which the inside heat will be dissipated and lost.[2,4]

Household specifics– this set of parameters covers all variables and constants specific to the house construction and position. The major effect here is the volume that requires to be heated, the insulation that will prevent the thermal energy to "escape" and the required inside temperature.[2,4]

Fuel specifics–shows the availability and prices of the fuels specific to the area. Those parameters describe not only the

properties of the fuel on which the system will run but also the rates and tariffs of the electrical energy, since they will be used as a basis when the system injects power. [2,4]

B. MATLAB – Simulink model

The study presented in the current paper, describing the usefulness and efficiency of the µCHP is determined by the a dedicated MATLAB Simulink model. The model structure and block diagram is presented on figure 2. The model has a basic and simple composition and only partly covers the above mentioned parameters.

It can be divided into several functional blocks:

• Simulation of the electrical consumption and system – describes the energy consumption of the household, based on a statistical evaluation. It shows the electrical energy needs in order to describe how they can be addressed by the µCHP system.

• House – presented by its volume, thermal insulation and window area.

• µCHP system – presented by a burner, organic rankine cycle block (as for this specific system composition), a thermostat, a generator and a convertor part. The µCHP is controlled using hysteresis which activates and deactivates the system based on the outside temperature and the desired indoor temperature. The rate of witch heat and power are

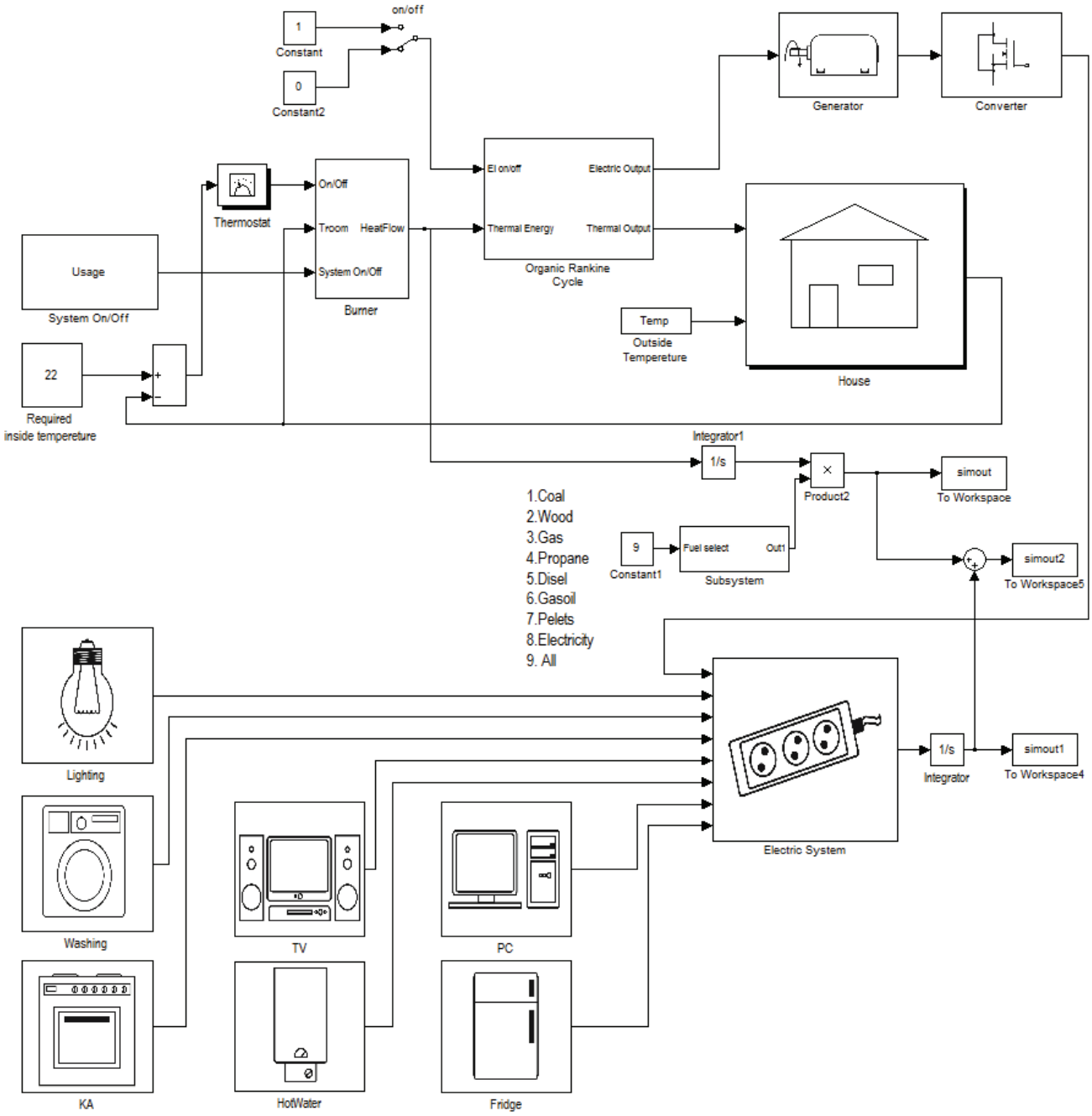


Fig. 2. Model of a µCHP for domestic use

generated is 80% heat, 20% power.

- Fuel section system – allows a broad range of fuels and their properties to be loaded in the model, so an optimal fuel for a given case can be selected.

The model was confirmed and verified by running it and comparing it with existing data on commercial conventional heating and µCHP systems.

III. Main Simulation Parameters

The simulation used to analyze the µCHP and the results presented in this paper are based on:

A house with the following parameters: area equal to 100m^2; height 2.80m; window area 300m^2;

The weather conditions that are used are for the area of the Technical University of Varna, where a weather station is operational and day to day data is available. Mean temperatures for each day of 2010 are used.

Fuel prices and energy values are provided by a local supplier. Fuel data used in the simulations is presented in Table I.

TABLE I
FUEL PRICES AND ENERGY VALUES

Fuel Type	Price	Energy Value[
Coal	75 EUR/t	3,72 kWh/kg
Wood	72,64 EUR/t	3,14 kWh/kg
Electricity	0,86 EUR/kWh	1,00 kWh
Nature gas	402 EUR/ 1000 nm3	9,01 kWh/m3
Butane	1050,8 EUR/t	12,80 kWh/kg
Diesel	1028,7 EUR/t	11,63 kWh/kg
Gasoil	950,15 EUR/t	10,98 kWh/kg
Pellets	190,02 EUR/t	4,88 kWh/kg

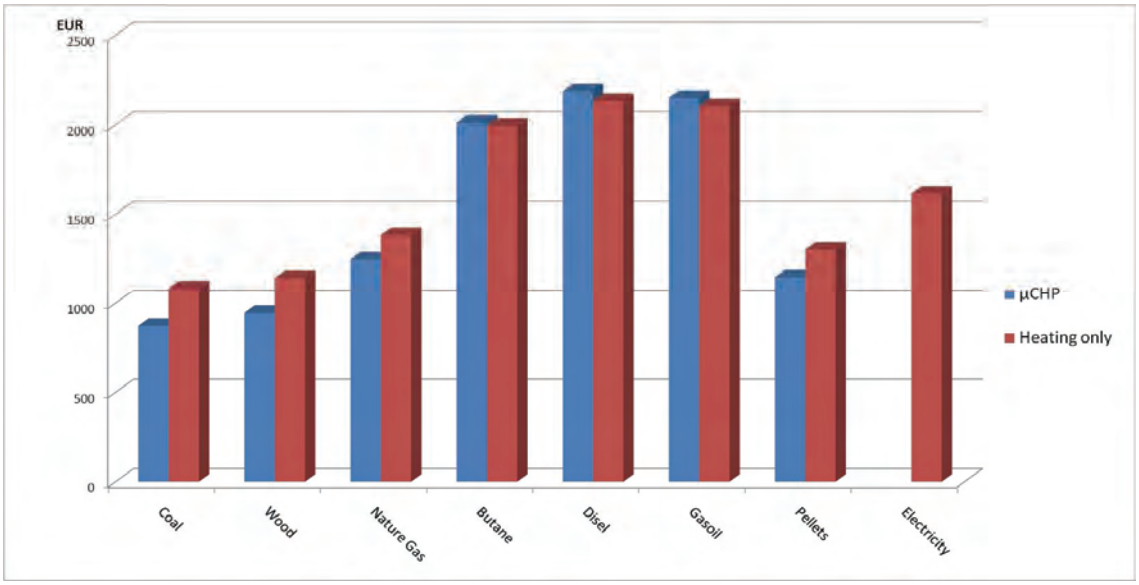


Fig. 3. Comparison between energy expenses of heating only system and µCHP

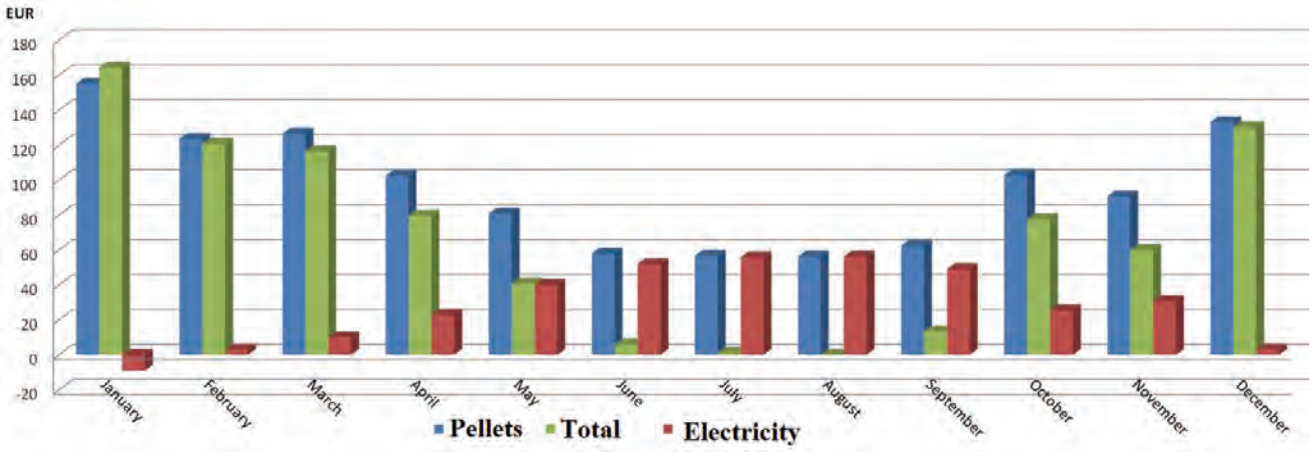


Fig. 4. Energy expenses of µCHP running on pellets

IV. Simulation Results

General simulation results are present at figures 3, 4, 5, 6 and 7

The simulation results show that the µCHP system has better economical efficiency then conventional heating systems – especially with the more inexpensive fuel types – coal, wood, and pellets (fig. 3). The total expenses when using µCHP show 13 to 20% less energy expenses compared to conventional heating systems (fig.5) and 28 to 46% less energy expenses compared to heating using electrical energy (fig. 6).

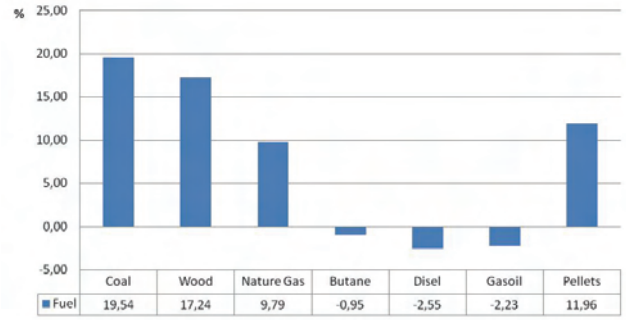


Fig. 5. Economical efficiency of µCHP systems relative conventional heating systems

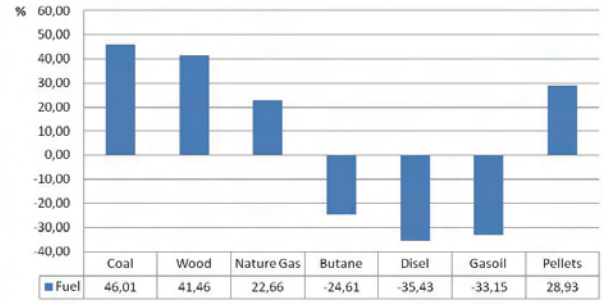


Fig. 6. Economical efficiency of µCHP systems relative to electrical energy

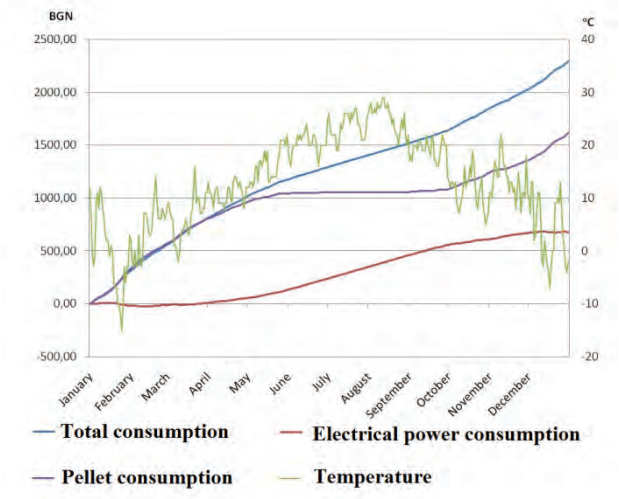


Fig. 7. Electrical and fuel energy expenses relative to the outside temperature

Wood and coal show the best financial values but also have some disadvantages, considering that both types cannot be described as renewable energy fuel. Pellets on the other hand can be produced form agricultural and wood waste, have less CO_2 emissions and produce less disposable waste (such as ash and sludge). That's why pallets where chosen to further the simulation.

Spanning the simulation for 1 year period shows that using µCHPcan, during the "cold" months eliminate or majorly reduce the electrical energy bill by generating electricity and consuming it on spot or feeding it in the electrical grid (fig. 4 and fig, 7).

V. Conclusions

Using the values and graphs generated by the model one can determine the general efficiency for a µCHP as well as the parameters that affect it. The presented results show that in this given case the reduced expenses, compared to a conventional heating system are not enough to cover the higher investment on a µCHP system – only 20% better performance.

It is clear however that the economical efficiency on the system depends on the number of cold days and the price of the electrical energy. Thus the performance of µCHP for this general case can be explained with the smaller number of cold days in Varna and the relatively inexpensive price of the electrical energy. Further investigation is considered where the µCHP will be simulated for different EU countries, various climatic and economic factors will be involved.

Acknowledgement

The authors would like to express their gratitude to the project DO-02-48/10.12.2008, National Scientific Fund, Bulgaria.

References

[1] www.energy.eu
[2] James E. Brumbaugh, "HVAC Fundamentals", Wiley Publishing, 2004, ISBN: 0-764-54206-0.
[3] V. Valchev, A. Marinov (2009), "Overview and comparison of renewable microgeneration with Combined Heat Power systems" – ICEST09, V. Tarnovo, Bulgaria, 2009, cd-rom.
[4] Robert A. Gustafson, "ASHREA 2000 HVAC Systems and Equipment Hanbook", ISBN 1-931862-47-8

Examination Parameters of SomeBasic Construction of the Brown Gas Generators

Rosen Vasilev[1], Ivailo Nedelchev[2] , Venko Venkov[3] and Angel Marinov[4]

Abstract – **Increasing fossil fuel consumption, enforce much more need of developing and using construction, which generate so called renewable sources of energy. One of these sources is water. There is an opportunity to use energy saved in it. Dissociation of the water with the help of electric current, create mixture of oxyhydrogen gas (brown gas), which can be used like alternative fuel.**

In this article are represented a few brown gas generator's constructions and some of its main parameters are examined. The experiments are conducted with generators working on DC current. The main parameters, observed in the experiments are: voltage, current, oxyhydrogen gas production and efficiency of the construction.

Keywords– **Brown gas, Generators, electrolyser.**

I. INTRODUCTION

One method to produce hydrogen is electrolysis. In this way applying electrical dc current in the solution or melt, on the electrodes the elements are dissociated into different components. This process represents breaking down of the chemical compounds. Anions (negative ions) are separated on the positive electrode anode, and cations (positive ions) on the negative – cathode. Electrolysis of water produces on the both electrodes two gases – hydrogen and oxygen. The hydrogen is separated on the cathode and oxygen – on the anode. They are in molecular state and saturate the solution with small bubbles in proximity to electrodes. After dissociation of the water molecule, generated gases mixes each other (if construction allows that) into combustible compound consist of two part hydrogen and one part oxygen. This is so called brown gas.

In the electrolyser begins to flow the follow processes:

On the cathode:

$$2e^- + 2H_2O \rightarrow H_2 + 2OH^- \quad (1)$$

On the anode:

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^- \quad (2)$$

This means that the cathode release two electrons and two hydroxyl ions and one molecule of hydrogen are produced from two molecules of water. In the other side anode accept four electrons and create four hydroxyl ions and one molecule of oxygen.

[1]Rosen Vasilev is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: rsnvasilev@abv.bg.

[2]Ivaylo Nedelchev is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: ivonedelchev@yahoo.com.

[3]Venko Venkov is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: rsnvasilev@abv.bg

[4]Angel Marinov is with the Faculty of Electronics, TU of Varna, 9010 Varna, Bulgaria, E-mail: igdrazil@abv.bg

The equation of that process can be expressed like that:

$$2H_2O \rightarrow O_2 + 2H_2 ; E = +1{,}229V , \quad (3)$$

what means, that for dissociation of the water is needed applying external electrical field equivalent to +1,23V. Practically for starting the process of the electrolysis and acceleration the brown gas generation, is necessary catalyser in the solution (KaOH, NaOH, NaSO4). For realization of a reliable electrolyser, is necessary not only good catalyser, but electrodes from durable

material resistible on the aggressive effect of the base and solution. In other case corrosion will destroy the electrodes.

II. STATEMENT

There exist a big variety of constructions of different electrolysers. They has equal principal of operation, but different number and shape of the electrodes, and its localization. Common type of the system for brown gas generation is shown on the Figure 1. It was made different types of electrolysers constructions in order to estimate their efficiency and reliability.

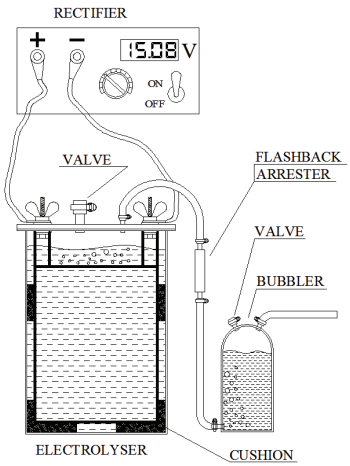


Fig. 1. Common type of the system for brown gas generation

The efficiency was evaluated using both Faraday's lows for electrolysis, which can be expressed with the equation:

$$m = \frac{Q}{F}.\frac{M}{z} , \quad (4)$$

where:

m is mass of the materials released over the electrodes.

Q – quantity of the electrical charge, used in the electrolysis,

M – molar mass of the elements,

z – valency,

F – Faraday's constant (F=96 485 C/mol),

and free Gibbs energy which is equal to:

$$G = -nFE, \quad (5)$$

where:

N is number of the molls of electrons, participated in the reaction.

F - Faraday's constant,

E – electrical potential between anode and cathode.

For estimation efficiency is introduced a quantity MMW – ml/W/min, which in fact is volume of created brown gas (in milliliters) divided on amount of energy needed for its production (in wattminutes):

1MMW=1ml/1Wmin.

According Faraday's lows is obtained that 100% efficiency will be reached when MMW is equal to 9,282.

For electrolysis cells were used electrodes made from stainless steel 316L type. That type consist molybdenum which support better redox reaction. For electrical measurements was used digital multimeter type MY-64 and current clamps type MS- 3302. It was tested constructions with anode, cathode and different numbers of neutral plates between them Figure 2.

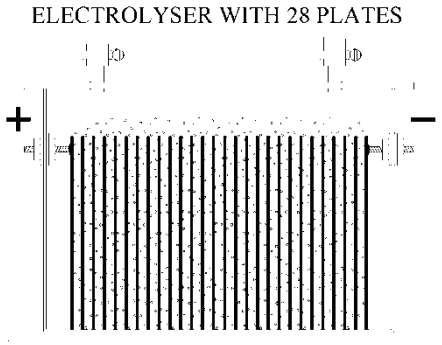


Fig. 2. Electrolyser with 28 plates

In this way from n number of electrodes are obtained n-1 numbers of cells. When voltage U is applied between anode and cathode, every internal cell gets Un/n-1 volts. One important parameter of an electrolyser is optimum drop of the voltage over each one cell which ensures maximum efficiency. The graph below (Figure.3) shows V-A characteristics of the constructions with different number of plates.

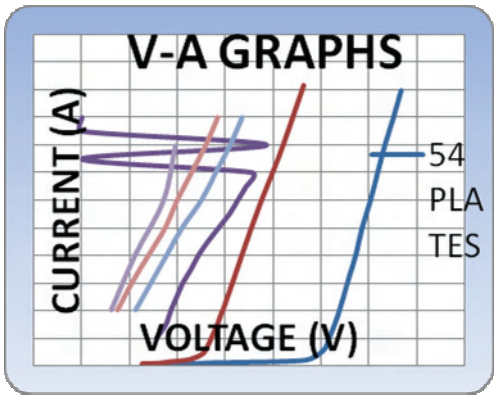


Fig. 3. V-A characteristics of the constructions with different number of plates

Dependencies at low voltage to the start of electrolysis are nonlinear, then increase release of gas and curves becomes almost linear as the steepness is maintained with increasing voltage. During the measurements it was found that the highest efficiency is obtained when on each cell of the electrolyser, a voltage drop is in the range (2V ÷ 2,2 V).

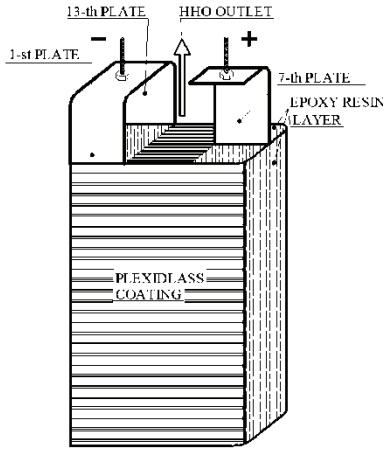


Fig. 4. The design of the work cell consisting of 13 plates

In electrolysis, a huge influence affects currents whose contours are closed in more distant electrodes. These currents represent a loss of efficiency, since their energy by polarizing electrolyte rather than producing gas. To avoid such losses, was designed electrolyser under which was minimized adverse currents, and in fact there are only losses in the cells themselves between adjacent electrodes. Thus, currents flowing around cells of the electrolyser are minimal and achieve less energy input per unit volume of produced gas. The design of the work cell consisting of 13 plates is shown in Figure 4, and the resulting performance data for design are in Table 1.

TABLE I
THE RESULTING FOR DESIGN OF FIGURE 4

1	2	3	4
I(A)	U(V)	V/t(ml/min)	MMW
current	voltage	flow	efficiency
2,6	10,8	110,1	3,92
2,8	10,9	134,9	4,42
5,5	12	279,8	4,24
7,5	12,7	414,33	4,35
10,5	13,2	591,82	4,27
13,2	14	778	4,21
15,4	14,4	929,17	4,19

It was eliminated not only the currents of the long sides between the electrodes, but some of those at the short sides. In the middle of them was left only a small slit that maintains circulation of electrolyte and the evacuation of gas production (Figure4).

Another type of design is shown on Figure 5.

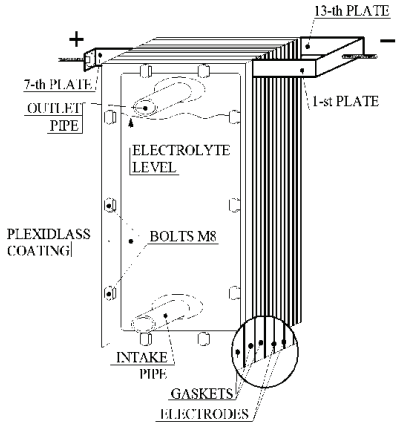


Fig. 5. The design of another type electrolyser

While previous types of electrodes are immersed in the electrolyte, this structure represents isolated cells with an electrolyte filled between them. These are the so-called dry electrolysers. Structures were tested with 4, 5 and 6 neutral plates i.e. 11, 13 and 15 plates. There were experiments with different distance between the plates - 2, 3 and 5mm.

TABLE II
THE RESULTING FOR DRY CELL-11 PLATES, 2MM GASKETS

1	2	3	4
I(A)	U(V)	V/t(ml/min)	MMW
current	voltage	flow	efficiency
3,4	10,25	185	5,29
6,35	11,55	356	4,85
15,8	14,25	808	3,59

TABLE III
THE RESULTING FOR DRY CELL-13 PLATES, 2MM GASKETS

1	2	3	4
I(A)	U(V)	V/t(ml/min)	MMW
current	voltage	flow	efficiency
2,5	11,85	170	5,71
7,4	12,75	525	5,56
10,2	13,3	750	5,28

It appears that reducing the distance between the plates is reduced and optimum voltage on each cell, yielding maximum efficiency. This leads to lower losses and hence to better performance of the electrolyser. On the other hand, the distance between the plates, cannot be too small (for example less than 2mm), as it hinders the conveyance of gas bubbles "adhered" on the plates, and thus slowing the process of electrolysis. Survey data are presented in Tables 2 ÷ 5. For better performance have to be ensured appropriate aperture on the plates for evacuation of the brown gas in order to prevent any undesired circumstances such like increasing of the pressure, leakage of electrolyte, slowdown the reaction.

TABLE IV
THE RESULTING FOR DRY CELL-15 PLATES, 2MM GASKETS

1	2	3	4
I(A)	U(V)	V/t(ml/min)	MMW
current	voltage	flow	efficiency
4,9	14,75	382	5,28
8,3	15,5	650	5,05
12,5	15,86	955	4,81

TABLE V
THE RESULTING FOR DRY CELL-13 PLATES, 5MM GASKETS

1	2	3	4
I(A)	U(V)	V/t(ml/min)	MMW
current	voltage	flow	efficiency
1,5	11,9	106	5,91
9	13,6	592	4,83
11,2	14,6	677	4,14

III. CONCLUSION

The effectiveness of an electrolyser improves when:

1. It works with smaller currents. This is explained by the smaller heat losses, which depend on the square of current, i.e. the lower the temperature of the electrolyte, the better is the effectiveness of the construction.

2. Reduces the distances between the plates, leading to lower ohmic losses, as well as those of polarization. The optimum distance is about 2mm. and depends on many parameters.

3. Reducing any adverse currents between distant plates, which do not participate in gas production.

In general, "dry" cell is more efficient, but their parameters are more unstable in terms of working currents and are less reliable because of the danger of leakage through gaskets. Immersion cell that was examined with more stable parameters throughout the operating range and better reliability, but inferior in efficiency by about 10%.

REFERENCES

[1] Patrick J. Kelly, “A practical guide to “Free Energy” devices”,pp.2200, ebook.
[2] F. M. Kanarev ,"Water as a new source of energy", ebook.
[3] http://h2science.blogspot.com/.
[4] http://en.wikipedia.org/wiki/Gibbs_free_energy

Calculation of PVGIS Solar Data for the Territory of Serbia

Dušan Ž. Djurdjević[1]

Abstract – **In this paper data for annual solar irradiation and potential power production by PV power installations for the territory of the Republic of Serbia is calculated by using the PVGIS interactive on-line calculator. The calculations performed can serve as a helpful guide for initial practical activities in the Serbian PV power engineering field.**

Keywords– **Renewable energy sources (RES), Photovoltaics (PVs), Serbia, PVGIS.**

I. INTRODUCTION

Threads of the global climate changes and environmental pollution due to the use of conventional energy sources have enhancedresearch in the field of the renewable energy sources (RES). Photovoltaic (PV) solar electricity generation iscertainly the most attractive method and possibility to solve global energy problem in the world-wide scale. As the direct energy conversion technology, PV solar power systems are noiseless and environmentally almost completely friendly, having a huge potential to win the race with the traditional energy sources and other RES possibilities.

Various PV estimation tools offer solar radiation and other climatic data useful for an assessment of the PV potential for specific location world-wide: European Solar Radiation Atlas (ESRA), SoDa, NASA SSE, Meteonorm, NREL – US dynamic solar atlas, etc. One of the most popular and easy-accessible solar resource and tools for an assessment of PV potentials and systems is PVGIS© [1]. Amongst other solar radiation databases, PVGIS has advantages as an open data and software PV estimation tool with an excellent geographical grid resolution (1km x 1km) and map based user-friendly interface, providing easy-understandable information for PV geographical assessments. The estimated accuracy of PVGIS calculations is proven to be within several percents [2]. Detailed geographical, climatic and other data make PVGIS on-line calculator ideally suited, not only for non-professionals and initial PV system estimations, but also even for serious PV systems design as part of the integrated management of distributed energy generation, for specifically selected locations in Europe, and most recently for Africa, as well.

In this paper maps of average solar irradiation and solar data for the specificlocation within territory of Serbia arranged by using PVGIS © on-line interactive calculator [1] are presented and discussed.More extensive analysis and results about practical data for solar irradiation and estimated power production by a PV system set up within the Serbian territory can be found in [3]. Although utilization of RES in Serbia is so far limited to micro and mini hydro power-plants, Serbia has large unused potential for production of energy from RES [4] (biomass and biogas resources, geothermal, wind energy potential, non-utilized hydro-power potential and solar energy resources).

[1]Dušan Ž. Djurdjević is with the Faculty of Technical Sciences, Knjaza Miloša 7, 38220 Kosovska Mitrovica, Serbia, E-mail: dusan.djurdjevic@pr.ac.rs

II. PVGIS

PVGIS (*Photovoltaic Geographical Information System*– PVGIS © European Communities, 2001-2008) is a part of the SOLAREC action aimed at contributing to the implementation of renewable energy in the EU, [5]. SOLAREC is an internally-funded project on PV solar energy for the 7th Framework Programme. PVGIS has been developed at the JRC (Joint Research Centre) of the European Commission within its Renewable Energies Unit since 2001 as a research GIS oriented tool for the performance assessment of solar PV systems in European geographical regions. At the very beginning PVGIS was planned to be an in-house decision support system, fortunately access to the PVGIS database and estimations has been made freely available to professionals and the general European public through web-based interactive applications. PVGIS is aimed at providing data to analyse the technical, environmental and socio-economic factors of solar PV electricity generation in Europe and to support systems for policy-making in EU countries. More about PVGIS and the data sources and methodology used can be found on the PVGIS official web-site [1] and references [2,3,6-8].

PVGIS methodology takes into account not only solar radiation data, it considers PV module surface inclination and orientation and shadowing effect of the local terrain features (e.g. when the direct irradiation component is shadowed by the mountains), therefore PVGIS is a powerful PV implementation assessment tool that takes into account the dynamic nature of interactions between solar radiation, climate, atmosphere, the earth's surface and the PV technology used. Several fast web applications (written in C language) enable an easy estimation of the PV electricity generation potential for selected specific locations in Europe.

In [3] PVGIS interactive on-line calculator is used to calculate the yearly total of solar irradiation and PV power estimation for the territory of Serbia for PV modules placed in horizontal (e.g. roofs), vertical (e.g. south-facing buildings facades) and optimally-inclined planes (for maximizing solar energy harvesting in grid-connected PV power plants). Some results of analysis performed in [3] are presented here.

A typical PVGIS value for the performance ratio (PV system losses) of PV systems with modules from mono- or polycrystalline silicon [9,10] is taken to be 0.75, [1,6].

In this paper the version PVGIS-3 is used. The PVGIS-3 data set is based on measurements made on the ground in the period 1981-1990 which are then interpolated between points to get radiation values at any point. A new version PVGIS-CMSAF has been recently introduced which uses the new databases for the solar radiation data provided by the Climate Monitoring Satellite Application Facility (CMSAF) from the period 1998-2010, [1]. According to the possible wrong terrestrial measurements and to the fact that the amount of solar radiation has increased over Europe in the last 30 years, calculations with new PVGIS-CMSAF give higher values than with the older PVGIS-3. For the territory of Serbia PVGIS-CMSAF gives up to 5% higher values for the solar irradiation data.

III. PVGIS Data and Solar Maps for the Territory of Serbia

PVGIS interactive on-line calculator is used to calculate the yearly total of solar irradiation and PV power estimation for the territory of Serbia for PV modules placed in optimally inclined and oriented plane (giving the"optimal" solar irradiation for maximizing solar energy harvesting) and two-axis sun-tracking PV systems.

Fig. 1 shows PVGIS calculated data for average yearly values of total global irradiation in optimally inclined planes in kWh/m^2 for the territory of Serbia. It is clear from Fig. 1 that average solar irradiation is not dependent on geographical latitude only. There are regional differences in global irradiation due to terrain features and climatic conditions.

The properties of the Serbian territory from the point of view of PV utilization are [3,11]:

- Serbia belongs to the South-Eastern European region with a diverse landscape and mainly continental climate with hot and often rainless summers, but windy and snowy winters.
- Yearly sum of "optimal" total solar irradiation for the territory of Serbia varies from 1380 kWh/m^2 in the north up to 1720 kWh/m^2 in the south.
- Serbia has some territorial units with favourable climatic conditions but mainly good conditions for solar PV electricity production.
- Serbia can be divided in the three main regions in respect to the level of the yearly sum of "optimal" total solar irradiation: 1) *Northern and easternmost regions* (about 25% of Serbia) with the "optimal" total solar irradiation less than 1500 kWh/m^2; 2) *Central region*(about 60% of Serbia) with the "optimal" total solar irradiation within a range of 1500 kWh/m^2 to 1600 kWh/m^2; and 3) *South-eastern region*(about 15% of Serbia) where the "optimal" total solar irradiation exceeds 1600 kWh/m^2.
- The seasonal variation of PV electricity yields is significant for the territory of Serbia, with November, December and January being the worst and July and August the best months for PV electicity harvesting (about 3 times better sunny conditions during summer months).

A summary of conclusions on the basis of PVGIS data analysis for the territory of Serbia is:

- The optimum inclination angle of south-facing PV modulesis mainly 33^0 to 35^0.
- Significant daily variations in solar radiation and seasonal horizon-heights of the sun's position in the skysuggest that the sun-tracking PV systems could be considered as the favourable solution in Serbia for PV investments in the future.
- For a performance ratio of the PV system at 0.75 and a PV module conversion efficiency of 15%, the required area for 1kW of installed PV power is about 9 m^2 (around 3.0 m x 3.0 m). It means that in real circumstances in Serbia, market available standard PV equipment assembled from fixed optimally inclined and oriented PV modules and mounted on area of about 9 m^2 (installed PV power of 1kW) harvests on average between 1150 kWh and 1200 kWh of electricity annually, while the daily average value is about 3.2 kWh.
- Grid-connected PV solar systems with optimally inclined and oriented PV modules and with 1kW of installed PV power could yield on average: 1550 kWh x 0.75 x 23 c€/kWh $\simeq$ 270 € annually. For a guaranteed period of 12 years it is nearly 3,300 €.Feed-in tariff (FIT) rate for PV produced energy from 23 c€/kWh is adopted in calculation.
- A two-axis sun-tracking PV solar system with installed 1kW could yield 30% to 35% more money, on average $\simeq$ 360 € annually or for a guaranteed period of 12 years about 4,300 €.

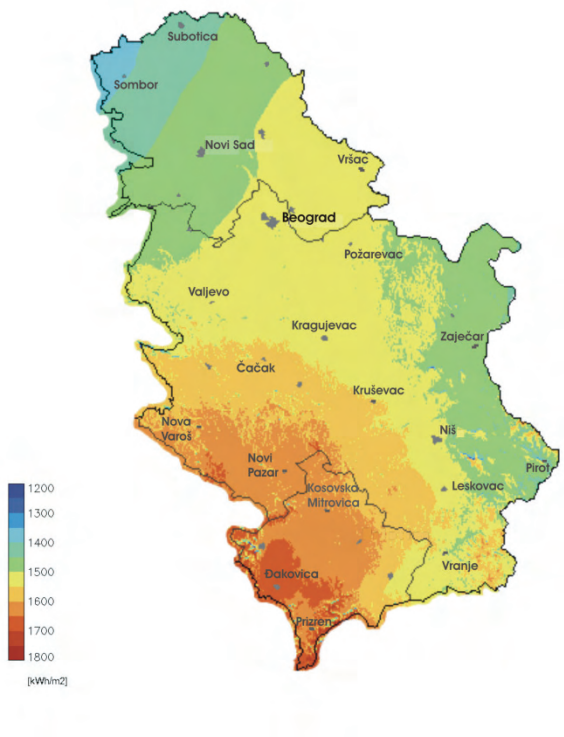


Fig. 1. Yearly sum of total solar irradiation incident on optimally inclined south-oriented PV modules in kWh/m^2 for the territory of Serbia. Adapted for Serbia from PVGIS © European Communities, 2001-2008,http://re.ec.europa.eu/pvgis/, [1,3].

Table 1 gives a daily average of possible PV electricity production in Serbia during a year for fixed optimally inclined and two-axis sun-tracking PV systems [1,3,11]. It is obvious that the central and the south-western parts of the Serbian territory have slight advantage for PV energy harvesting.

TABLE I
AVERAGE VALUES OF POSSIBLE PV ELECTRICITY PRODUCTION FOR THE SERBIAN TERRITORY

Relative location in Serbia	Yearly average values of PV electricity production per day, [kWh/day]	
	fixed optimally inclined PV systems	two-axis sun-tracking PV systems
Farthest North	3.0	3.9
North	3.1	4.1
Farthest South-East	3.1	4.0
South-East	3.2	4.2
Central	3.2	4.3
South-West	3.4	4.5
Farthest South-West	3.5	4.7

IV. PVGIS CALCULATED DATA FOR THE CITY OF NIŠ (SOUTH-EAST REGION OF SERBIA)

For the host city of the ICEST-2011 Conference, the city of Niš,which is located in the south-east region of Serbia (latitude: 43°18'59" North, longitude: 21°53'59" East), some illustrative PVGIS calculations are given in Figs. 2 to 6.The optimum inclination angle of south-facing PV modulesis calculated to be 32^0 with relative orientation -3^0.

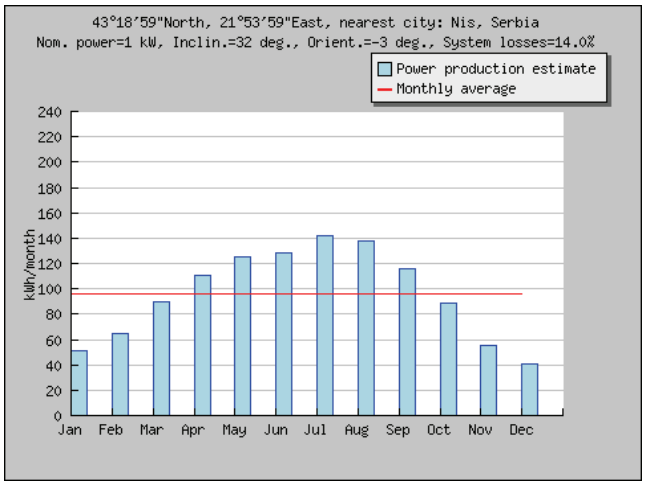


Fig. 2. PVGIS calculation of the estimated amount of electric power which can be expected each month from a PV system with nominal installed power of 1 kW, with stationary PV modules placed in an optimally inclined and oriented plane, placed in the city of Niš, [1].

There are many other aspects which have to be considered in a PV assessment analysis of the specific location planned for setting up the PV power system. Some of them, like surface inclination and orientation, shadowing effect of the local terrainfeatures (e.g. mountains), etc., are incorporated in PVGIS software already. However, possible shadowing effect of the local urban features (e.g. buildings), winter and snow lasting conditions, local pollution and dust conditions (e.g. the vicinity of highways and factories as possible dust sources), etc., have to be carefully analyzed for particular location.

It is important to highlight that besides solar irradiation parameters calculated for the specific location, there are other parameters which have to be taken into consideration within a feasibility study of setting up a grid-connected solar PV system in Serbia [3], such as: technical parameters of the planned solar PV modules and panels, the market prices of solar PV equipment, the cost of design, mounting and maintenance of a solar PV system, and eventually the Government driven RES policy parameters, [3,12,13].

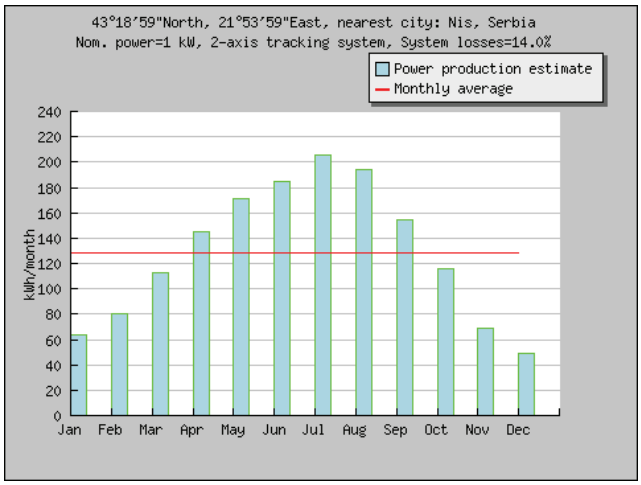


Fig. 3. PVGIS calculation of the estimated amount of electric power which can be expected each month from a PV system with nominal installed power of 1 kW, with a two-axis sun-tracking PV system, placed in the city of Niš, [1].

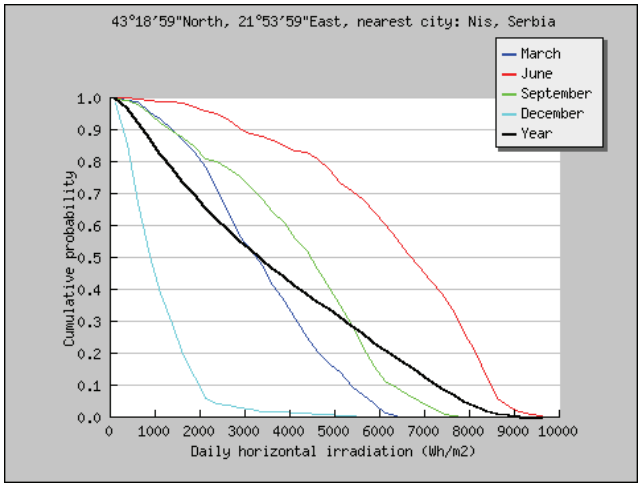


Fig. 4. Probability distribution of daily horizontal solar irradiation,in Wh/m^2, for the city of Niš, [1].

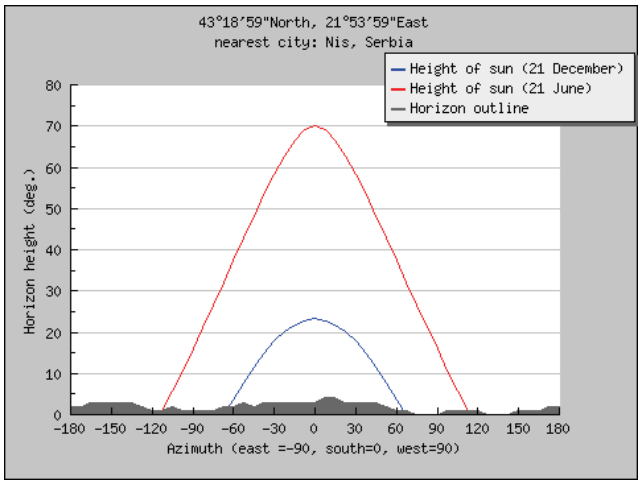


Fig. 5. Seasonal positions of the sun in the sky above the city of Niš during 21st December (the "worst" month and day for the PV electricity production) and 21st June (the "best" month and day for the PV electricity production), [1].

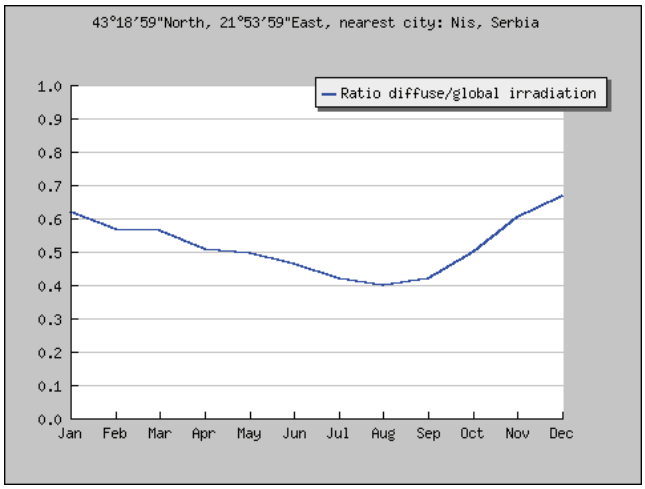


Fig. 6. The ratio between the diffuse component and global solar irradiation for the location of the city of Niš,[1].

V. Conclusion

PVGIS based maps and data for average annual solar irradiation for the specific location within the territory of Serbia are presented and analysed in order to achieve useful details and assessment for the potentials for solar PVs utilization in Serbia. The PVGIS figures and tables (some examples are presented in this paper) can serve as guidelines for the basic necessary data for solar radiation and design of PV grid-connected systems in the Republic of Serbia.

The results obtained from the PVGIS database and on-line calculator can differ from those provided by other providers of solar PV data. For detailed and much more trusted data one has to compare PVGIS calculations with data offered by other similar services. It is obvious that every serious investor in PV power engineering in Serbia would take into consideration data from several databases from different specialized companies.

The overall conclusion is that the Republic of Serbia has the favourable solar irradiation and climatic conditions for solar PV electricity generation; hence the utilization of solar PV electricity generation in Serbia has to be almost certainly expected in the future.

Acknowledgement

This paper was based on research conducted within the Project (Project code: TR33046)funded bythe Ministry of Science and Technological Development of the Republic of Serbia (MSTDRS).

References

[1] PVGIS ©European Communities, 2001-2008 [internet], available at: <http://ie.jrc.ec.europa.eu>.

[2] M. Šúri, T.A. Huld, E.D. Dunlop and H.A. Ossenbrink, "Potential of solar electricity generation in the European Union member states and candidate countries", Solar Energy, vol. 81, pp. 1295–1305, 2007.

[3] D.Z. Djurdjevic, "Perspectives and assessments of solar PV power engineering in the Republic of Serbia", Renewable and Sustainable Energy Reviews, vol. 15, pp. 2431-2446, 2011.

[4] M. Golušin, Z. Tešić and A. Ostojić,"The analysis of the renewable energy production sector in Serbia", Renewable and Sustainable Energy Reviews, vol. 14, pp. 1477-1483, 2010.

[5] European Commission, Joint Research Centre, Institute for Energy, *Research, Solar Cell Production and Market Implementation ofPhotovoltaics* [internet]. PV Status Report, August 2010, available at: <http://ie.jrc.ec.europa.eu>.

[6] M. Šúri, T. Huld, T. Cebecauer and E.D. Dunlop, "Geographic Aspects of Photovoltaicsin Europe: Contribution of the PVGIS Web Site", IEEE Journal of Selected Topics in AppliedEarth Observations and Remote Sensing, no. 1, pp. 34-41, 2008.

[7] M. Šúri, E.D. Dunlop and T.A. Huld, "PV-GIS: A web based solarradiation database for the calculation of PV potential in Europe", International Journal of Sustainable Energy, vol. 24, no. 2, pp. 55–67, 2005.

[8] T. Huld, M. Šúri, E.D. Dunlop, "Geographical variation of the conversion efficiency of crystalline silicon photovoltaic modules in Europe",Progress inPhotovoltaics: Research and Applications, vol. 16, pp. 585-607, 2008.

[9] A. Goetzberger, V.U. Hoffmann, *Photovoltaic Solar Energy Generation*, Berlin Heidelberg, Germany, Springer-Verlag, 2005.

[10] D.Z. Djurdjevic andM. Jevtic,"Issues and prospects of solar power engineering", Energija, ekonomija, ekologija, vol. 11, no. 5, pp. 59-66, 2009.

[11] D.Z. Djurdjevic, "Status of solar photovoltaic power engineering in the Republic of Serbia", (accepted for Conference), ECOS-2011 Conference, July 2011, Novi Sad, Serbia.

[12] S. Lüthi, "Effective deployment of photovoltaics in the Mediterraneancountries: Balancing policy risk and return", Solar Energy, vol. 84, pp. 1059–1071, 2010.

[13] D.Z. Djurdjevic, "Assessment of the Present Serbian RES Policy", Energy Sources, Part B: Economics, Planning, and Policy, (accepted for publication), 2011.

Session PO6:

PO6 - COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES

Issues of Migration from IPv4 to IPv6

Gjorgi Mikarovski[1], Aleksandar Kotevski[2], Ilija Jolevski[3]

Abstract – **Internet Protocol (IP) has become the standard communication tool, ever since its introduction in 1970's. This resulted inmajortechnological developments of the internetworking devices. IPv6 is a new version of Internet Protocol, developed by the Internet Engineering Task force (IETF) as an evolutionary step from IPv4. The most important purpose of IPv6 is to allow a greater number of connected hosts than allowed by IPv4. While IPv4 allows 32 bits for an Internet Protocol address, and can therefore support 232 (4,294,967,296) addresses, IPv6 uses 128-bit addresses, so the new address space supports 2128 (approximately 340 undecillion or 3.4×1038) addresses. In addition, IPv6 reduces packet processing overhead and increases scalability. Together, these improvements allow a greater exchange of data traffic.**

Keywords – **communication, traffic, network, protocol, migrations, dual-stack, tunnelling, IP/ICMP, NAT-PT, 6 to 4, 4 over 6.**

I. INTRODUCTION

IETF designed IPv6 to allow interoperability between devices that use the IPv4 stack and devices that use the IPv6 stack. It is usually possible to install IPv6 on internet devices without losing IPv4 capability, so organizations can perform incremental upgrades from IPv4 to IPv6 while avoiding disruption of service during the transition.

Since IPv4 and IPv6 packets are not directly compatible, therefore a technique known as translators are used that translate the IPv4 packets into IPv6 and vice versa. But translators tend to slow the network. Translation between IPv4 and IPv6 can take place at three levels i.e. IP level, transport level and the application level.

II. OVERVIEW OF IPV4 AND IPV6

The IPv4 addresses are represented with dot decimal notation. The addresses are divided into two parts, the network ID and the host ID. The network ID is of 8 bits, therefore it can addresses up to 256 hosts. To overcome this limit, five different classes namely A, B, C, D & E were created and were named as classful addressing. Classless Inter Domain Routing (CIDR) replaced the classful addressing, which allowed the re-division of class A, B & C networks. Class A, B & Care reserved for use by private networks.

[1]Gjorgi Mikarovski is with the Faculty of Technical Science, Ivo Lola Ribar bb 7000 Bitola Macedonia, E-mail: gjorgi.mikarovski@tfb.uklo.edu.mk

[2]Aleksandar Kotevski is with the Faculty of Technical Science, Ivo Lola Ribar bb 7000 Bitola Macedonia, E-mail: aleksandar.kotevski@uklo.edu.mk

[3]Ilija Jolevski is with the Faculty of Technical Science, Ivo Lola Ribar bb 7000 Bitola Macedonia, E-mail: ilija.jolevski@uklo.edu.mk

On the contrary, IPv6 is designed to overcome the deficiencies of IPv4 by expanding the available IP's pool and by incorporating features such as IPsec, quality of service (QOS), efficient routing and mobile communications. These new features can be used to develop new E-Commerce businesses, increase broadband penetration and to enhance the mobile communication.

The transition from IPv4 to IPv6 will take place in three stages i.e. substitution, diffusion and complete transformation. In substitution IPv6 will substitute the IPv4. In this phase organizations implementing IPv6 in their infrastructure will operate in a dual stack environment. In diffusion, new applications will be developed using IPv6 that will be more innovative and economical. In diffusion,IPv4 will be obsolete and new hardware will run entirely on IPv6. The complete transition from IPv4 to IPv6 is expected to take many years.

IPv6 deployment issues

- IPv4 and IPv6 do not interoperate
  - IPv4 applications do not work with IPv6
  - IPv4 nodes cannot communicate with IPv6 nodes
- The applications have to be modified
  - a lot of work still has to be done.
- It is likely that IPv4 and IPv6 will coexistfor a long period of time
  - How to enable communications among IPv6islands isolated in the IPv4 world?
  - How to enable communications between the existing IPv4 world and the new IPv6 world?

III. MIGRATION MECHANISMS

In order to replace the IPv4 protocol, a few years ago the IPv6 protocol has been intro-duced to the internet community. However, IPv4 is still dominating most of the internet infrastructure. We expect that the migration between both protocols will not take place in a rapid way, especially not in the U.S. and Europe. Due to the serious lack of IP addresses in Asia there is more need for IPv6, that's why the migration process will be faster in that region.

The migration mechanisms described in this report are based on the transition. These are: Dual IP layer operation, Tunneling, IPv4-compatible IPv6 addresses, 6to4, 6over4.

Furthermore, there exist other transition mechanisms. Network manufacturers have deployed various proprietary mechanisms, like the Intra Site Automatic Tunnel Addressing Protocol (ISATAP).

Dual IP Layer operation - the most used migration approach nowadays is the dual IP layer operation, also called the dual stack method. A host with a dual stack can interoperate with both IPv4 and IPv6 nodes using IPv4 or IPv6 packets. Dual

stack has the possibility to disable one of the IP stacks for operational reasons. A node configured with a dual stack can make decisions on TCP connections based on the IP header of the TCP packet:

- The IPv4 protocol stack will be used if the destination address used by the application is an IPv4 address;
- The IPv6 protocol stack will be used if the destination address used by the applications an IPv6 address;
- Encapsulation of an IPv6 packet inside an IPv4 packet will occur if the destination address used by the application is an IPv6 address with embedded IPv4 address.

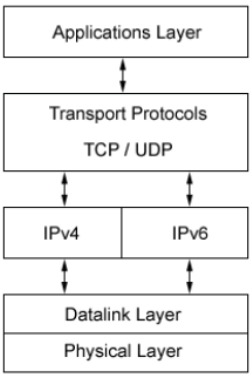


Fig 1. The dual stack approach

Dual stack makes it possible to continue provide access to IPv4 resources, while adding IPv6 functionality. The IP address acquisition in the dual stack nodes occurs by using IPv4 mechanisms like DHCP or IPv6 mechanisms, as for instance the stateless address auto-configuration. Hosts and routers that support dual stack may use tunneling mechanisms to route IPv6 tra c over IPv4 networks.

Issues with simple dual-stack

- it does not reduce the demand for globally routable IPv4 addresses
- it increases network complexity due to the need for a double (IPv4/IPv6) routing infrastructure

Other dual-stack approaches

DSTM (Dual Stack Transition Mechanism)

- deployment of dual-stack nodes with dynamically assigned IPv4 addresses
- IPv4 over IPv6 tunneling to avoid the need for a dual-stack routing infrastructure

Application Level Gateways (ALG)

- the client is IPv6-only and the communication with the IPv4 world goes through a dual-stack proxy

A dual-stack alternative is NAT-PT (NAT - Protocol Translator)

Tunneling - because IPv6 will be developed over the IPv4 infrastructure, tunneling provides a way to use the existing routing infrastructure to carry IPv6 tra c. Tunneling IPv6 packets over IPv4 infrastructure is done by encapsulating IPv6 packets inside IPv4 packets as shown in figure 2. The IPv6 header contains the address of the final destination and the IPv4 header contains the address of the tunnel endpoint. The encapsulation of IPv6 packets within IPv4 packets can occur in the following ways:

Router-to-Router: An IPv4 infrastructure will tunnel IPv6 traffc between directly linked IPv6/IPv4 routers.

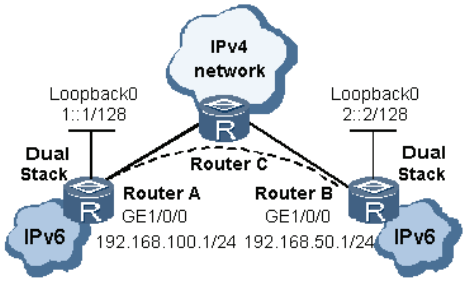


Fig 2. Router to router

Host to Router: IPv6/IPv4 hosts will tunnel IPv6 tra c to an IPv6/IPv4 router via an IPv4 infrastructure.

Fig 3. Host to router

Tunneling of IPv6 over IPv4 - these are amongst the most basic techniques that can be deployed in order to allow operation of two or more protocols on the network. This technique involves the encapsulation of IPv6 packets with in IPv4 header. A tunnel is a link between two IPv4 end points that must be configured by specifyingthe IPv6 destinations for which the packets are to be encapsulated, and the remote IPv4 end point to which they must be sent.

Issues with simple tunneling

- configured tunneling requires heavy manual configuration and therefore does not scale well
- automatic tunneling is not the solution because it can be used only between individual hosts

IPv4-compatible IPv6 - are also known as 6over4. In this mechanism, ex addresses are used to create IPv4-compatible IPv6 addresses. These addresses are by a 96bit zeros prefix followed by the 32bits IPv4 address. In this approach, IPv4 addresses become a virtual link-layer address by using I cast group. Neighbor Discovery takes place by mapping IPv6 multicast address multicast addresses. The router must be configured as 6over4 in order to make IPv routing possible. The hard requirements and poor scalability characterize this implementation.

6 to 4 - It is a method of constructing the IPv6 address directly from the IPv4 address. This mechanism enables sites to communicate over the IPv4 Internet without using explicit tunnels while still communicating with IPv6 relay routers. 6 to 4 treats IPv4 Internet as a unicast point-to-point link layer and specify an encapsulation mechanism for transmitting IPv6 packets using the prefix. This mechanism is implemented entirely in border routers and is thus becoming a

standard feature of router software. Below is thediagram given for 6 to 4

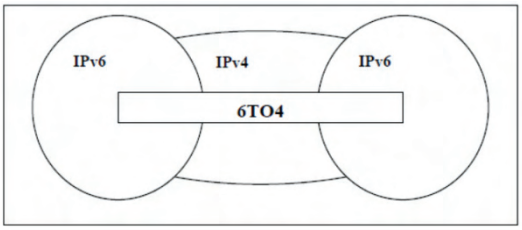


Fig 4. 6to4

6 over 4 - This mechanism facilitates IPv6 connectivitywith in a site that lacks any IPv6 infrastructure. It describes the frame format for IPv6 packets as well as method of forming IPv6 link local addresses over IPv4 multicast domains. It allows IPv6 hosts to become functional if at least one IPv6 router is located in the same domain. This technique is helpful for sites that still have no IPv6 networks but wish to deploy it or test it. 6 over 4 have received very limited support from the major vendors; only Microsoft and Nokia have implementations of 6 over 4.

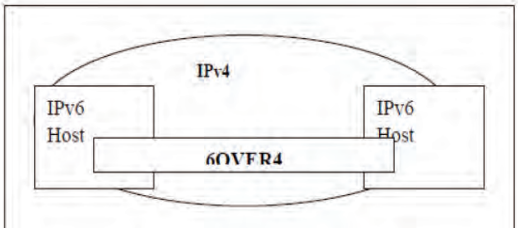


Fig 5. 6over4

IV. TRANSLATION MECHANISMS

Once an IPv4 network connects to an IPv6 network, there is still the need to communicate between the IPv6-only nodes with the IPv4 nodes that are not yet IPv6 enabled. There are various translation mechanisms developed that enable cross-protocol communication. The difference between all the translation mechanisms is where the translation actually takes place in the various layers of the TCP (or UDP)/IP reference model.The different translation mechanisms that we describe in this report, and also have been proposed as a RFC to the IETF are categorized as follows.

*Network layer:*Stateless IP/ICMP Translation Algorithm (SIIT), NAT-PT
*Transport layer:*Transport Relay Translator (TRT)
*Application layer:*Bump in the API (BIA)

Stateless IP/ICMP Translation Algorithm - the stateless IP/ICMP translation algorithm (SIIT) can be used in the transition between IPv4 and IPv6. There are two types of translation:

IPv4 to IPv6 translation
When an IPv4-only node sends a packet to another node, whose destination is an IPv6 ad-dress in another network, the SIIT algorithm detects this. The SIIT algorithm translates the IPv4 header of the packet to an IPv6 header, and discards the IPv4 header. Subsequently the packet is forwarded to the destination IPv6 address. The noticeable dierences between the IPv4 and IPv6 header are in the following fields:

Source field The IPv4 address source address of the IPv4 host is converted to an IPv6 address where the low-order 32 bits is the IPv4 source address. The high-order 96 bits is the IPv4-mapped prefix (::ffff:0:0/96).

*Destination field*the destination address in the IPv4 header is converted to an IPv6 address where the low-order 32 bits is the IPv4 destination address. The high-order 96 bits is the IPv4-translated prefix (::ffff:0:0:0/96).

IPv6 to IPv4 translation
In this translation type, the IPv6 header is translated to IPv4, and the packet is forwarded through IPv4. The issue here is the MTU handling of the network link. IPv6 ensures that the minimal link MTU of an IPv6 packet is 1280 bytes, while IPv4 uses packets with a minimal size of 68 bytes. When a translated IPv4 packet with a bigger MTU than 68 bytes arrives at an IPv4 router on a network with a minimal MTU of 68 bytes, the router fragments the translated IPv4 packet and forwards it over IPv4.

Network Address Translation and Protocol Translation (NAT-PT) is a transition mechanism to provide transparent routing between IPv4 and IPv6 end nodes. NAT-PT uses a combination of network address and protocol translation. The difference between NAT-PT and NAT in IPv4 is that translation does not happen between private and global addresses, but between IPv4 and IPv6 addresses.

Traditional NAT-PT With traditional NAT-PT, a host can only initiate a TCP (as well UDP and ICMP) session from the IPv6 network to connect an IPv4 host outside the network. These kinds of sessions are unidirectional. There are two variations of traditional NAT-PT, namely Basic NAT-PT and NAPT-PT.

Basic NAT-PTwith Basic NAT-PT a /96 prefix is reserved that can be routed in the IPv6 domain. All IPv6 traffic with the /96 destination prefix will be routed to the NAT-PT border router and translated to IPv4.

NAPT-PT stands for Network Address Port Translation and Protocol Translation and ex-tends the notion of translation one step further. NAPT-PT allows multiple IPv6 hosts to share a single IPv4 address to communicate transparently with the hosts outside the NAPT-PT router.NAPT-PT can be used in addition to basic NAT-PT to extend the pool of external map-ping addresses in conjunction with port translation. By assigning the TCP sessions from the initiating IPv6 hosts to ports in conjunction with IPv4 addresses, a maximum of 63000 TCP and 63000 UDP sessions per IPv4 are available to assign. There are 254 IPv4 addresses available, which are enough possibilities for address mapping.

Bidirectional NAT-PT combined with a DNS Application Layer Gateway provides bi-directional connectivity between an IPv6 domain and an IPv4 domain outside the NAT-PT border router. This mechanism employs the possibility to initiate TCP sessions from as well IPv6-only hosts as from an

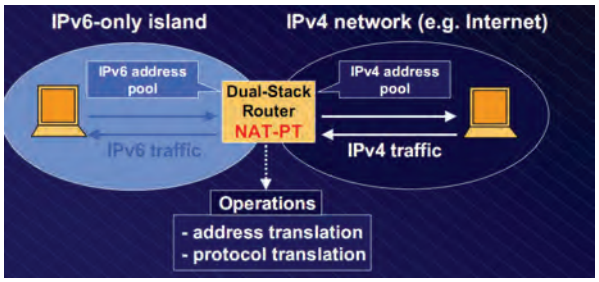


Fig 6. Dual-stack alternative

Issues with NAT-PT

- More or less the same as IPv4 NATs
- some applications may not work (need for ALGs)
- IPsec, Mobile IP, etc. fail (no e2e transparency)
- performance degradation
- single point of failure
- need for strict coordination with DNS for automatic translation state initialization

But unlike IPv4 NATs, NAT-PTs are just a temporary solution and after the transition has been completed the NAT-PT box may be removed

Transport Relay Translator- the transport relay translator (TRT) mechanism functions on the transport layer of the TCP/IP reference model. It allows IPv6-only nodes to communicate with IPv4-only nodes by translating TCP-over-IPv6 to TCP-over-IPv4, and the other way round. The TRT mechanism works the same for UDP track. The TRT system can be located on a dual stack host or router. When an initiating IPv6 host wants to communicate with an IPv4 host it needs an IPv6 destination address. All TCP traffic from the IPv6 host goes through the TRT system, which functions as a traffic relay server. When the TRT system receives incoming TCP traffic from an IPv6 source host (X6) to an IPv4 destination host (Y4), it makes an IPv6 connection with the initiating IPv6 host. The TRT system is configured with a dummy IPv6 prefix like C6::Y4/64, where Y4 is the destination IPv4 address. The initiating IPv6 host has the ability to connect to the IPv4 host through the IPv6 address C6::Y4.After that, the relay server makes a connection w the IPv4 host and forwards the TCP traffic to the IPv4 host. When the relay server receives traffic from the IPv4 host, it establish in the same way a virtual connection with address C6::/64, and forwards the traffic to the IPv6 host.

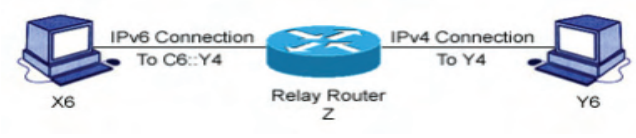


Fig 7. TRT Connection

Bump in the API- an alternative for the Bump in the Stack mechanism is the Bump in the API (BIA) mechanism. The main difference between the two is that BIA does not do any header translation, but performs its translation between the IPv4 API and IPv6 API on the application layer. The BIA API consists of three components, which are inserted between the socket API module and the TCP/IP module in a dual stack host. These components are: name resolver, address mapper, function mapper

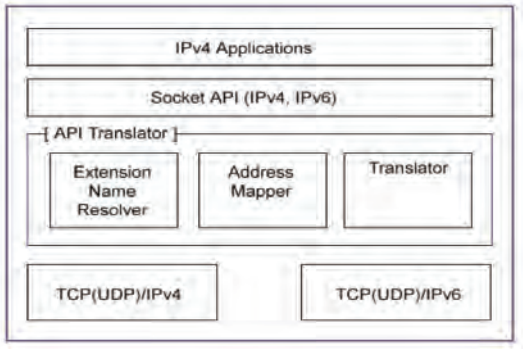


Fig 8. Architecture of the dual stack host using BIA

The name resolver and address mapper are the same as used in BIS. The main di erence between BIS en BIA happens in the translation function. The function mapper in BIA captures IPv4 socket API function calls and converts them to new IPv6 socket API function calls. BIA translates in the same way from IPv6 to IPv4.

V. Conclusion

We believe that running both protocol versions at the same time based on the dual stack approach while deploying IPv6 in phases, is the key to a successful IPv6 migration. IPv6 connectivity in a dual stack network environment means that current net-work security measures are not valid anymore for IPv6. Concerning viruses and worms, we don't expect these will be representing a potential IPv6 vulnerability since Antivirus software works independent of IPv4 or IPv6 unless the used software has been updated with IPv4 functionality.

VI. References

[1] Nordmark, "Stateless IP/ICMP Translation Algorithm (SIIT)", RFC 2765,
[2] February 2000.
[3] Tsirtsis and Srisuresh, "Network Address Translation - Protocol
[4] Translation (NAT-PT)", RFC 2766, February 2000.
[5] Tsuchiya, Higuchi and Atarashi, "Dual Stack hosts"
[6] Bump-in-the-Stack" technique (BIS)", RFC 2767, February 2000.
[7] Hagino and Yamamoto, "An IPv6-to-IPv4 Transport Relay Translator",
[8] IETF ipng working group http://www.ietf.org/html.charters/ipngwg-charter.html
[9] IETF ngtrans working group http://www.ietf.org/html.charters/ngtrans-charter.htm

Change of the National Top-Level Domain and its Influence to Some Spam Detection Characteristics

Slobodan Mitrović[1], Slaviša Aćimović[2], Slađana Janković[3], Norbert Pavlović[4], Sanjin Milinković[5]

Abstract – **In this paper some experiences related to spam detection and change of national top-level domain in Republic of Serbia are presented. Email and spam detection statistics as well as efficiency of several spam detection tests before and after this change are considered. Presented statistics are based, as example, on the MTA log and spam-filter data of the Faculty of Transport and Traffic Engineering Computer center in Belgrade.**

Keywords – **email, spam detection, national top-level domain, statistics**

I. INTRODUCTION

One of the most important Internet services is electronic mail (email) service that uses Simple Mail Transfer Protocol (SMTP) for transfer of messages. This type of service has some security disadvantages which make it suitable for spam delivering. During last decade spam is identified as one of major threats for business or other kinds of activities, especially in systems with large number of employees, such as national corporations (i.e. Railways) or Internet corporations that host webmail services (i.e. Google, Yahoo, Microsoft (Live mail), etc.). This was the primary motive for many institutions to build email filtering mechanisms and monitoring service that collects data about email traffic.

Some major changes, such as change of national top-level domain (nTLD) have crucial influence to many features of the Internet traffic of the corresponding country. This reflects to email traffic structure, and also to efficiency of spam detection filters. In Section II the concept of used email filter and email monitoring is briefly presented, as well as collected statistics before and after nTLD change. Section III presents some of the most successful anti-spam filtering tests before this change and their efficiency afterwards. This will be followed by corresponding conclusions presented in Section IV.

[1]Slobodan Mitrović is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: s.jankovic@sf.bg.ac.rs.

[2]Slaviša Aćimović is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: slavisa@sf.bg.ac.rs.

[3]Slađana Janković is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: s.jankovic@sf.bg.ac.rs.

[4]Norbert Pavlović is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: n.pavlovic@sf.bg.ac.rs.

[5]Sanjin Milinković is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: s.milinkovic@sf.bg.ac.rs..

II. EMAIL STATISTICS BEFORE AND AFTER NTLD CHANGE

Republic od Serbia as successor of former Yugoslavia used national TLD "***.YU***" until 2009 when IANA assigned new TLD "***.RS***", with transition period that ended on March, 31st of 2010.

During last 15 years, many of users' email addresses were ended in spammers' databases because of various security issues, which leaded to increase of unwanted email traffic amount. This was primary reason for our Computer center to engage security measures, in order to make email service more usable and friendly for end users. The concept of security measures is based following structure: our email system consists of *Postfix* MTA [1] integrated with *SpamAssassin* anti-spam filter [2] and *Syslog-NG* logging engine [3] that collects log data for all activities related to email traffic. Further, collected log data is statistically processed by Mailgraph [4], which uses RRDtools [5]. This is internally called *PSRM integration* [6], which has many similarities with other standardized *Postfix* integrations.
This means that incoming messages are tested against *Postfix SMTP* rejection rules that consist of HELO, sender and recipient restrictions. The goal of this standard procedure is rejecting sessions from any system that fails to identify itself. Further, those messages that passed this primary filter are tested with anti-virus and anti-spam filters, respectively. Infected messages are passing through cleaning procedure. Message that is identified as spam, if not rejected by primary filter, ends in mailbox called *spamcontainer*. Content of the message is modified with brief description how the message has been qualified as spam. Original content is placed as attachment.

In this way, increase of local email traffic and also unnecessary increase of amount of local users' mailbox is avoided. In case of *false-positive* message, end user can look for it by searching the content of *spamcontainer* mailbox. Further, this mailbox is used for collecting information related to spam characteristics.

Statistics that are presented in this paper are collected in the last 12 months before change of national TLD and also in the first 12 months after this change. Those statistics that are related to amount of detected spam and infected messages are given in Table 1.

TABLE I
EMAIL STATISTICS BEFORE AND AFTER THE CHANGE OF NATIONAL TLD

Message type / period	March 2009 - March 2010	April 2010 - March 2011
Total	2.692.324	291.899
Rejected by primary filter	2.032.849	159.680
Messages with virus content	6.624	482
Identified spam	471.515	8.695
Regular messages	181.336	123.042

Statistics presented in Table 1. show significant decrease of amount of incoming messages after change of national TLD, which is graphically presented in Fig 1. It could be also noticed that more than 93% of all incoming messages have been treated as unwanted before the change of national TLD. This ratio decreases to almost 58% in the first year after the change of national TLD.

Some additional issues related to presented statistics must be taken into consideration:

- Various end user surveys show presence about 4% of *false negatives* and 0.5% of *false positives* in corresponding mailboxes, in period before this change.
- There have been increased amount of rejected regular emails in first two weeks after the change of national TLD, because recipients' email addresses have not been updated by many uninformed senders that use domestic Internet Service providers (ISP). This is important issue, because many domestic ISPs didn't remove all DNS records related to old TLD. This is also shown in Fig. 1.

Furthermore, it have been noticed increased ratio of *false positives* after change national TLD, because *SpamAssassin* filter trigger have been tuned with *combined score* that is not appropriate to be used in new situation. In order to improve results, some additional assessments had to be done.

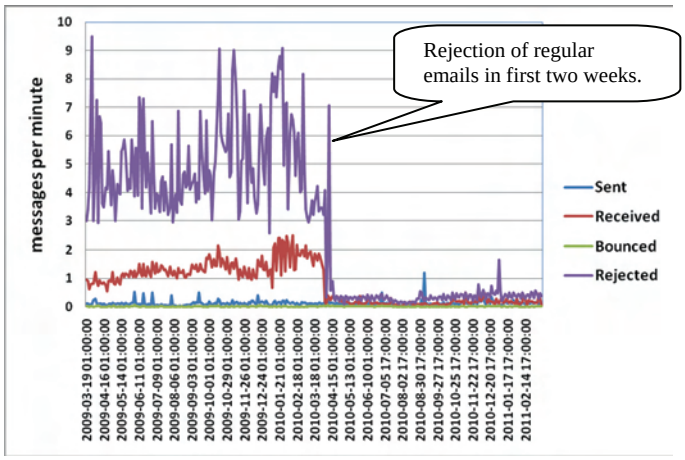


Fig. 1. Email statistcs before and after change of national top-level domain

III. SPAMASSASSIN FILTERING BEFORE AND AFTER NTLD CHANGE

SpamAssassin is used in our *PSRM integration* as secondary mail filter, for the purpose of spam determination. This flexible and powerful set of Perl programs, unlike older spam filtering approaches, uses the *combined score* from multiple types of checks to determine if a given message is spam [2]. Its primary features are [2]:

- Header tests
- Body phrase tests
- Bayesian filtering
- Automatic address whitelist/blacklist (AutoWhitelist)
- Manual address whitelist/blacklist (ManualWhitelist)
- Collaborative spam identification databases (DCC, Pyzor, Razor2)
- DNS Blocklists, also known as "RBLs" or "*Realtime Blackhole Lists*"
- Character sets and locales.

The scores are assigned using a neural network trained with error back propagation (*Perceptron*). Both systems attempt to optimize the efficiency of the rules that are run in terms of minimizing the number of *false positives* and *false negatives* [2].

During years before change of national TLD *SpamAssassin* engine that belongs to our *PSRM integration* used *combined score* trigger value of **6.5**, which was determined as optimal for our email service in terms of minimizing the number of *false positives* and *false negatives.* Significant increase of *false positives* in first several weeks after change of nTLD indicated a need for determination of new optimal *combined score* triggers value. Hence, it has been lowered to value of 4, but after only one week it has been additionally lowered to value of 3. Lowering of this value has been done on empiric basis because of emerging need for decreasing *false positives* rate.

Spamcontainer mailbox content was analyzed in order to identify those *Spamassassin* tests that have crucial weight in scoring process of spam determination, before and after nTLD change. In this purpose, *spamcontainer* messages are divided in two groups, called *Before* and *After*. Mailbox group *Before* was analyzed first. This group contains 417515 spam messages that have been identified in last 12 months before nTLD change. In first turn, there have been extracted four groups of spam messages:

- Messages that contain keyword *Viagra*. These messages were automatically redirected to *spamcontainer*, upon header checks.
- Messages that contain keyword *Rolex*. These messages were also automatically redirected to *spamcontainer*, upon header checks.
- Messages that were identified by *SpamAssassin* as type *Nigerian 419*.

- Messsages cought by blacklist which was created within *SpamAssassin* with score value of 100.

In second turn, there have been extracted another nine groups of spam messages, regarding frequency of test with score that was crucial for message to be qualified as spam:

- HELO_DYNAMIC_IPADDR - *Relay HELO'd using suspicious hostname* (score 4.4). An untrusted relay used an IP address as a HELO argument during a SMTP transaction.
- FORGED_MUA_OUTLOOK - (score 4.2). Spammer's client is trying to pretend to be an MS Outlook
- FH_DATE_PAST_20XX - *The date is grossly in the future* (score 3.6)
- BAYES_99 (score 3.5)
- BAYES_95 (score 3.0)
- HTML_IMAGE_ONLY (score 2.6)
- HELO_DYNAMIC_DHCP-*Relay HELO'd using suspicious hostname* (score 2.6). An untrusted relay used a hostname (FQDN) as a HELO argument during a SMTP transaction that appears to suggest a dynamically allocated hostname. [2].
- BAYES_80 (score 2.6)

Spam statistics related to these tests are presented in Fig. 2.

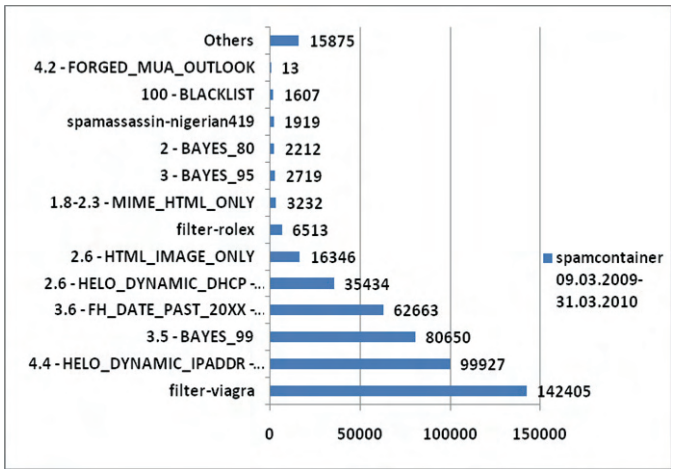


Fig. 2. Spam detection statistics before change of national top-level domain

These nine tests with score that was crucial for message to be qualified as spam, as well as four specific tests were used to analyze mailbox group *After*. This mailbox group contains 8695 spam messages that have been identified in first 12 months after change of national TLD. Statistics related to tests applied on this mailbox group are shown in Fig. 3.

In order to determine types of spam that has not been affected by change of national TLD, presence of determined tests in mailbox groups *Before* and *After* has been compared, as it shown in Fig. 4.

It could be noticed that only one test has almost identical level of presence on both mailboxes. It is test called HELO_DYNAMIC_IPADDR with score value of 4.4. Other

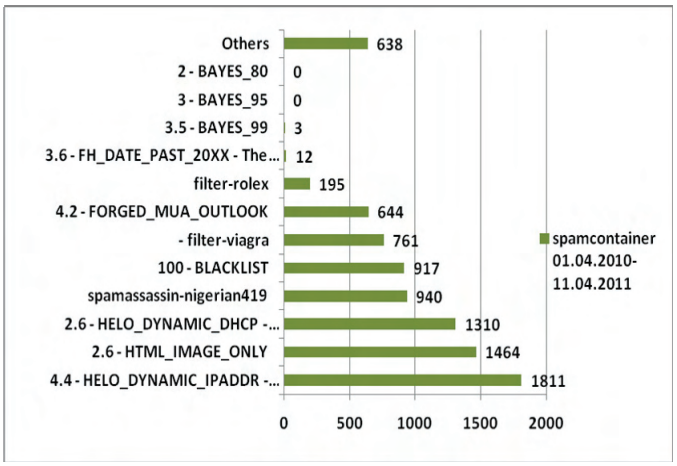


Fig. 3. Spam detection statistics after change of national top-level domain

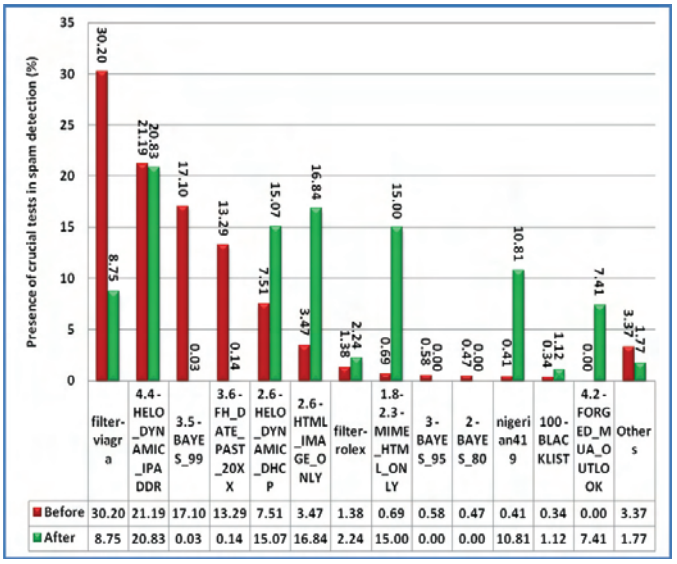


Fig. 4. Presence of determined tests in mailbox groups *Before* and *After*

tests have almost opposite levels of presence, with exception of *Nigerian 419* spam messages which level of presence is seems to be independent of other tests' occurring frequencies. This can be logically explained in terms of nature of this spam type [7].

It could be also noticed that tests related to *Bayes* analysis lost crucial role in spam detection in *After* mailbox group, comparing to their efficiency before change of national TLD. In that period BAYES_99 test has been the most successful test, while BAYES_95 and BAYES_80 just "helped" other tests in reaching of *combined score* trigger value for targeted spam.

Each other *SpamAssassin* test has insignificant level of presence in *Before* mailbox group comparing to levels of those nine tests. Similar situation could be found in mailbox group *After*, with exception of BAYES_95 and BAYES_80 tests which are included in comparative purposes.

IV. Conclusion

In this paper the change of national top-level domain and its influence to email delivery and spam filtering is considered. In the first several weeks after this change, significant amount of regular emails that have been originated from domestic senders has been rejected, because of inconsistent DNS records.

SpamAssassin filter maximized its efficiency in spam determination, using neural network trained with error back propagation as well as optimal *combined score* trigger value regarding amount of spam attacks during last several years before change of national TLD. This efficiency has been disrupted after nTLD change, which is confirmed by several important issues, such as increased number of *false positives*, that involved the need for modification of *combined score* trigger value. Improving efficiency of Bayesian filter in presence of decreased number of spam attacks would be subject of some future analysis.

Change of national TLD involved significant decrease of spam attacks, because spammers email address listings obviously became out-of-date. This is new chance for network administrators to increase security measures in order to protect end users from email addresses harvesting.

Acknowledgement

This work is partially supported by the Ministry of the Science and technological development of the Republic of Serbia under No. 036012.

References

[1] www.postfix.org, (03.04.2011)
[2] http://wiki.apache.org/spamassassin/ (05.04.2011)
[3] http://balabit.com/network-security/syslog-ng, (03.04.2011)
[4] http://mailgraph.schweikert.ch, (03.04.2011)
[5] http://oss.oetiker.ch/rrdtool/doc/rrdtool.en.html, (03.04.2011)
[6] S.Mitrović, V.Radojičić, “A new approach in tracking efficiency of anti-spam filter”, XLV International Scientific Conference on Information, Communication and Energy Systems and Technologies, Proceedings of Papers: ICEST 2010, Ohrid, Makedonija, str. 345-348
[7] Cukier W.,Nesselroth E.,Cody S.,Genre, "Narrative and the "Nigerian Letter" in Electronic Mail", 40th Annual Hawaii International Conference on System Sciences, Proceedings of Papers: HICSS-40, 2007, Waikoloa, HI, pp. 70-70.
[8] Dagon, D.; Lee, W. “Global Internet Monitoring Using Passive DNS”, Cybersecurity Applications & Technology Conference for Homeland Security, Proceedings of Papers: CATCH '09, 2009. pp. 163 – 168

Hybrid ARQ Schemes: Problems and Perspectives

G. Marinova[1], I. Penev[2]

Abstract - **The diversity of Hybrid Automatic Repeat Request (HARQ) schemes is determined by their intensive development. In the paper is made attempt to classify HARQ systems. The problems are introduced on base of their description in a structural and an algorithmic direction.**

Keywords - **ARQ, HARQ I, HARQ II, HARQ III, RB-HARQ**.

I. Introduction

The development of modern communication systems is based on active development of Hybrid Automatic Repeat Request (HARQ) systems. There are two main techniques for error control in data communication systems: Forward-error correction (FEC) and Automatic repeat request (ARQ) [1, 7].

Improving the characteristics of communication channel with FEC systems is achieved by adding additional information to data transmitted. Typically block codes, convolutional codes or turbo codes are used in the FEC system**.**

Automatic repeat request (ARQ) is used in nonstationary channels with high noise level. In ARQ a cyclic redundancy check (CRC) code is applied for error detection. In case of errors the ARQ system sends a request for retransmission of erroneous data [14].

Throughput in ARQ systems is calculated using the follwing expression [1]

$$T_r = E[T] = \frac{1}{1-P_r}, \quad (1)$$

where *Pr* is probability for retransmission of received packets.

Using ARQ schemes cause increasing the effectiveness of throughput of the physical channels [14].

The sum of probabilities is equal to one [16]:

$$P_e^i + P_{ARQ}^i + P_c^i = 1, \quad (2)$$

where i represent the i^{th} retransmission, P_e^i resulting in a block error P_{ARQ}^i is probability of retransmission and P_c^i is correctly decoded the packet.

The aim is to reduce losses from transmission of additional information for the purposes of protocol interaction in the channel level. This approach is applicable only in the cases of duplex channels.

The combination of both FEC and ARQ techniques are known as HARQ and it is applied in modern data communication systems. If there are insignificant errors in the HARQ system works as pure FEC system. Otherwise, it is run the functions of the ARQ system.

[1]G. K. Marinova is with the Department of Computer Science and Engineering, Technical University of Varna, Studentska Street 1, Varna 9010, Bulgaria, E-mail: gin_kaleva@abv.bg

[2]Ivaylo Penev, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: ivailopenev@yahoo.com

II. Analysis of HARQ Shemes

Traditional HARQ schemes are designed to improve performance of data communication systems that are under the influence of interference.

According to the operation mode, there are three basic types of HARQ systems:

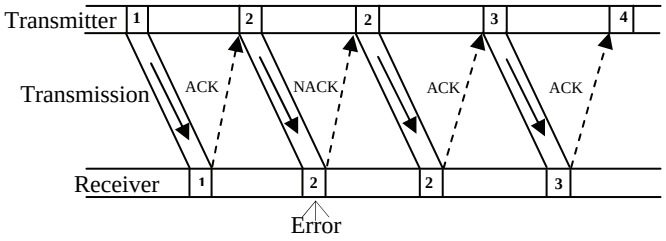


Fig.1"Stop and wait" ARQ scheme.

"Stop and wait" scheme is shown on Fig.1**.** In this system operating mode for the transmission of a block or codeword is required (ACK or NACK) acknowledgement of the previous word. Next codeword is not sent until the previous is not received correctly. Buffering of one packet is required.

The Basic "Stop and wait" scheme disadvantage is the reduced throughput.

“Go-back – N” scheme. The system requires buffering of more than one packet. In the case of NACK acknowledgement all subsequent packets are ignored. The process continues until correcting the missing packet.

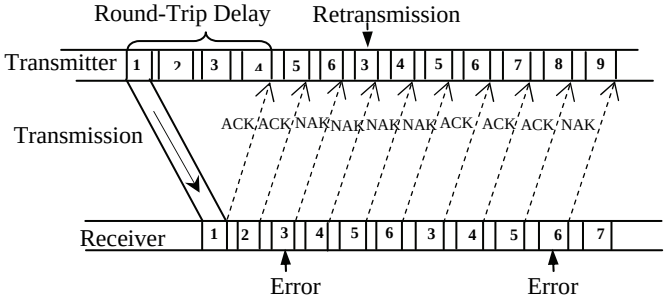


Fig. 2“Go-back – N” ARQ scheme.

“Selective repeat” ARQ scheme. The process of work in this system is that only the missing packets are transmitted.

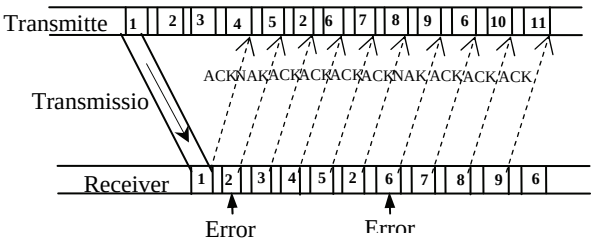


Fig. 3 “Selective repeat” ARQ scheme.

HARQ systems are used to increase the efficiency and the throughput of digital systems. These systems provide higher reliability than a FEC system and have higher throughput compared to an ARQ system.

There are three types of HARQ systems:

A. Type-I HARQ

Type-I HARQ scheme is used for simultaneous error detection and error correction [16].

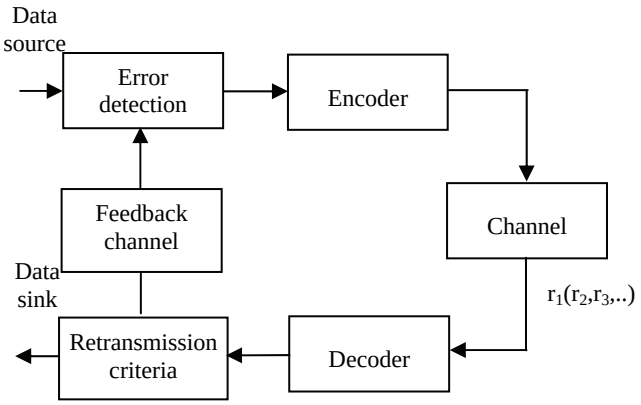


Fig. 4 General HARQ scheme

In this scheme two codes are used: one code for detection and other for error correction. Upon receiving of the packet containing errors the receiver first tries to correct it. If the correction failed, the packet is rejected and is retransmitted. In this system the code rate is fixed. This disadvantage can be avoided by optimizing the channel conditions. Type I HARQ is suitable for systems in which the constant noise and interference in the channel are present. In mobile channels, where the bit error rate changes, this scheme has following disadvantages:

- If the channel is very noisy it is possible the correcting capability of code to be insufficient;
- frequency of retransmission is increased and the throughput of the HARQ system is reduced;
- Type I HARQ provides a higher throughput of the ARQ scheme;
- limitation of throughput because of adaptive state changes in the channel.

B. Type II HARQ

In Type II HARQ the buffer memory is required.

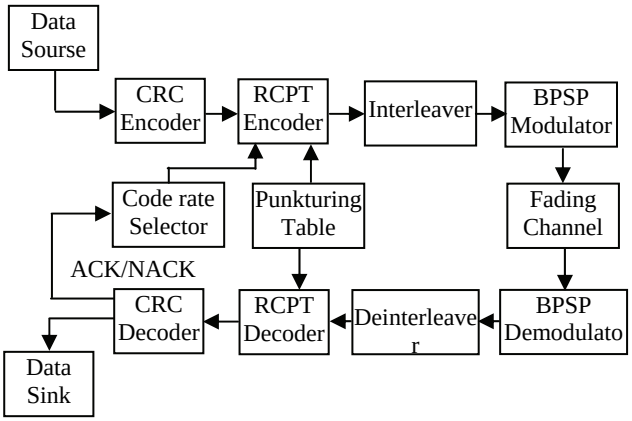


Fig. 5 Type II HARQ scheme

There are two basic systems of Type II HARQ.

- Chase combining;
- Incremental redundancy;

The general idea in Chase combining is following [2]:

- sending a number of copies for each packet of encoded data;
- enables the decoder to combine many parts of copies;
- decoding by measuring the signal-to-noise ratio of a prior decoding.

The essence of the Type II HARQ with incremental redundancy consists in the following:

- the received word is stored in a buffer in the case of errors;
- NACK is sent to the transmitter;
- the transmitter transmittes an additional number of encoded bits to check the original message and an attempt is made to correct the errors.

The main disadvantages of this scheme are:

- additional number of code symbols are sent with the packet which increase the size of transmitted information. A retransmission format depends on the applied strategy and the code for recovery of errors;
- a buffer with a large size is required which increases the cost of the system;
- greater complexity in decoding compared to type I HARQ.

The advantages are:

- better adaptation to the channel characteristics;
- outperform the Chase algorithm at high retransmission rates;
- throughput improves compared to Type I HARQ.

C. Type III HARQ

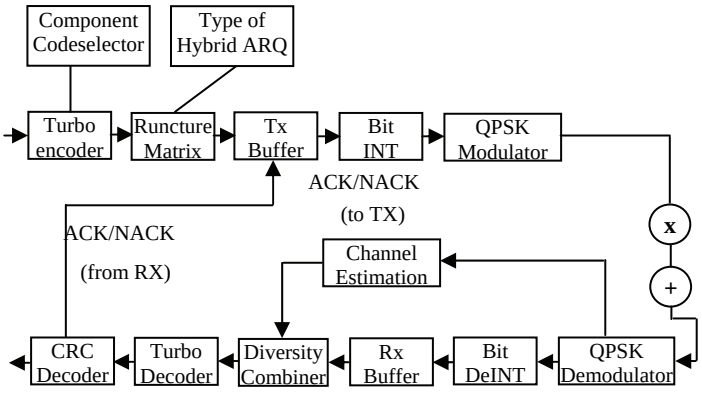


Fig. 6 Type III HARQ scheme

Type III HARQ is based on self-decoding and the source of information is needed for information extracting. The system has adaptable structure and determines an adaptive amount of additional information. At satisfactory channel conditions FEC code is used. The scheme also can be combined with incorrect data stored. According to the incremental redundancy used the schemes could be divided into two groups:

- with one version of the incremental redundancy (soft combining of the incremental redundancy);
- with several versions of the incremental redundancy (packets with detected errors are stored. The decoder combines the copies according to the ratio of signal to noise).

The main disadvantages of this scheme are:

- amount of redundancy information is increased in noisy channels;
- complex algorithms for coding and decoding are required in the above cases;

Type III HARQ has lower throughput than Type II HARQ in good channel conditions.

The advantages are:

- the throughput is improved compared to type I HARQ;
- the efficiency is improved compared to Type II HARQ[11];
- Type III HARQ has adaptive structure, i.e. it reduces the amount of redundancies to a minimum.

D. Reliability-Based HARQ

System model of Reliability-Based HARQ is shown in (fig. 7) [4]

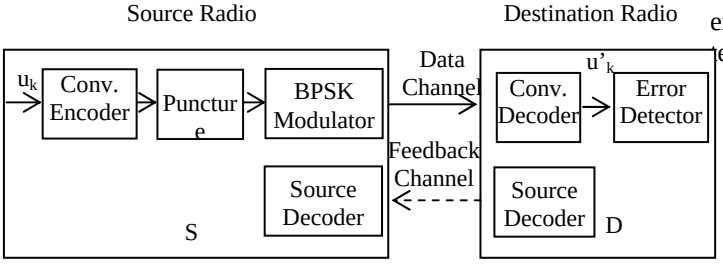


Fig. 7 System model for HARQ with convolutional codes

HARQ protocol is used in the model. This protocol which incorporates reliability information of data is called Reliability-Based HARQ. Bits are selectively retransmitted to the receiver in Reliable-Based HARQ. This is based on the estimated reliability of bits. Typically soft–input/soft-output decoding algorithms are used in the Reliable-Based HARQ. One of the possible algorithms is the maximum a posteriori probability (MAP) algorithm and its approximations [4]. In the log- MAP for every information bit u_k logarithm of the posteriori input probability is calculated by [15]

$$L(u_k) = \log\left(\frac{P(u_k = +1|y)}{P(u_k = -1|y)}\right), \quad (3)$$

where y is produced the codeword with noise.

The probability of bit error for an information bit can be estimated as the minimum of a posteriori probabilities, which is given by

$$P_b = \frac{1}{1 + e^{|L(u_k|y)|}} \quad (4)$$

The performance of different RB-HARQ schemes can then be compared with the channel capacity. Presented a technique to approximate the word error probability given log-likelihood ratios provided for the information bits. No error detecting code is needed when using the word error probability as a reliability criterion for retransmissions. The estimation can be improved by applying subblock by subblock decoding. The word error probability can be used as a reliability measure for ARQ protocols to decide whether a frame or codeword can be considered reliable enough to be accepted. The exact bit error probability for u_k can be calculated using the magnitude of the corresponding log-likelihood ratio

$$P_{b,k} = \frac{1}{1 + e^{|\tilde{u}_k|}}, \quad (5)$$

The average bit error probability P_b of the whole word can be obtained as

$$P_b = \frac{1}{K}\sum_{k=1}^{K} P_{b,k}, \quad (6)$$

This technique has the ability to improve performance and minimize the amount of retransmited bits. In recent years, more widely application the soft decoding founds, because exists a possibility to use interactive decoding. There are three techniques for reliability estimation [3,16]:

- log-likelihood ratios;
- determination the likelihood of the error bits;
- estimation of the error by using the reliability information.

The main disadvantages of this scheme are:

- according to [11] amount of additional information has increased significantly. To achieve a high reliability more complex algorithms are required;
- this type of system requires more memory which is required to maintain the combination of transmitted and retransmitted packets;
- the bad channel conditions and convolutional codes are extended time for analysis [14].

The advantages of this scheme are:

- the volume of calculations or decoding time is reduced [15];
- RB-HARQ is used interactive feedback to achieve higher reliability and performance.
- using convolutional codes in this systems are decreased the size of the retransmission requests.

III. CONCLUSION

The development of HARQ systems is passed different stages: first- HARQ with different modes of transmission - "Stop and Wait", "Go-back-N" and "Selective Repeat"; second - HARQ Type I, Type II, Type III and third- current RB-HARQ systems. This is due to the following:

- the intensive development of theoretical elaborations of efficient algorithms for encoding and decoding.
- the development of the technological base for the transmission, processing and storing of information.

Generally, the development of HARQ systems is expected in the following areas:

- new achievements in the field of coding and decoding of information,
- the elaboration and implementation of fast and efficient decoding algorithms,
- increasing of the reliability of transmission and accepting of information in the above two directions,
- evaluation the performance of RB-HARQ for fading channels and compare it with the performance of conventional HARQ techniques.
- examination the effect of node density on the interference power and overall system performance.

Finally, to achieve significant results in the communication the joint development of four main HARQ systems is required.

REFERENCE

[1] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Pearson Prentice Hall, 2005.

[2] D. Chase, "Code Combining: A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," IEEE Trans. on Commun., Vol. 33, pp. 593-607, 1985.

[3] J. M. Shea "Reliability-Based Hybrid ARQ", IEE Electronics Letters, , vol. 38, no. 13, pp. 644–645, 2002.

[4] A. Roongta and J. M. Shea, "Reliability-based hybrid ARQ using convolutional codes," in Proc. IEEE International Conference on Communications (ICC 2003), Anchorage, Alaska, pp. 2889–2893., 2003.

[5] V. Tripathi, E. Visotsky, R. Peterson, and M. L. Honig, "Reliability-based type II hybrid ARQ schemes," in Proc. IEEE Int. Conf. on Comm. (ICC 2003), Anchorage, Alaska, USA, 2003.

[6] E. Visotsky, Y. Sun, V. Tripathi, M. L. Honig, and R. Peterson, "Reliability-based incremental redundancy with convolutional codes," IEEE Trans. Commun., vol. 53, pp. 987–997, 2005.

[7] J. Moreira., P. Farrell "*ESSENTIALS of error-control coding*" John Wiley & Sons Ltd, 2006.

[8] V. K. Garg "*Wireless Communications and Networking"* Morgan Kaufmann Publishers, San Francisco, 2007.

[9] C.W. Wong, J. M. Shea, and Tan F. Wong "Secret Sharing in Fast Fading Channels based on Reliability-Based Hybrid ARQ" Military Communications Conference,. MILCOM 2008. IEEE pp.1-7, 2008.

[10] Y Inaba, T Saito, T Ohtsuki "Reliability-Based Hybrid ARQ (RB-HARQ) Schemes Using Low-Density Parity-Check (LDPC) Codes" IEICE TRANS. COMMUN., VOL.E89–B, NO.4, pp.1170-1177, 2006

[11] S. Lin, D. J. Costello and M. Miller, "Automatic-Repeat-Request error-control schemes," IEEE Commun. Mag., vol. 22, No. 12, pp. 5–17, 1984.

[12] S. Kallel, "Complementary punctured convolutional codes and their applications," IEEE Trans. Commun., vol. 43, pp. 2005–2009, 1995.

[13] M. Liinaharja, *"Studies on the Performance of some ARQ Schemes"* Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology, Communications Laboratory 2006.

[14] Xin Li, T. F. Wong, and J. M. Shea „Performance Analysis for Collaborative Decoding with Least-Reliable-Bits Exchange on AWGN Channels" IEEE Transactions on Communications, vol. 56, No. 1, 2008.

[15] E. Uhlemann *"Hybrid ARQ using Serially Concatenated Block Codes for Real-Time Communication an Iterative Decoding Approach"* Dissertation for the degree of Doctor of Licentiate of Engineering Halmstad University Department of Computer Engineering, 2000.

[16] J. C. Fricke, H. Schoeneich, and P. Hoeher "Reliability-Based HARQ using Word Error Probabilities", in Proc. NEWCOM-ACORN Joint Workshop, Vienna, Austria, 2006.

A Model for Integration of Railway Information Systems Based on Cloud Computing Technology

Slađana Janković[1], Snežana Mladenović[2], Slobodan Mitrović[3], Norbert Pavlović[4], Slaviša Aćimović[5]

Abstract - **This paper proposes the *Enterprise Application Integration (EAI)* modelin the field of railways, based on combining information integration and portal integrationin the *cloud computing*technological environment. Information integrationis carried outin a common *SQL Azure* database. The portal integrationis enabled with*Windows Azure hosted service*. The proposed model was implemented in a case studyof integration of information systems that are used for the railroad crossings management inthe Serbian Railways.**

Keywords – **Railway information systems, EAI, Cloud computing, SQL Azure database, Windows Azure hosted service.**

I. INTRODUCTION

Railways corporations(Railways) are the very complex systems, so they are organizationally divided into a great number ofunits, such as directorates, sectors, services, etc.Each organizational unit uses applications and databases designed to meet their specific needs. It is not uncommon that one particular entity from the real system is modeled in multiple databases that are used in different organizational units [1]. An organizational unit of the railways often use the data generated and updated by another organizational unit [2]. For example Sector for transportation of goods and passengers base their work on the data given by the Directorate of Infrastructure (data on construction, electrotechnical, telecommunication and transport infrastructure and its maintenance), the Department for maintenance of rolling stock and the Department for towing trains.In addition, each unit uses its own databases and applications. This creates redundancy and inconsistency of data. This means that cooperation between organizational units are often based on reports that have different syntax and semantics [3].To avoid redundancy and inconsistency of data and to avoid incompatible reports, necessary to enable the integration of databases and *Enterprise Application Integration* (*EAI*) using different organizational units of the railway.

In the second section of paper will be proposed model of *Enterprise Application Integration*that can be used on the Railways. The third section describes the implementation of the proposed model in the case study of IS which is used for the railroad crossingsmanagement in the Serbian Railways. Finally, the paper concludes the most important results achieved by the integration of railway IS according to proposed model.

II. A MODEL FOR INTEGRATION OF RAILWAYIS

We propose a model of integration of railways IS which is based on combining information integration and portal based integration in the cloud computing technological environment. Cloud computing is an area in which highly scalable IT facilities are provided as a service, which is delivered via Web [4]. Cloud includes both software and hardware services. Using the cloud, if not for free, is charged only for what is used and how it is used. This is known as *pay-as-you-go*. In order to use all the services and opportunities that come from the cloud, user must have a Web browser and an Internet connection and also must be logged in. The user does not need to worry for the latest software version or whether the database is updated.

Information integration will be achieved through developing a common database within *SQL Azure* platform. Portal based integration will be realized through development and use of the Web portal that will be hosted as a *Windows Azure* service. Common database and Web portal will be designed to meet the needs of all organizational units within the railways.Microsoft SQL Azure Database is a cloud-based relational database service built on SQL Server technologies. It provides a highly available, scalable, multi-tenant database service hosted by Microsoft in the cloud. Developers do not have to install, setup, patch or manage any software. High availability and fault tolerance is built-in and no physical administration is required. SQL Azure Database supports Transact-SQL(T-SQL). SQL Azure Database can help reduce costs by integrating with existing toolsets and providing symmetry with on-premises and cloud databases.

There are several ways to incorporate SQL Azure in applications [5], however there are two application patterns to access the SQL Azure data:

1. On-Premises Applications,
2. Hosted Applications residing in Cloud.

In a traditional on-premise application, the application code and database are located in the same physical data center. SQL Azure and the Azure Services Platform offer many alternatives to that architecture. The Fig. 1. demonstrates two

[1]Slađana Janković is with the Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe 305, 11000 Belgrade,Serbia,E-mail: s.jankovic@sf.bg.ac.rs.

[2]Snežana Mladenović is with the Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: snezanam@sf.bg.ac.rs.

[3]Slobodan Mitrović is with the Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: s.mitrovic@sf.bg.ac.rs.

[4]Norbert Pavlović is with the Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: n.pavlovic@sf.bg.ac.rs.

[5]Slaviša Aćimović is with the Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, E-mail: s.acimovic@sf.bg.ac.rs.

generalized alternatives available for how application can access SQL Azure data.In Scenario Aon the left, application code remains on the premises of corporate data center, but the database resides in SQL Azure. Application code uses client libraries to access database(s) in SQL Azure.

In Scenario Bon the right, application code is hosted in the Windows Azure and database resides in SQL Azure. Application can use the same client libraries to access database(s) in SQL Azure as are available in Scenario A. The Scenario Bclient premises may represent an end user's Web browser that is used to access Web application. The Scenario B client premises may also be a desktop application that uses the benefits of the Entity Data Model and the ADO.NET Data Services client to access data that is hosted in SQL Azure.

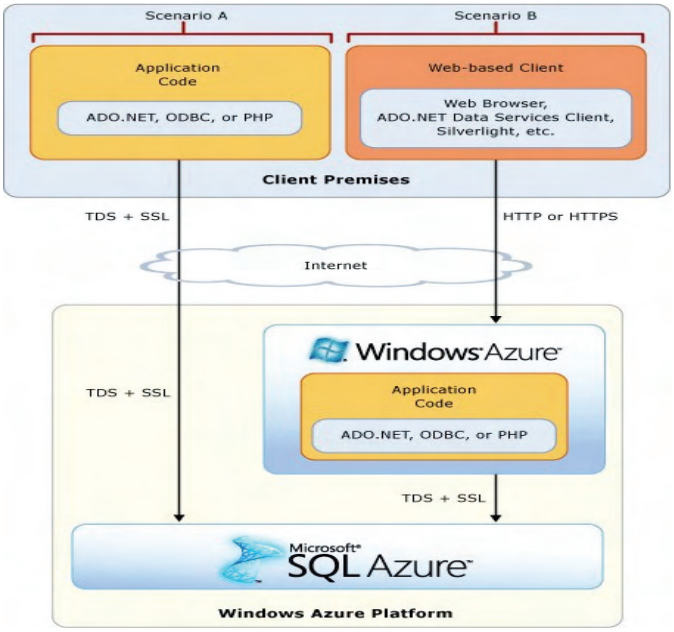


Fig. 1. Scenarios to access the SQL Azure data

Our railway IS integration model based on Scenario B. Scenario would be implemented as follows:

1. To consolidate the data updated and used by different sectors within the railway, proposed the integration of data within a SQL database Azure;
2. Application that includes data access and business logic will be implemented as a Web portal and hosted as a service on Windows Azure platform;
3. Client application will be a Web browser that will allow access to the Web application.

There are three categories of cloud computing services [6]:

- *Software-as-a-service, SaaS* - software that is implemented as a hosted service and can be accessed via the Internet,
- *Platform-as-a-service, PaaS* - platform that can be used for deployment of applications provided by the clients or service provider partners,
- *Infrastructure-as-a-service, IaaS* - computer infrastructure, such as servers and storage.

The model we propose combines all three categories of services. Using SQL database Azure means using SQL Azure platform for development and database management, ie. PaaS and also the using of the infrastructure, ie. IaaS. The using of Web application as Windows Azure hosted service, means using of the SaaS. Cloud computing offers four deployment models: Private Cloud, Public Cloud, Community Cloud, Hybrid Cloud. The proposed integration model of therailways IS involves the Community cloud delivery model. This means that only those computers which belong to Serbian Railways IP range should have right of access to a hosted service.

III. Implementation of the Proposed Model

A. *Architecture of The System*

The proposed model involves the development of the classic n-tier applicationwith the following layers:

- Database layer – SQL Azure database.
- Data access layer – Windows Azure Hosted Service, ADO.NET Entity Framework,
- Presentation layer – ASP.NET application.

B. *SQL Azure Database*

Individual organizational units within the Serbian Railways update the data related to railroad crossings for which they are responsible. This doesn't mean that other types of data are not necessary. In order to properly determine the appropriate safety measures for a railroad crossing, it is necessary to dispose of data about its current state, the actual volume of traffic and safety parameters. All these information categories are merged into a single *SQL Azure* database of Serbian railroad crossings. One of 16 database tables is shown in Fig. 2.

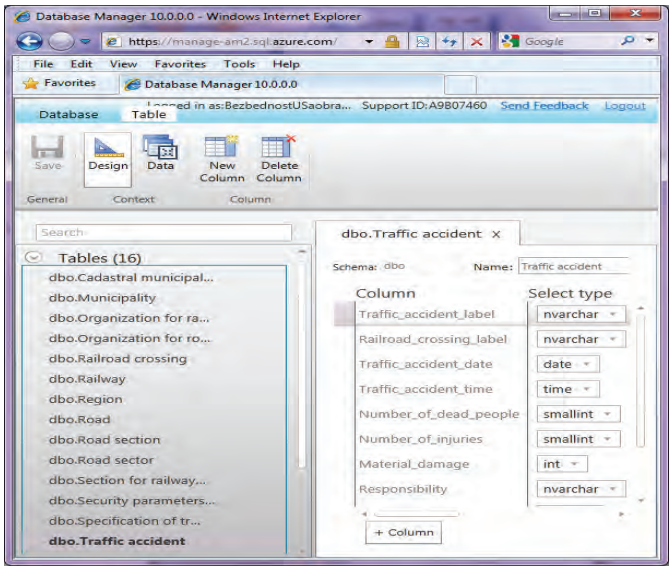


Fig. 2. Traffic accident table in SQL Azure database

C. *Windows Azure Hosted Service*

Windows Azure platform hosts the service calles *Serbian railroad crossings*, supplied with domain http://srrc.cloudapp.net (Fig. 3). This service is developed as WindowsAzureProject in IDE Visual Studio 2010. A Windows Azure application can be created using three kinds of roles: Web roles, intended primarily for running Web-based applications; Worker roles, designed to run a variety of code like a simulation, or video processing; VM roles, which can run a user-provided Windows Server 2008 R2 image. In our case study, the Windows Azure application is realized asWebRole ASP.NET application called SERBIAN RAILROAD CROSSINGS PORTAL(SRCP).

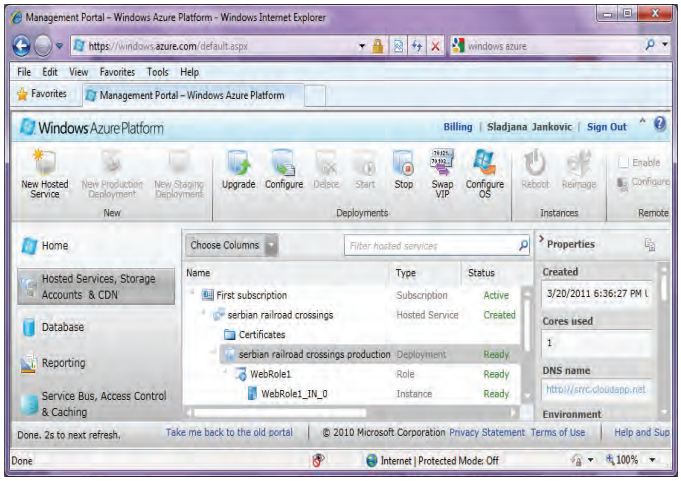


Fig. 3. Serbian railroad crossings - Windows Azure hosted service

The web role provides support for presenting a user facing frontend through IIS. SRCP uses the ADO.NET Entity Framework - a set of technologies in ADO.NET that support the development of data-oriented applications. The EntityDataSource control allows us to bind Web controls on a page to data in our Entity Data Model (Fig. 4).

D. *Web Based Client*

Client Web application *SERBIAN RAILROAD CROSSINGS* PORTAL allows users to view and/or update content of the *SQL AzureSerbian railroadcrossings* database. The main application menu consists of: *Railways, Roads, Railroad crossings, Traffic and Traffic Safety*. For example, *Traffic* menu consist of:

- *Traffic load on railroad crossings,*
- *Traffic load on railroad crossings – structure,*
- *Traffic load on railroad crossings – average retentions.*

*Traffic Safety*menu consist of:

- *Traffic accidents on railroad crossings,*
- *Security parameters of railroad crossings,*
- *Security parameters – responsibility structure.*

Fig. 5. shows Web application page designed for viewing the safety parameters of railroad crossings. Safety parameters are calculated on the basis of data on traffic accidents that occurred on a railroad crossing, which is stored in the same SQL Azure database. It is possible to calculate the safety parameters for the transition to the selected year, or calculate and display the safety parameters for all crossings.

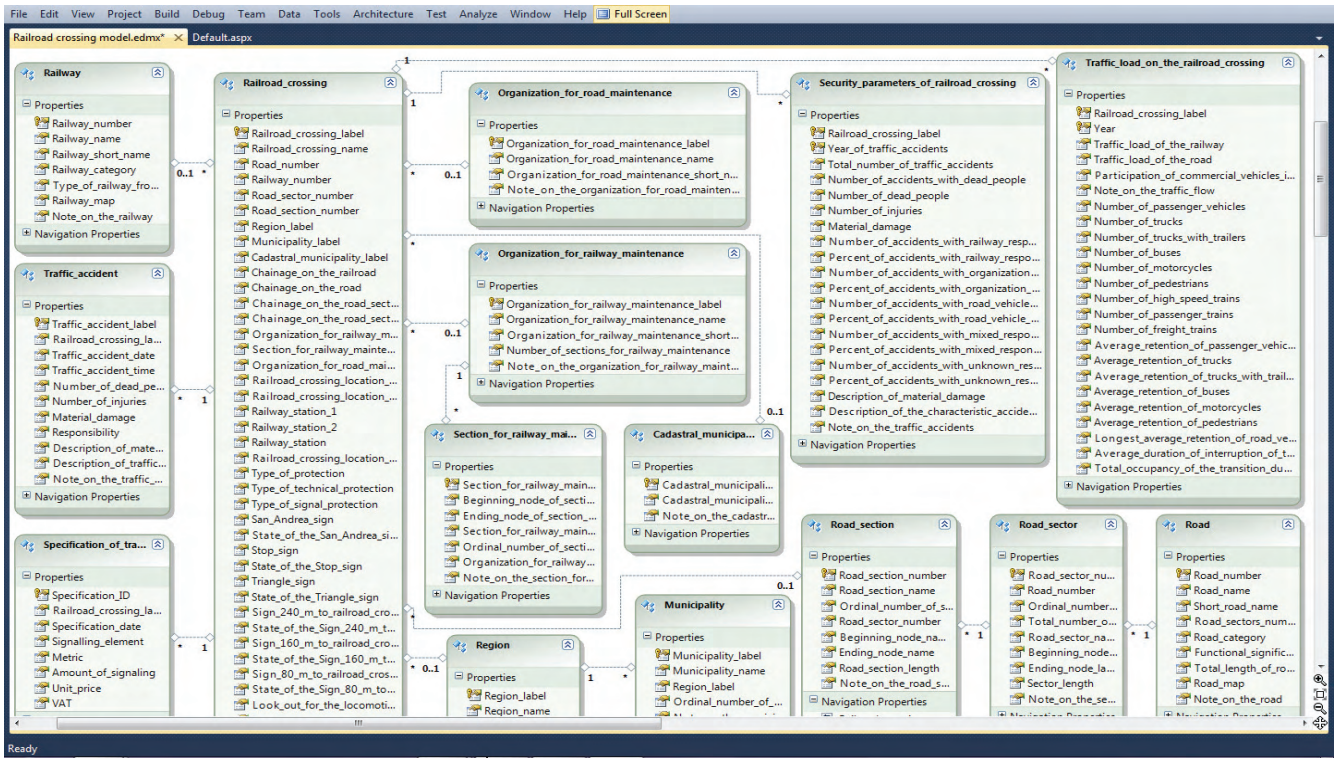


Fig. 4. Entity Data Model of ASP.NET application SERBIAN RAILROAD CROSSINGS PORTAL

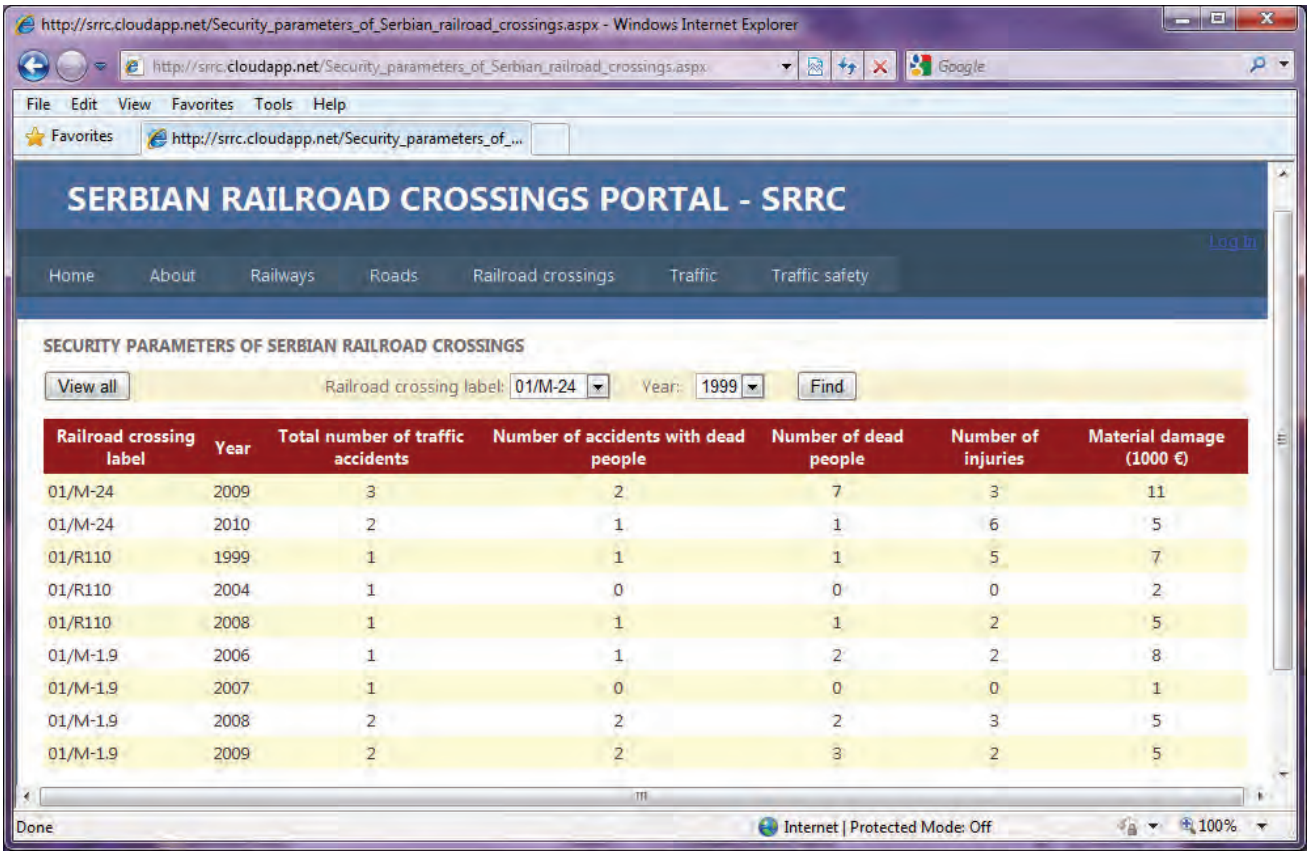


Fig. 5. SERBIAN RAILROAD CROSSINGS PORTAL – Security parameters of Serbian railroad crossings

IV. Conclusion

SQL Azure Database allows consolidation of multiple data sources in the cloud and enable secure access from multiple locations, desktop and/or devices.Hosting application code in Windows Azure is beneficial to the performance of application because it minimizes the network latency associated with application's data requests to SQL Azure.

Development of database on SQL Azure platform as well as Web applications hosted on Windows Azure platform allows different railways organizational units to use the same database, update only the data for which they are responsible, generate uniform reports. This eliminates the redundancy and ensures data integrity. End users of database and Web application should use only a Web browser on their computers.

Cloud computing technology was designed primarily to increase efficiency [7]. Hence, the proposed integration model emphases following benefits: the flexibility, scalability, portability, usage time saving as well as response times, minimization the risk of poor implementation of physical infrastructure, fixed costs reduction and acceleration of innovation process. The most serious drawback of the proposed model and technological solution is its dependence of cloud providers and partial loss of control over the system of governance. Judging by the large international companies that rely on cloud computing technology the risk should be awarded.

Acknowledgement

Thiswork is partially supported by the Ministry of the Science and technological development of the Republic of Serbia under No. 036012.

References

[1] S. Janković, S. Mladenović, “Some Aspects of Interoperability of Enterprises and Their Applications”, SYMOPIS 2008, Conference Proceedings, pp. 71-74, Soko Banja, Serbia, 2008.

[2] S. Janković, “Interoperability of transport business systemsbased on the integration of service-oriented B2B applications”, InfoM, vol. 36/2010,pp. 4-12, 2010.

[3] W. Lee, S. Tseng, W. Shieh, “Collaborative real-time traffic information generation and sharing framework for the intelligent transportation system”, Information Sciences, vol. 180, pp. 62-70, 2010.

[4] D. S. Linthicum, *Cloud Computing and SOA Convergence in Your Enterprise*, Boston, Pearson Education, 2010.

[5] D. Betts, S. Densmore, R. Dunn, M. Narumoto, E. Pace, M. Woloski, *Developing Applications for the Cloud on the Microsoft® Windows Azure™ Platform*, Microsoft, 2010.

[6] J. Krishnaswamy, *Microsoft SQL Azure: Enterprise Application Development*, Birmingham – Mumbai, Packt Publishing, 2010.

[7] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, A. Ghalsasi, “Cloud computing - The business perspective”, Decision Support Systems, vol. 51, pp. 176-189, 2011.

Efficient Implementation of Hashing in BDD Package

Miloš M. Radmanović[1]

Abstract – **The efficient manipulation of Boolean functions is an important component of CAD tasks. Binary Decision Diagram (BDD) packages have always been sensitive to hash design. This paper describes the use of various hash keys in implementation of basic BDD algorithms. The various hash key strategies have been implemented within a BDD package. In this paper, I experimentally performed a detailed analysis of hashing in BDD package using BDD algorithms computation and direct performance monitoring. The ultimate goal is to exceed the computation performance for various BDD algorithms of BDD packages.**

Keywords – **Boolean functions, BDD package algorithms,hash key strategies.**

I. INTRODUCTION

Binary Decision Diagrams (BDDs) have become the dominant data structure for representing Boolean functions in computer-aided-design (CAD)applications [1]. They are widely used in various areas of CAD: logic synthesis, testing, simulation, design and simulation verification [2]. [3], [4].

In practice, the memory required by large BDDs is typically the limiting factor for CAD tools. In same cases, especially with various BDD algorithms, run time is also a limiting factor. Therefore, considerable research has been intended for more efficient BDD algorithms implementation [4], [5], [6], [7], [8].

Various BDD algorithms are usually built on the top of a BDD package. Many BDD package implementations have been built in a variety of programming languages (C, C++, Lisp, Java) [9], [10], [11], [12]. The most of BDD packages are freely available in public domains (on the Web). The packages CAL (UC Berkeley, USA), CMU (CMU/ATT, USA), and CUDD (Boulder, CO, USA)are famous.The packages IBM (IBM Watson, USA), Tiger (Bull/DEC/Xorix, USA) and TUD (Darmstadt, Germany) are popular [13]. The choice of the BDD package might be guided by the following aspects: functionality, software interface, robustness, reliability, portability, support and performance. Naturally, performance is of concern [14], [15], [16], [17], [18], [19]. But, from performance comparative studies, BDD packages behave similarly as long as they are not put to theextreme [13]. They are many parameters that influence the run-time of the BDD package, for example: programming language, software and hardware platform, BDD node structure, type of garbage collection, unique and operation hash table size, and hash keystrategiesof tables [1], [20], [21].

The concept of hash miss complexity of the BDD package has been introduced by paper [21]. It also has been shown that run-time of BDD algorithms depends on BDD node and operations hashing strategies of the BDD package.

This paper describes the use of various hash keys strategies in implementation of a BDD package and their influence on the run-time of the basic BDD algorithms on the top of a BDD package. The various hash key strategies have been implemented within a BDD package.I build a simple software tool for evaluating algorithms depending on hash key of the BDD package. Using this software tool, I experimentally performed a detailed analysis of hashing in BDD package. The analysis of hashing usesbasic BDD algorithms computation and direct performance monitoring of the BDD package.

This paper is organized as follows: Section 2 shortly introduces the BDD representation of Boolean function.Section 3 describes the BDD package technique includingbasic BDD algorithms. Section 4presents the hash key strategies of the BDD packages. Section 5shows experimental analysis of hashing in BDD package and gives some examples of using various hash keys. Some concluding remarks end the paper.

II. BINARY DECISION DIAGRAM

The Binary Decision Diagram (BDD) is a graphic representation of the Shannon's expansion of a Boolean function. The concept of BDD was first proposed by Lee in 1959 [22]. It has been developed into a useful data structure by Akers [23] and later by Bryant [2], who introduced a concept of reducedordered BDD (ROBDD), and a set of efficientoperators for their manipulation.

The BDD is directed acyclic graph that contain non-terminal nodes, two terminal nodes, and edges. Non-terminal nodes are labeled with variables x_i and have two outgoing edges. Outgoing edges are labeled '0' and '1' values of variable x_i. Terminal nodes contain values '0' and '1'. The truth table entry of Boolean function labels edges from the root node to the corresponding terminal node. An example of the BDD representation for the function $f(x_1, x_2, x_3) = x_1 + x_2\bar{x}_3$ is shown in Figure 1.

The strength of BDDs is that they can represent Boolean function data with high level of redundancy in a compact form, as long as the data is encoded in such way that the redundancy is exposed.

It is well known that the size of the BDD for a given Boolean function depends on the variable order chosen for the function.

Other variants of BDD have been developed to address some of the problems with BDDs. Boolean expression diagrams (BEDs) are a generalization of BDDs that can represent any Boolean circuit in linear space.Zero-suppressed binary decision diagrams (ZDDs) are similar to BDDs but use a different reduction rule. ZDDs are generally more efficient for representing sparse sets. Binary moment diagrams

[1]MilošM. Radmanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.radmanovic@gmail.com

(BMDs) are a generalization of BDDs to linear functions over other domains than Booleans, such as integers or real numbers [24].

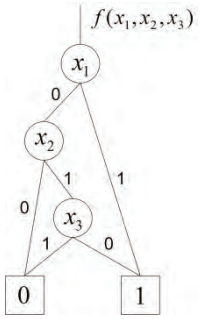


Fig. 1. BDD for the function $f(x_1,x_2,x_3) = x_1 + x_2\bar{x}_3$.

III. BDD PACKAGE TECHNIQUE

In this section I describedthree data structures that are fundamental to most BDD packages.

The main principles of BDD package implementation are [2], [3], [13]:

- BDD nodes are data structures that contain a variable identifier, “1” and “0” children pointers and “next” pointer that links nodes together belonging to the same collision chain in the unique table.
- BDD nodes are kept in a unique table.
- BDD operations are sped up by using an operation table.
- The order of variables may be changed to reduce the total number of nodes.
- Recycling of nodes is easily implemented by keeping a reference count for each node.
- BDD packages use breadth-first traversal of the directed acyclic graphs.

It is obvious what needs to be stored for each node in a BDD package data structure: the variable field that labels the node and two edges field to point children nodes. There should obviously also be a field to identifier non terminal nodes from terminals. BDD traversal uses a “visited” field.Garbage collection might use a reference counter field.Chaining of nodes in a hash table needs a “next” field. The decisions made in defining the basic node data structure have an immediate impact on several related aspects. All this data requires certain memory space and needs to be packed into a node structure or object [14].

The unique table maps a triple of (*v*, *G*, *H*) of BDD node, where *v* is the variable identifier, *G* is the node connected to the "1" edge, and *H* is the node connected to the "0" edge. Each node in the BDD has an entry in the unique table. Before a new node is added to the BDD, a lookup in the unique table determines if the node for that function already exists [3], [14].

The If-Then-Else or *ITE* operator forms the core of the BDD package. *ITE* is a Boolean function defined for inputs *F*, *G* and *H*which computes: “If*F* Then *G* Else *H*”. ITE operations can be used to implement all two-variable Boolean operations.The operation table records results of operations on BDDs. Typically it stores the fact that *R* = *op*(*F*,*G*,*H*), i.e., the BDD *R*is the result of the operation *op* applied to three BDD arguments *F*, *G*, and *H*, in a operation table [3], [14].

Garbage collection invalidates entries that refer to dead nodes. Most BDD packages use a reference counting garbage collector.

Originally designed to support the standard operations on Boolean functions, BDD packages have grow to build in various algorithms. Basic BDD algorithms are based on a recursive formulation that leads to a depth-first traversal of the directed acyclic graphs representing the BDD. The depth-first traversal visits the nodes of the BDD on a path-by-path basis. The large in-degree of a typical BDD node makes it is impossible to assign contiguous memory locations for the BDD nodes along a path. Therefore, the recursive depth-first traversal leads to an extremely disorderly memory access pattern [5].

The most of BDD algorithms are the result ofsome other basic BDD algorithms yet to be completed. A comprehensiveset of BDD manipulation algorithms are implemented usingthe above techniques. The most common BDD algorithms are, for example,BDD construction,BDDsaddition,and BDD printing [2], [3], [13].

IV. HASH KEY STRATEGIES OF THE BDD PACKAGES

The unique table is usually implemented as a hash table [25]. For greater flexibility, open hashing with collision chains is normally used. The collision lists can be kept sorted [1] to reduce the number of memory accesses required for a lookup on average. The unique table might be divided in subtables, one for each variable [14].

The size of the hash table is either a prime number or a power of two. For hashed insertion, the hash table is used in a two-way associative manner: the hash function calculates an index *k* (hash key) for a node (v, G, H) and the node is found either at *k* hash entry or in the collision list. The hash key should obviously be composed of the memory position of the node and its successors or by defining a signature for each node which consists of the variable associated with that node and a pseudo random number [1].

It should be noticed that the hash key calculation of the unique table can be defined as follows:

$$UTH(v,G,H) = (m(G)\ op\ m(H)) \mod hs \quad (1)$$

where *m*(*G*) and *m*(*H*) denote memory addresses of the nodes *G* and *H*, *hs* denotes hash table size, and *op* represents some logic or arithmetic operation.

The operation table is usually implemented as a unique table [25]. The hash function calculates an index *k* (hash key) for anoperation (*F*, *G*, *H*) and the operation is found either at *k* hash entry or in the collision list.

It should be noticed that the hash key calculation of the operation table can be defined as follows:

$$OTH(F,G,H) = (m(F)\ op\ m(G)\ op\ m(H)) \mod hs \quad (2)$$

where *m*(*F*), *m*(*G*) and *m*(*H*) denote memory addresses of the nodes *F*, *G* and *H*, *hs* denotes hash table size, and symbol *op* represents some logic or arithmetic operation.

Experimentally, it is proved that efficient operation for symbol *op* can be + or $\oplus$ [1], [8], [9], [13], [14].

From previous consideration, it is evident that various hash key strategies can be defined:

(1) $UTH(v,G,H) = (m(G) + m(H)) \mod hs$

$OTH(F,G,H) = (m(F) + m(G) + m(H)) \mod hs$

(2) $UTH(v,G,H) = (m(G) + m(H)) \mod hs$

$OTH(F,G,H) = (m(F) \oplus m(G) \oplus m(H)) \mod hs$

(3) $UTH(v,G,H) = (m(G) \oplus m(H)) \mod hs$

$OTH(F,G,H) = (m(F) + m(G) + m(H)) \mod hs$

(4) $UTH(v,G,H) = (m(G) \oplus m(H)) \mod hs$

$OTH(F,G,H) = (m(F) \oplus m(G) \oplus m(H)) \mod hs$

V. Experimental Results

It is interesting to consider proposed hashing strategies that could be applied to BDD package and tested on basic BDD algorithms.

Below I give tables of BDD package performanceusing different basic BDD algorithms. I performed the testing on a PC Pentium IV on 2,66 GHz with 4GB of RAM (MS Windows 7). The memory usage for all tests was limited to 2 GB.The size of unique and operation table was limited to 65536 entries. The garbage collection was activated if the total number of BDD nodes and operations in memory became greater than 524288. All benchmarks were used in the Espresso-mv or pla format [31].

Table 1, Table 2, and Table 3 gives the complete list of experimental results of BDD package time statistics for hash key strategies proposed in the previous section tested on BDD construction, BDD addition and BDD printing algorithms, respectively.

All times in all tables are given in seconds. It should be noticed that hash key strategies in previous section and in headers of all tables have same denotation with symbols (1), (2), (3), and (4). Last four columnsfor each table presents comparable time performance of BDD package hash key strategies.

There are 15 of 19 benchmarks in table 1 for which algorithm has minimum computation time for strategy (4). According to that fact, the BDD construction algorithm is in most cases efficient for strategy (4).

There are 14 of 19 benchmarks in table 2for which algorithm hasminimum computationtime for strategy (1). According to that fact, the BDD addition algorithm is in most cases efficient for strategy (1).

There are only 17 of 19 benchmarks in table 3for which algorithm has minimum computationtime for strategy (2). According to that fact, the BDD printing algorithm is in most cases efficient for strategy (2).

For comparison, the hash key strategy (3) was least efficient for all BDD basic algorithms.

This BDD basic algorithms dependency of proposed hash key strategies is a major advantage of this paper.

TABLE I
BDD Package Time Statistic For Various Hash Key Strategies Tested on BDD Construction Algorithm

Fun. name	inp/out/cubes	algorithm performance [s] (1)	(2)	(3)	(4)
alu4	14 / 8 / 1028	0.20	0.15	0.18	0.15
apex1	45 / 45 / 206	5.78	5.18	5.04	4.80
apex2	39 / 3 / 1035	3.38	3.31	3.32	3.40
apex4	9/19/438	0.08	0.05	0.04	0.04
apex5	117/88/1227	0.34	0.30	0.29	0.27
b12	15/9/431	0.01	0.01	0.01	0.01
clip	9/5/167	0.01	0.01	0.01	0.01
con1	7/2/9	0.02	0.01	0.01	0.01
cordic	23/2/1206	0.08	0.06	0.05	0.03
cps	24/109/654	0.16	0.15	0.15	0.14
duke2	22/29/87	0.03	0.03	0.02	0.03
e64	65/65/65	0.02	0.01	0.01	0.01
ex4p	128/28/620	0.01	0.01	0.01	0.01
ex1010	10/10/1024	0.02	0.02	0.02	0.02
misex2	25/18/29	0.05	0.05	0.05	0.05
misex3	14/14/1848	0.24	0.20	0.20	0.21
misex3c	14/14/305	0.05	0.03	0.02	0.04
table3	14/14/135	0.03	0.02	0.01	0.01
table5	17/15/158	0.02	0.01	0.01	0.01

TABLE II
BDD Package Time Statistic For Various Hash Key Strategies Tested on BDD Addition Algorithm

Fun. name	inp/out/cubes	algorithm performance [s] (1)	(2)	(3)	(4)
alu4	14 / 8 / 1028	0.40	0.44	0.42	0.45
apex1	45 / 45 / 206	9.33	9.25	9.53	9.22
apex2	39 / 3 / 1035	5.57	5.75	5.74	5.63
apex4	9/19/438	0.31	0.26	0.29	0.31
apex5	117/88/1227	0.31	0.43	0.35	0.33
b12	15/9/431	0.02	0.03	0.03	0.04
clip	9/5/167	0.03	0.02	0.03	0.02
con1	7/2/9	0.03	0.04	0.04	0.04
cordic	23/2/1206	0.21	0.26	0.25	0.23
cps	24/109/654	0.42	0.51	0.57	0.59
duke2	22/29/87	0.07	0.07	0.07	0.07
e64	65/65/65	0.06	0.08	0.09	0.12
ex4p	128/28/620	0.11	0.16	0.19	0.21
ex1010	10/10/1024	0.16	0.18	0.25	0.24
misex2	25/18/29	0.11	0.12	0.13	0.12
misex3	14/14/1848	0.66	0.68	0.63	0.61
misex3c	14/14/305	0.14	0.12	0.11	0.10
table3	14/14/135	0.08	0.11	0.11	0.12
table5	17/15/158	0.04	0.05	0.05	0.06

TABLE III
BDD PACKAGE TIME STATISTIC FOR VARIOUS HASH KEY STRATEGIES TESTED ON BDD PRINTING ALGORITHM

Fun. name	inp/out/cubes	Algorithm performance [s] (1)	(2)	(3)	(4)
alu4	14 / 8 / 1028	0.11	0.10	0.11	0.10
apex1	45 / 45 / 206	2.25	2.21	2.23	2.22
apex2	39 / 3 / 1035	1.81	1.80	1.83	1.84
apex4	9/19/438	0.03	0.02	0.02	0.02
apex5	117/88/1227	0.16	0.15	0.16	0.16
b12	15/9/431	0.01	0.01	0.01	0.01
clip	9/5/167	0.01	0.01	0.01	0.01
con1	7/2/9	0.01	0.01	0.01	0.01
cordic	23/2/1206	0.04	0.03	0.04	0.04
cps	24/109/654	0.07	0.06	0.07	0.06
duke2	22/29/87	0.02	0.02	0.02	0.01
e64	65/65/65	0.01	0.01	0.02	0.02
ex4p	128/28/620	0.01	0.01	0.01	0.01
ex1010	10/10/1024	0.02	0.01	0.02	0.02
misex2	25/18/29	0.03	0.04	0.04	0.03
misex3	14/14/1848	0.16	0.13	0.14	0.13
misex3c	14/14/305	0.02	0.01	0.02	0.02
table3	14/14/135	0.02	0.01	0.02	0.02
table5	17/15/158	0.02	0.01	0.02	0.02

VI. CONCLUSION

This paper describes the use of various hash keys in implementation of basic BDD algorithmson the top of a BDD package.I experimentally performed a detailed analysis of hashing in BDD package using BDD algorithms computation and direct performance monitoring. The ultimate goal is to exceed the computation performance for various BDD algorithms of BDD packages.

From experimental results, it is evident that different hash key strategies can be used in different points of BDD package computation. Especially, hash key strategies in basic BDD ccomputationsare promising. Further work will be devoted to deeper exploiting these possibilities as well as exploiting different hash key strategies of BDD packages.

REFERENCES

[1] D. Long, “The Design of Cache-friendly BDD Library“,Proceedings of the 1998 IEEE/ACM international Conference on CAD, pp. 639 - 645, 1998.

[2] R. Bryant, "Graph-based Algorithms for Boolean Function Manipulation", IEEE Trans. Computers, vol C-35, pp. 667-691, 1986.

[3] K. Brace, R. Rudell, R. Bryant, "Efficient implementation of a BDD package", Proc. Design Automation Conf., pp. 40-45, 1990.

[4] R. Rudell, "Dynamic Variable Ordering for Binary Decision Diagrams", Proc. Conf. on CAD, pp. 42-47, 1993.

[5] J. Sangavi, R. Ranjan, R. Bryton, A. Sangiovanni-Vincentelli, "High Performance BDD Package Based on Exploiting Memory Hierarchy", Proc. of the Design Automation Conf., 1996.

[6] H. Ochi, K. Yasuoka, S. Yajima, "Breadth-First Manipulation of Very Large Binary-Decision Diagrams", Proc. Int. Conf. on CAD, pp. 48-55, 1993.

[7] P. Ashar, M. Cheong, "Efficient Breadth-First Manipulation of Binary Decision Diagrams", Proc. Int. Conf. on CAD, pp. 622-627, 1994.

[8] G. Janssen, “Design of Pointerless BDD Package,“10th Int. Workshop on Logic & Synthesis Granlibakken, Lake Tahoe, CA, 2001.

[9] F. Somenzi, "CUDD: CU decision diagram package", Public Software, University of Colorado, Boulder, CO, 1997, http://vlsi.colorado.edu/~fabio/.

[10] T. Stornetta, F. Brewer, “Implementation of an Efficient Parallel BDD Package”, Proc. of 33rd Design Automation Conference, pp. 641–644, 1996.

[11] R. Sumners, "Correctness Proof of a BDD Manager in the Context of Satisfiability Checking", Proc. of ACL2 Workshop 2000, Technical Report TR-00-29, 2000.

[12] C. Krieger, "A Java 1 Implementation of a BDD Package", University of Utah, 1998, http://www.cs.colostate.edu/~krieger/.

[13] G. Janssen, “A Consumer Report on BDD Packages,“Proc. of the 16th Symposium on Integrated Circuits and Systems Design, pp.217-223, 2003.

[14] F. Somenzi, “Efficient Manipulation of Decision Diagrams,” Int. J. Software Tools for Technology Transfer (STTT), vol. 3, no.2, pp. 171-181, 2001.

[15] B. Yang, R. Bryant, D. O’Hallaron, A. Biere, O. Coudert, G. Janssen, R. Ranjan, F. Somenzi, “A Study of BDD Performance in Model Checking,” Int. Proc. FMCAD, Palo Alto, CA, pp. 255-289, 1998.

[16] S. Manne, D. Grunwald, F. Somenzi, "Remembrance of Things Past: Locality and memory in BDDs", Int. Proc. of the 34th ACM/IEEE Design Automation Conference, 1997, pp. 196-201.

[17] M. Lam, Context-Sensitive Pointers Analysis Using Binary Decision Diagrams, John Whaley, 2007.

[18] M. Sentovich, "A Brief Study of BDD Package Performance". Int. Proceedings of the Formal Methods on CAD, 1996, pp. 389-403.

[19] S. Minato, N. Ishiura, S. Jajima, "Shared Binary Decision Diagram with Attributed Edges for Efficient Boolean Function Manipulation", Int. Proc. of the 27th ACM/IEEE Design Automation Conference, 1990, pp. 52-57.

[20] B. Yang, Y., Chen, R. Bryant, D. O'Hallaron, "Space and time efficient BDD construction via working set control". Int. 1998 Proceedings of Asia and South Pacific Design Automation Conference , 1998, pp. 423-432.

[21] N. Klarlund, T. Rauhe, "BDD algorithms and cache misses", BRICS Report Series RS-96-5, Department of Computer Science, University of Aarhus, 1996.

[22] C. Lee, “Representation of Switching Circuits by Binary Decision Programs,” Bell System Technical Journal, vol. 38, pp. 985-999, 1959.

[23] S. Ackers, “Binary Decision Diagrams”, IEEE Trans. on Computers, vol. C-27(6), pp. 509-516, 1978.

[24] T. Sasao, M. Fujita, “Representations of Discrete Functions”, Kluwer Academic Publishers, Boston, 1996.

[25] A. Aho, J. Hopcroft, J. Ullman,“Data Structures and Algorithms“,Addison-Wesley, Mass., USA, 1983.

[26] R. Rudell, Espresso Misc. Reference Manual Pages, 1993, http://embedded.eecs.berkeley.edu/pubs/downloads/espresso/index.htm

Using Shared Multi-Terminal Binary Decision Diagrams for Image Representation

Miloš M. Radmanović[1]

Abstract – **Compactness criteria for 2D image representations are important because of their increasing role in many computer applications. This paper discusses the use of shared multi-terminal binary decision diagrams (SMBDDs) to represent bitmapped image. In order to investigate and compare performance of the SMTBDD image representation, I have developed a specialized tool for building an SMTBDD from bitmapped image. I have demonstrated that SMTBDDs are an efficient representation for every special-case image.I do not purport to "prove" in any real sense that SMTBDDs are a superior representation of general images. This paper is not the end, but rather the beginning of my research.**

Keywords – **Boolean functions, shared multi-terminal binary decision diagram, image representation, software tool**

I. Introduction

Binary Decision Diagrams (BDDs) are a data structure that has been used for years to provide an efficient representation of Boolean functions. BDDs were introduced by Akers in 1959 [1]. In the early 1980's, Bryant [2] demonstrated how a BDD could be modified to become a canonical representation of a Boolean function. In 1990's, Coudert, and Madre [3] demonstrated that BDD could represent sets of finite-state machine states efficiently.

In [4], it is observed that for a BDD is generally thought to take only terminals 0 and 1.It is shown that a BDD can have integer terminals (Multi-terminal BDD - MTBDD). In the late 1990's, there are many explorations of the relationship between BDDs and matrices. Fujita, McGeer, and Yang [5] have demonstrated that MTBDDs are the space-optimal representation of dense, sparse,and permutation matrices. In [6], it is proposed a new compression scheme to reduce the huge size of the inverted files in a large information retrieval system without loosing the querying efficiency. The basic idea is to transform the inverted list into a logic function, to represent that function in the form of BDD, and then store the BDD directly in the inverted file.

In [7], it is shown how images can be compressed using Ordered Binary Decision Diagram (OBDD). The mechanism of sharing identical sub-OBDDs representing subimages is also useful when compressing related image sequences such as movies.However, the compression ratios obtained with this approach are low. In [8], it is presented a coding algorithm for OBDDs which provided better compression ratios. It has also been shown that performances of image operations on BDDs depend on the size of the BDDs [9].In [10], it is shown that the geometric transformations of an image represented by a BDD can be expressed using only BDD manipulation process.

In this paper, I discuss the use of Shared MTBDD (SMTBDD) [10] to represent images. I take ideas introduced in [5], [7], and [8], and extend them to SMTBDD. The basic idea of my approach is to represent the image to be coded with a Boolean function, and then simplify and code it efficiently with SMTBDD. In order to investigate and compare performance of SMTBDD image representation, I have developed a specialized tool for building SMTBDD from bitmapped image.

This paper is organized as follows: Section 2 shortly introduces the SMTBDDs. Section 3 describes SMTBDD bitmap image representation. Section 4 describes the software tool for building SMTBDDs from bitmap images and shows experimental analysis of SMTBDD representation for some benchmark bitmap images. I finally give some conclusions in section 5.

II. Shared Multi-Terminal BDD

Binary Decision Diagrams (BDDs) are data structure convenient for representation of discrete functions. Due to that, BDD have become widely used for a variety of CAD applications, including symbolic simulation, verification of combinational logic, and verification of sequential circuits, see for instance [11], [12].

BDDs are derived by the reduction of the corresponding binary decision trees (BDTs). The reduction is performed by sharing the isomorphic subtrees and deleting the redundant information in the BDT using the suitably defined reduction rules [2].

Multi-terminal binary DDs (MTBDDs) are generalization of BDDs. They can represent Boolean functions by the corresponding integer equivalent functions [11]. This technique is useful for various areas in computer science.

Multiple-output switching functions are represented by shared BDDs (SBDDs) or shared MTBDDs (SMTBDDs) having a separate root node for each output.

An example of SMTBDD for the multiple-output function with outputs f_0 and $f_1$is shown in the following figure.The truth vector with integer values of the function f_0 is F_0 = [2, 2, 4, 2, 4, 4, 4, 4]^T and of the function f_1 is F_1=[2, 2, 4, 2, 4, 3, 4, 4]^T.

It is obvious that this SMTBDD is quite smaller then two MTBDDs for the functions f_0 and f_1 in the number of nodes since there are shared values of the vectors of f_0 and f_1.This feature is essential in SMTBDD representations and will be highly exploited in SMTBDD representation of bitmapped image.

[1]Miloš M. Radmanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos.radmanovic@gmail.com

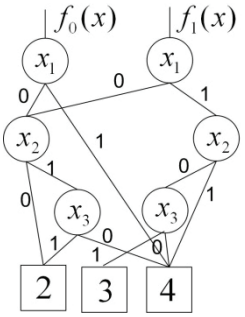


Fig. 1. An example of SMTBDD for the functions f_0 and f_1.

III. SMTBDD Bitmap Image Representation

Typically,the black and white bitmapped image can be representing as a matrix of size $n \times m$ whose entries correspond to pixels having either value 0 or 1.For gray-scaled bitmapped images, matrix entries correspond to pixel values between 0 and 255. For RGB color bitmapped images, R, G, and B matrices entries correspond to pixel values between 0 and 255.

To transform a bitmapped image to a BDD, it requirescreationof a set of Boolean function variables to each entry of the image matrix and consideration of each pixel asa minterm representation of the Boolean function. Thereafter, for image matrix of size $n \times m$,it is necessary $\lceil \log_2 n \rceil + \lceil \log_2 m \rceil$ variables [9]. Then, mintern representation of the Boolean function can be transform to BDD using BDD construction algorithms proposed by researchers [2], [3], [4]. An example of BDD representation for black and white bitmapped image is shown in the following figure.

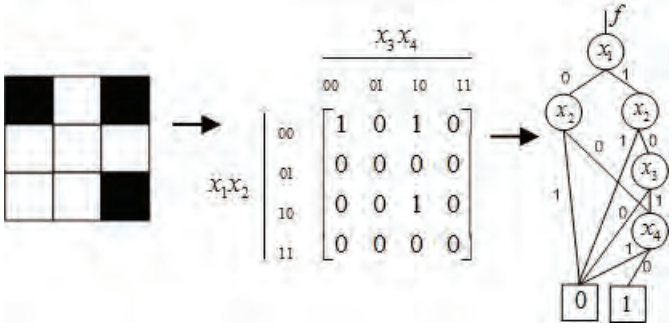


Fig. 2. An example of BDD representation for black and white bitmapped image of size 3×3

For gray-scaled bitmapped images, mintern representation of Boolean functions with integer values can be transform to MTBDD using MTBDD construction algorithms proposed by researchers [5]. An example of MTBDD representation for gray-scaled bitmapped image is shown in the following figure.

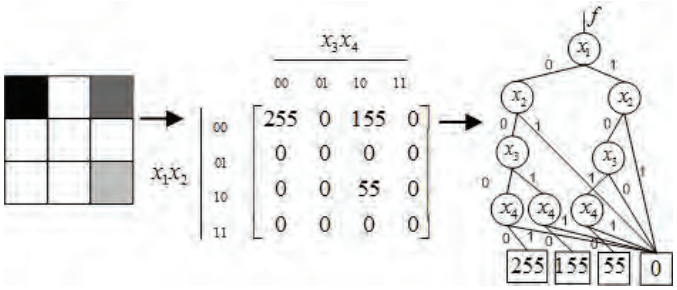


Fig. 3. An example of MTBDD representation for gray-scaled bitmapped image of size 3×3

For RGB color bitmapped images, I propose mintern representation of multi-output Boolean functions with integer values. This representation can be transform to SMTBDD using SMTBDD construction algorithms proposed by researchers [10]. An example of SMTBDD representation for RGB color bitmapped image is shown in the following figure.

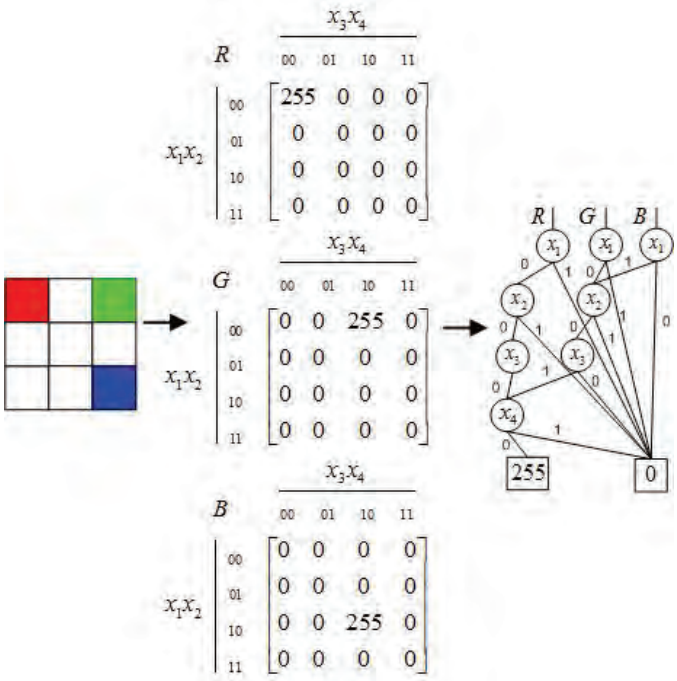


Fig. 4. An example of SMTBDD representation for RGB colorbitmapped image of size 3×3

It is obvious that SMTBDD representation for RGB color bitmapped image from previous example is quite smaller then three MTBDDs for the multi-output function with outputs R, G and Bin the number of nodes since there are shared values betweenR, G and B matrices.

It should be observed that SMTBDD from previous example is compact since there are many constant subvectors of four consecutive 0 in R, G and B matrices.

IV. Experimental Results

In order to investigate and compare performance of SMTBDD image representation, I have developed a specialized tool for building SMTBDD from bitmapped RGB color images.

Specialized software tool is written in MS Visual C++ and use MFC technology [13]. It consists of three basic modules: (1) BDD module for SMTBDD representation of Boolean functions [14], (2) Image transformation module for minterm representation of RGB color bitmapped images, and (3) Interaction module that allow user to interact with software tool.

Below I give a list of RGB color image benchmarksof various types. Benchmarksare based on the collection of famous image benchmarks freely available in the public domains[15], [16]. The benchmarks then have been categorized in four categories: benchmarks of detailed type (Figure 5), benchmarks of screen test type (Figure 6), benchmarks of text type (Figure 7) and, benchmarks of texture type (Figure 8). This allows judging the quality of SMTBDD image representations and gives a better overview of representation performance. Descriptions of benchmarks are given in table 1.

I performed the testing on a PC Pentium IV on 2,66 GHz with 4GB of RAM on MS Windows 7 platform. The memory usage for tool was limited to 2 GB, and space statistics of benchmarks is presented in table 1. All benchmarks are 24-bit RGB bitmapped images.

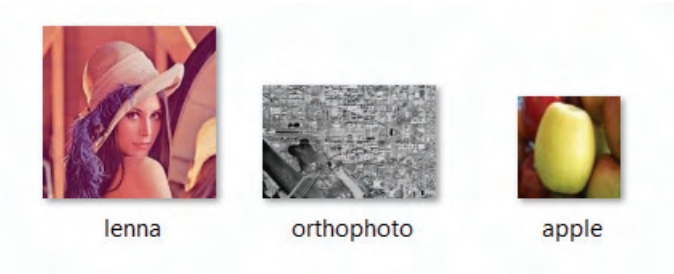


Fig. 5. Benchmarks - RGB color images of detailed type

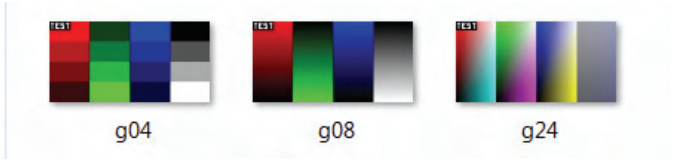


Fig. 6. Benchmarks - RGB color images of screen test type

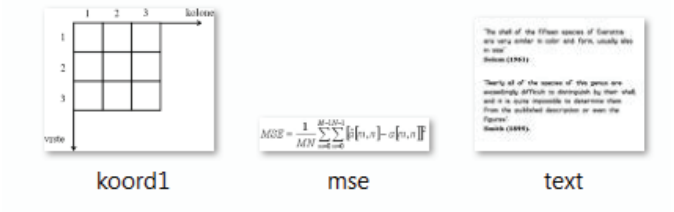


Fig. 7. Benchmarks - RGB color images of text type

Fig. 8. Benchmarks - RGB color images of texture type

TABLE I
SPACE STATISTIC OF SMTBDD REPRESENTATION FOR RGB COLOR BITMAPPED IMAGES

Image	Size [pixels]	Image size [elem.]	SMTBDD size [nodes]	Comp. ratio [%]
lenna	204 x 204	124848	73461	58
orthophoto	512 x 256	393216	86564	22
apple	64 x 64	12288	9731	79
g04	128 x 64	24576	411	1
g08	128 x 64	24576	2516	10
g24	128 x 64	24576	3109	12
koord1	389 x 324	378108	2168	0.5
mse	285 x 54	46170	1611	3
text	111 x 86	28638	4154	14
brain	249 x 203	151641	68143	45
textile	225 x 225	151875	108590	71
texture	259 x 194	150738	82314	55

Table 1 describes SMTBDD space statistics using proposed representation of RGB color bitmapped images. The third column of the table presents the total number of elements in three matrices (R, G and B) of the matrix representation for bitmapped images. The fourth column of the table presents total number of nodes in SMTBDD representation for bitmapped images. Differences in space performances between RGB matrices and SMTBDD image representation are emphasized by percentage values in the last column of the table.

According to percentage values of differences, in most cases for details and texture type images, it is shown that SMTBDD is not compact representation. But, in all cases for screen test and text type images, it is shown that SMTBDD is very efficient representation.

V. Conclusion

This paper describedthe use of SMTBDDs for compact representation of RGB color bitmapped images. In order to investigate performance of SMTBDD image representations, I have developed a specialized software tool for building SMTBDD from matrix representation of bitmapped image. I just report the fact that first experimentation with SMTBDD image representation is quite satisfactory. It opens the possibility to further research of SMTBDD representation for some special-case images. I do not purport to “prove” in any

real sense that SMTBDDs are a superior representation of general images. This paper is not the end, but rather the beginning of my research.

REFERENCES

[1] S. Ackers, “Binary Decision Diagrams”, IEEE Trans. on Computers, vol. C-27(6), pp. 509-516, 1978.

[2] R. Bryant, "Graph-based Algorithms for Boolean Function Manipulation", IEEE Trans. Computers, vol. C-35, pp. 667-691, 1986.

[3] O. Coudert and J.C. Madre, “A unified framework for the formal verification of sequential circuits,” IEEE Int. Conf. on CAD, 1990.

[4] E. Clarke, M. Fujita, K. McMillan, J. Yang, X. Zhao, “Spectral Transforms for Large Boolean Functions with applications to technology mapping”, Proc. of the 30th Int. Design Automation Conf., 1993.

[5] M. Fujita, P. McGeer, J. Jang, “Multi-Terminal Binary Decision Diagrams: An Efficient Data Structure for Matrix Representation”, Journal of Formal Methods in System Design”, vol. 10, is. 2-3, 1997.

[6] C. Lai, T. Chen, “Compressing Inverted Files in Scalable Information Systems by Binary Decision Diagram Encoding”, ACM/IEEE 2001 Conference, 2001.

[7] M. Starkey, R. E. Bryant,”Using Ordered Binary Decision Diagrams for Compressing Images and Image Sequences” Technical Report, CMU-CS-95-105, CarnegieMellonUniversity, 1995.

[8] P. Mateu-Villarroya, J. Prades-Nebot, “Sequential Logic Compression of Images”, Proc. of IEEE Int. Conf. on Image Processing (ICIP'01), Tessaloniki, Greece, pp. 479-482, 2001.

[9] C. Lursinap, K. Kanchanasut, T. Siriboon, "Basic Binary Decision Diagram Operation for Image Processing" Advances in Computer Science - ASIAN'97, Lecture Notes in Computer Science, 1997, vol. 1345, pp. 368-370, 1997.

[10] L. Robert, G. Malandain, "Fast Binary Image Processing Using Binary Decision Diagrams", Computer Vision and Image Understanding, Elsevier, vol. 72-1, pp. 1-9, 1998.

[11] T. Sasao, M. Fujita, “Representations of Discrete Functions”, Kluwer Academic Publishers, Boston, 1996.

[12] G. Hachtel. F. Somenzi, Logic Synthesis and Verification Algorithms, Kluwer Academic Publishers, 1996.

[13] Microsof Visual Studio 6.0, Microsoft Corporation, http://msdn.microsoft.com/en-us/library/ms950417.aspx

[14] Stgroup CIITLAB Software, http://stgroup.elfak.edu.rs/home/node/4

[15] Computer Vision Test Images, http://www.cs.cmu.edu/~cil/v-images.html

[16] BMP Suite, http://entropymine.com/jason/bmpsuite/

Facebook as a Learning Platform

Enes Sukić[1], Miljana Maksimović[2], Leonid Stoimenov[3]

Abstract – **In this paper we analyzed the need to introduce new, more modern approaches to traditional education. Sudden developments of interactive services that are specifically integrated into the social networks, especially on Facebook, represent an inexhaustible opportunity to use them in educational purposes. The paper analyzed the possibilities of using Facebook as an educational platform, through a concrete example of creating a separate group related to the course of Computing and Information technology. Emphasis is placed on the huge potential of educational applications, including complete e-learning systems, which are free to use on Facebook. As Facebook is naturally open for student population, educators must not miss opportunity and not use this extraordinary integrated system in the service of knowledge transfer.**

Keywords – **Education, Facebook, E-learning, Web 2.0.**

I. Introduction

Exceptionally fast development of information and communication technologies and their impact on all spheres of life have caused various changes in approaches even when it comes to the traditional systems such as education. In recent years, we have witnessed how the Internet has become more interactive. Services based on Web 2.0 technologies have contributed communication to become versatile and easier to use, so that it is easy to say that the Internet has become the primary means of communication. There are many different products of this type of Internet such as social networks, blogs, forums, video and other services, which can be used for the purpose of transferring knowledge and online learning.

Modern concepts of learning require more open communication between learners and teachers, which is very simple if communication takes place through social networks. The greatest potential for these purposes has Facebook, because of largely familiar environment. As most of the students visit social network sites almost daily, many educational institutions recognized the opportunity to take advantage of this free resource through advertising and software packages for e-learning that are designed to function on Facebook. This gives students special opportunity to have an informal entertaining interest and follow their school work and events, all without leaving their favorite social network.

[1]Enes Sukić is PhD student at the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: esukic@yahoo.com

[2]Miljana Maksimović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: miljana.maksimovic@elfak.ni.ac.rs

[3]Leonid V. Stoimenov is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: leonid.stoimenov@elfak.ni.ac.rs

II. Related Work

Permanent migration of people spread knowledge beyond classrooms, but Internet is increasing the knowledge without movement of people and products. Distance learning makes it easier for learners and teachers to communicate and share knowledge. The main idea is that learners need computer and access to learning platform, but not without "adequate professional and technological methods, quality and organizational standards" [1].

E-learning, whether through applications based on LMS (Learning Management System) or CMS (Content Management System) e-learning systems or online learning through specific services such as Moodle [2], greatly contributes to the quality of education. Social networks bring together a huge number of users who try to connect in smaller online communities around common interests [3]. Students in a course or within a department are very much interested to exchange knowledge, experience, materials, and to informally socialize within their interest groups. Professors are also interested to use such communities to achieve communication and help students master materials more easily and acquire knowledge more quickly in course they attend. Learning platform provides videos and lectures are available to learners at any time; collaboration between learners is increased and learning flow of each learner is accompanied by the teacher.

Social networks are also designed to ensure privacy of information that is placed within a group. Teacher controls access to the group. A group must be closed in order to function properly so that all the attention of group members is directed toward the goals of education. Creating such a space for learning does not mean ignoring the traditional approach to learning in terms of classroom or laboratory, which is still irreplaceable. This virtual space should be accepted as a supplement and an excellent opportunity for additional work that contributes to a faster and easier learning and mastering the course materials.

Hamann and Wilson [4] found that students who participated in a web-enhanced class outperformed those students in a traditional lecture format. This suggests that Internet based learning modules actively engage students in a manner unique from the traditional class lecture [5].

Within Web social networks, such as Facebook, MySpace or Ning [6], it is possible to integrate applications related to e-learning. Applications can be adapted to the course to which the group is attached, and in most cases it enables uploading and downloading of materials, various announcements, and ability to communicate in real time, and video section where teachers can post their lectures or additional materials related to the topics of the course. In social networks, within the Interest groups, these relationships and their mutual interactions are important:

- Teacher-student,
- Student-student,
- Student-teacher and
- Teacher-lecturer, in the same course.

Another advantage of using social networks for education is the possibility that students who are away at some point while studying can use social networks as a distance learning platform and follow the course material, while being absent from school [7]. This ability to access materials and more open and direct access to teachers give chance for less active and shy students to freely communicate with the teacher by asking questions directly through the email system integrated into the social network or direct chat, in the moments when this type of communication is allowed by the teacher.

With the appearance of first Web social networks, such as sixdegrees.com and livejournal.com [8], educators realized the opportunity to create professional profiles related to their course, where they would provide information and post materials. Even then, there were some important rules for using social networks in educational purposes:

- Teachers regulate who can access the group,
- The group must be professional, closely related to the field of teaching, with no unnecessary elements,
- Teachers must communicate with students in a professional way,
- Teachers must control posted materials, especially if the students are allowed to upload materials within a particular section.

III. Facebook in Education

Among the many social networks people can have access to, Facebook is probably one of the most popular one. It connects people with friends, relatives and associates by simply creating a Facebook account. Users can create their own profiles attaching any information of their interest [9]. This social network includes easy-to-use applications such as uploading videos, pictures, files while users get in touch with others in terms of friendship, business or pleasure. To create an account in Facebook is completely free. A person only needs a valid e-mail account, and fill out the Facebook registration page. Once the account is established, the information can be updated as desired [9].

We can see that Facebook has become the main channel of communication among youth and elderly population. First of all because it is free and because the constant improvement and addition of new services [10], then ease of use and accessibility from all known platforms, from mobile phones and other devices, the number has been increasing incredibly fast. Statistics show that Facebook has over 500 million registered users which highly contributed to its worldwide recognition as a brand in all areas of activity. This is why all major companies, associations, interest groups, including educational institutions use Facebook to present themselves and even show their activities on this social network.

Facebook has managed to integrate the most interesting Web 2.0 services that exist and are used separately on the Internet, and now receive much more importance when collected in one place. Out of these services and features, these are the most significant ones:

- The Wall, which is actually a modern blog, with possibility to comment contents.
- The ability to upload images and place comments.
- Upload of video and audio materials, and upload of existing video materials from popular video services, YouTube being the most important one.
- The option to create separate groups within the profile, and their administration.
- The ability to use various applications within Facebook.
- Using the internal mail system.
- Using advanced chat system.

The mere announcement that Facebook made about integrating Skype system completes this platform and makes it unique on the Internet [11].

This enormous abundance of services is irresistibly attractive for use in educational purposes. As we said earlier, the best option for teacher is to create a separate group for his course, which will be separate from his private information and contacts. The teacher will then inform his students and individuals interested to join such a group. Facebook offers all administrative features, through which teacher regulates who can access the group and what type of materials and comments students may post. Teacher can use Wall, Email or chat system, and students can use all resources for mutual communication and cooperation, which is of great importance for progress in learning [12]. They can also make appropriate comments; ask questions, and make suggestions about all the materials that the teacher has uploaded. This is why this contributes to the quality of interaction between teachers and students.

IV. Facebook as a Learning Platform

In order to examine issue which approve that Facebook is a good e-learning tool for IT students, a group for Computing and IT course has been created as closed group, which means that only members can see and add posts. The main page of the group is presented on Fig. 1. This group could be secret, which means that only members can see the group. In that case, administrators need to send invitation to every single student. Problems appear when they find out that there are two (or even more) people who had the same name and last name. For security reasons, it is better to allow users of Facebook to see the group and send invitation to join it. Given that Facebook has the ability to display to all students friends information that he has accessed a particular group, there have been few unauthorized requests to join the group that the teacher rejected in order to preserve privacy of this group and the pursuit of its primary objective and that is to assist target group in overcoming the material and raise education to a higher level. It is clear that even with a small group of students, better results in knowledge transfer can be achieved.

Fig. 1. The main page of the group

The Wall was used to post announcements, audio-video materials or links important for the course. There were some links which lead to additional resources, on Facebook and outside of it. Students have been invited to send messages and ask questions if any. The teacher has sent individual and collective mails to the students of the group to answer the question or to tell students about the dates when he will be online and when he can talk directly with them through internal chat system.

Facebook as social software platform contributes to e-learning through applications for a wide variety of academic purposes. These applications could be used and shared by students, teachers and librarians, to assist them in learning and administering.

Important part of learning platform is section where teachers can post their lectures or additional materials related to the topics of the course. There are a lot of Facebook applications on this field, but the most popular are YouTube Video Box and SlideShare. YouTube Video Box is tool for sharing videos on Facebook from an existing YouTube.com account. It is shown on Fig. 2. SlideShare (shown on Fig. 3.) is tool for sharing presentations from existing SlideShare.net account. In this way, teachers can ensure that lectures are held by prominent individuals, which allow students to follow world trends and tendencies.

Fig. 2. YouTube Video Box tool

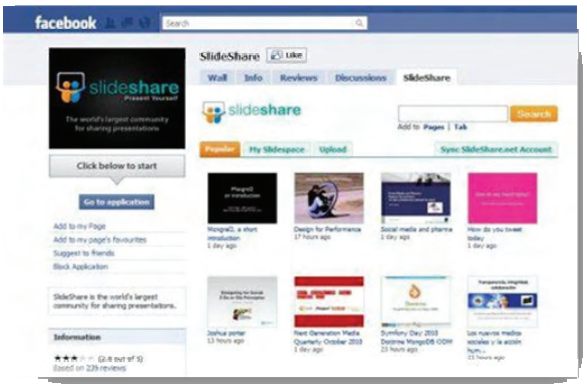


Fig. 3. SlideShare tool

To analyze the effectiveness of increasing education quality with the use of groups on Facebook, we monitored and surveyed 34 students. The goal was to make approve that Facebook could be successfully used to master the materials faster and in a more detailed manner by students. Also, through survey, the personal opinion and experience of students were presented.

Surveyed included a several yes/no questions. Some of them are:

Q1: Do you download and share the documents, video and audio material that the professor uploads into the group?
Q2: Do you think this group made studying easier?

On these questions all students answer “Yes”. This means that students are interested in this kind of learning. They have pointed out the importance of sharing general information, such as mid-term and exam grades, changes in the schedule and other info that the professor shares on the wall or sends via group email.

Online discussion helps students to exchange opinions and thereby eliminate the uncertainties about the material that can eventually occur. Usage of group to discuss, chat with professor and as a primary source of learning is shown on Fig. 4. Even 64.7% of the students actively participate in discussions on the group. Activity of the group can be valued equally as the activity in class, as an active student publicly expresses his opinions and showing interest in the material. Also, it activates the other students to discuss.

Results of monitoring show that 23.5% of the students chat with a professor. 29.4% of the students use the group as primary source of learning. Such a small engagement may be justified by the fact that this type of learning coupled with traditional classroom learning, with direct communication with the professor.

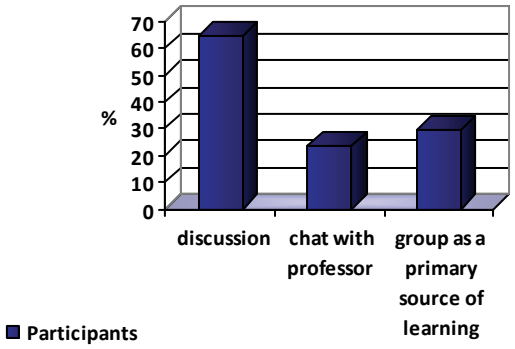


Fig. 1. Usage of group to discuss, chat with professor and as a primary source of learning

These results indicate remarkable advantages in terms of improving quality of knowledge transfer, in a completely new, more interactive and interesting way. In fact, the education policy should implement use of such resources that social networks like Facebook offer.

V. CONCLUSION

It is noticeable that the Internet through the development of Web 2.0 services is becoming more sociable and more communication-oriented media. The development of social networks, especially Facebook and its integration of many popular services in one place, led to the expansion of ideas of using this virtual space. Educators of all profiles have seen their chance, that in addition to traditional means of knowledge transfer, they can also use Facebook , at least, as a complement and easier way of uploading information and materials related to the course they teach in a modern and highly accessible way.

The authors of this text are convinced that the traditional ways of knowledge transfer remain dominant, but with a great need to implement new technologies, where Facebook has a great potential and power. Through excellent multimedia capabilities offered by Facebook and the possibility of creating special closed groups, it is possible to develop and use specialized e-learning applications that are integrated into the very social network. The authors believe that this new approach of using Facebook for educational purposes has great potential and that only in the years to come it will experience the full expansion.

ACKNOWLEDGEMENT

Research presented in this paper was partially funded by the Ministry of Science of the Republic of Serbia, within the project "Technology Enhanced Learning", No. III 47003.

REFERENCES

[1] F. Priem, R. De Craemer, J. Calu, F. Pedreschi, T. Zimmer, S. Saighi, J. Lilja “E-Learning in Science and Technology via a Common Learning Platform in a Lifelong Learning Project”, Eurodl, 2011.

[2] S. Hargadon, **“**Moodle: An Open Learning Content Management System for Schools”, CoSN K12 Open Technologies, 2008.

[3] A. Webster, “Visible Relations in Online Communities: Modeling and Using Social Networks”, University of Saskatchewan, Saskatoon, 2007.

[4] K. Hamann, B. M. Wilson, “Beyond search engines: Enhancing active learning using the internet”, Politics & Policy, 2003.

[5] C. Muñoz, T. Towner, “Opening Facebook: How to Use Facebook in the College Classroom, Society for Information Technology and Teacher Education conference”, SITE, Conference Proceedings, pp. 2623-2627, Charleston, SC, USA, 2009.

[6] J. C. Dunlap, P. R. Lowenthal, “Learning, unlearning, and relearning: Using Web 2.0 technologies to support the development of lifelong learning skills”, E-Infrastructures and Technologies for Lifelong Learning: Next Generation Environments, IGI Global, pp. 292-315, 2009.

[7] F. Baltar, I. Brunet, “The Use Of Facebook In Social Research.The Virtual Snowball Method Applied To The Study Of Immigrant Entrepreneurs In Spain”, INTED2011, Spain, 2011.

[8] C. C. García, R. Lago, ”Web 2.0: Innovative Experience In University Education Through Social Networks”, INTED2011 Proceedings, pp. 6320-6324, Spain, 2011.

[9] G.G. Quijano Zavala, R. Ferrer Méndez, R. A. May Melendez, **“**The Use Of Facebook To Develop Social Written Interaction Among Freshmen Students In English”, Inted2011, Spain, 2011.

[10] A. Szwelnik, “Embracing the Web 2.0 culture in Business Education – the new face of Facebook”, Oxford Brookes University, 2008.

[11] M. Shiels, “Facebook and Skype deal to dial friends and family”, BBC News, Silicon Valley, 2011.

[12] C. Reeve, “Social Media in Education”, 2010. http://playwithlearning.com/2010/07/01/social-media-in-education-part-1/

Data Mining on University Database

Jasmin Ramadani[1],Sime Arsenovski[2], Ruben Nuredini[3], Zoran Gacovski[4]

Abstract – **The paper is focusing on using Data Mining technologieson University Database of student data for extracting knowledge to improve the management of enrolment of new students. Using the data of current students placed in Oracle Database, the goal is to use Oracle Data Mining to classify the students in order to predictof distribution of the students based on their characteristics that will help to recognize better the target groups of future students and improve the enrolment process.**

Keywords –**Data Mining, Students, Oracle, ODM, Algorithm.**

I. INTRODUCTION

The enormous use of electronic databases on the higher education institutions has brought to large amount of data that are stored and not converted to valuable knowledge. The data is hiding the knowledge which means that we have to use some techniques to extract and use that information. The best suited technique for this purpose is the Data Mining. It can be defined as an analysis of large amounts of data in order to find undiscovered relations and to present them on a new way that they will be understandable and useful [1].Some authors are using the term Data Mining as a synonym for Knowledge Discoverywhile others consider the Data Mining as a part from that process. The process of Data Mining can be used to find new information hidden in the databases which can be used to support many issues of the educational process. In the paper we will describe the process of Data mining on Oracle database of student records in order to extract valuable knowledge.

II. KNOWLEDGE DISCOVERY PROCESS

As central part of the process of Knowledge Discovery are the methods of data mining but there are also other parts that are important. The process (Fig.1) can be simply represented as a composition of three main elements [2]:

1. Preparation of data
2. Algorithm of Data Mining
3. Analysis of the data

[1]Jasmin Ramadani is with the Faculty of Information and Communication Technologies at FONUniversity b.b, 1000 Skopje, Macedonia,E-mail: jasmin.ramadani@fon.edu.mk

[2]Prof. Dr. Sime Arsenovski is with the Faculty of Information and Communication Technologies at FONUniversity b.b, 1000 Skopje, Macedonia, E-mail: sime.arsenovski@fon.edu.mk

[3]Ruben Nuredini is with the Faculty of Information and Communication Technologies at FONUniversity b.b, 1000 Skopje, Macedonia, E-mail: ruben.nuredini@fon.edu.mk

[4]Zoran Gacovski is with the Faculty of Information and Communication Technologies at FON University b.b, 1000 Skopje, Macedonia, E-mail: zgacovski@yahoo.com

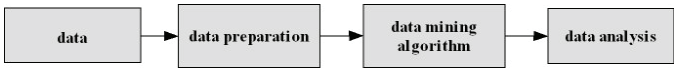


Fig. 1 The process of Knowledge Discovery

On the beginning of the process we must deal with the data, its form and condition. Based on that, we will prepare the data for the process of Data Mining. This is eminent if there are some irregularities in the database. After the Data Mining is done the data should be analyzed so it can be used and properly presented. The process of Knowledge Discovery is an interactive process where every step has it own importance although the step of Data Mining is the most important.

III. TECHNIQUES

There are various techniques of Data Mining that can be used to find different patterns in the data. The number of techniques can vary, but we can identify five main techniques of Data Mining [3]:

- Classification
- Prediction
- Clustering
- Association
- Summarization

In the paper we will use the classification technique to get the probabilities for the distribution of the students.The classification is a very frequent technique of Data Mining and it checks the characteristics of a database object after which it sets the object in a class which was defined previous. The classification organizes the data in classes on basis of certain attributes. The goal is to make a model that could be used on unclassified datawhich takes the attributes as input and it gives the class as output. The number of classes can be two, n but also we can have many classes.

IV. ALGORITHMS

There are many algorithms that can be used in the process of Data mining. What algorithm we will use it depends from the technique that we use, the conditions and the technology we use. For the technique of classification most used algorithms are Decision Trees and Naïve Bayes.

For creation of the model of Data Mining in this paper we use the technique of classification which is using the "Naïve Bayes" algorithm.For the Naïve Bayes algorithm we are making classification using the Bayesian classification which is statistic classification of data [5]. We take the calculation of the possibility that one object belongs to one of the classes [6]. To calculate the possibility we use the Bayesian theorem:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (1)$$

where *P(X)* is the probability that the event will happen, *P(Y)* is the probability for the event Y while *P(X|Y)* is the prior probability for the event Xunder condition that the event Y has happened and *P(Y|X)* is the posterior probability of X conditioned by Y.

V. Oracle Data Mining

The Oracle Data Mining (ODM) is very capable tool for the process of Data Mining. The software is part of the Oracle Database Enterprise Edition. The ODM package is constructed of two parts [8]:

- Data Mining API
- Data Mining Server

The first part contains java classes and methods for creatingmodels of Data Mining. The second part is server side component which is making the Data Mining.

ODM supports the most used techniques of Data Mining. It is chosen to be used for the process of Data Mining in this casebecause it has many advantages. The complete process is done inside the Oracle Database which means that there is no need to export the data in to other environments. This is a very important advantage because in this way the process is more secure, more stable and the number of errors is smaller. It is very important to keep the database in the system which will remove the possibility of error and inconsistence which can appear if the database is moved. Also ODM is avoiding the possibility of using old data which is nit update because it uses the current database condition.The complete preparation of the data like cleaning and transforming is done with the ODM tools.Another great advantage is that if we use ODM, the models are staying in the Oracle database.

The Oracle Data Miner is another part of ODM which is the graphic interface that enables the visual creation of the process of Data Miningand it gives access to all the activities and functions.

VI. Data Mining Integration in Higher Education

The higher education institutions are storing large quantities of data about the students.This means that the institutions are storing data that are strategic resourcethat can be used for improving of the quality of different processes in the institution [9].The Data Mining as a process is well used in other areas like banking, trade, insurance where the data amount is very large. This means that we can try to implement the Data Mining in the educational system. In order to extract the hidden knowledge for the database we will use the techniques of Data Mining. The knowledge found can be used to support decisions and resolve problems in the management of the institution like the increasing of the efficiency, marketing decisions, enrolment of new students etc. In our case we want to find information that will help the University to get some information from its own database about the characteristics of the students on several campuses that will lead to improvements.

VII. Goal of the Research

The universities and faculties are defining different strategies they will use in the process of enrolment of new students. The higher education institutions could use the data collected when the student enroll to find some relations and knowledge that can predict some directions in the next year of enrolment. The data collected contains personal and demographic data like gender, place of living, department of studying, faculty etc. This kind ofinformation can help the institution to concentrate the efforts of marketing using the demographic factors. The classification of the students based on the above mentioned characteristics can group and predict the future student behavior. In this paper we have chosen to define the grouping of the students on the faculties of Detectives and Security and the Faculty of Economics based on the campus location. This could help the University to predict what percent of the students are going to enroll on different campuses on the mentioned faculties. The faculties are chosen based on their greater popularity among all the faculties on the University. We will try to build the models of data mining using the following students attributes in the database:

- Campus
- Faculty

Where the attribute campus contains the city where the campus is located and the attribute faculty is showing the name of the faculties. We will try to show any tends about that how the students are divided on base of the location of the campus on particular faculty. We will use the Data Mining to find the trends and to predict the distribution of the students. The benefits of the research will be the information we will get that will show the management of the institution about the possible interests of the students to enroll on particular students on the different campuses. According to that information the management can decide either to maintain or close some faculties on different campuses and to find out which faculties need more marketing on the campuses where there is lack of interest,

VIII. Modelling

After resolving the problem we will choose this case is concerning the possible management of the enrolment strategybased on the data enrolled students in 2009 on the FON University in Skopje. The data is stored in Oracle Database where with ODM we will try to show the information which was extracted.For the model we will use the attributes faculty and campus which are stored in the database of the students.

The data is prepared where we are treating the missing values in the tables of the database and other irregularities such the duplicate entries in some attributes like the name of the Faculties where it was found that there are different names for the same faculty in the student database. This needs correction of the duplicates in order to avoid the errors in the results.

After this step we choose the model of Data Mining that will be used. ODM gives support for several techniques and algorithms.After the analysis of the data and the possible outcomes it is chosen to use the model of classification using the Naïve Bayes algorithm.

The model of Data Mining is created in Oracle Data Miner 11g Release 2 which can be found in the Oracle Sql Developer 3.0.

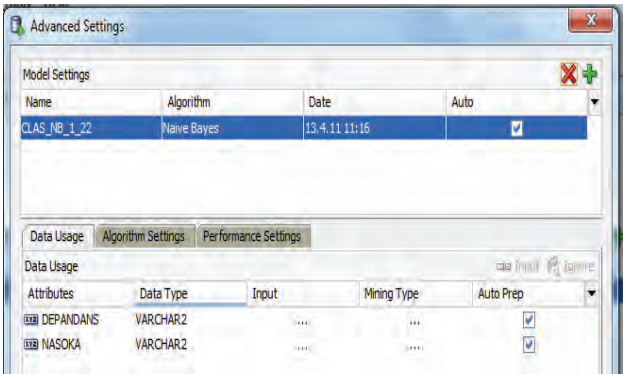


Fig. 2 Settings of the Data Mining Model

The result will show is it possible to group the students on basis of the city where the campus is placedand they study and the faculty that they will enroll to study. The prediction can be made using the attributes and the data from the current student database. The model should show what faculties are popular on different campuses of the University.

This information can help us to predict the distribution of the students on different faculties. As an example we will show how the students on the Faculty of Detectives and Security (Table1) and the Faculty of Economics (Table2) are dispersed on the different campuses in Skopje, Struga, Strumica and Gostivar.

TABLE 1.PROBABILITIES FOR THE FACULTY OF DETECTIVES AND SECURITY

Attribute	Value	Probability
Campus	Skopje	61,07784431
Campus	Gostivar	20,95808383
Campus	Strumica	12,57485030
Campus	Struga	5,38922156

The results in Table 1 show that there is high probability of about 61 percent that the future students of the Faculty of Detectives and Security will study on the campus in Skopje. That means that more than the half of all students of this faculty will be studying in the campus in Skopje. The smallest percent is found for the campus in Struga which shows that on this campus we cannot expect significant interest.This means that there is place for actions that should be taken by the management to increase the number of enrolled students of this faculty on the other campuses of the University

TABLE 2. PROBABILITIES FOR THE FACULTY OF ECONOMICS

Attribute	Value	Probability
Campus	Skopje	79.33333333
Campus	Struga	10.00000000
Campus	Gostivar	8.000000000
Campus	Strumica	2.666666667

From the results in the Table2 for the Faculty of Economics we can see that the biggest percent of probability found for the campus in Skopje. This shows that according to the technique of classification and the algorithm used that most for the students on the Faculty of Economics, about 79 percent will be studying on the campus un Skopje. In contrary there is very small percent which is about 2.66 percent that there will be students on the campus in Strumica. The information can affect the of the process of enrollment by concentrating the resources on the campuses where there interest about the particular faculty is smaller

IX. CONCLUSION

The process of Data Mining can help the Higher education institutions to use the knowledge hidden in the data they store every year about their students. The data does not show the possible all the relations and information that can help us to find valuables information and recognize future trends. In our case in this paper the Data Mining model shows some information about that how thecharacteristics of the students are connected with the city of the campus where they study and the faculty they are attending. Based on the current data using the model of classification in the Data Mining we can classify the students and predict the future distribution of the students on the campuses divided by the faculty. As an output we have the percent of students on every campus for one faculty.The information can help the University to determine which faculties on different campuses are popular and which are not. This could improve the process of marketing and enrolment of new students showing in which places there is need for more activities in order to increase the number of students. The benefit of the use of Data Mining for the purposes of improvement of the enrollment process will be in the information we will get which shows the trends of the future of the distribution of the students on the different campuses. This will show the justification of the existence of some faculties on different campuses and the need for greater marketing efforts and improvements to increase the popularity of different faculties.

REFERENCES

[1] D. Hand, H. Mannila, P. Smyth, “Principles of Data Mining”, MIT Press, 2001.

[2] S. Sumathi, S.N. Sivanandam,"Introducтion to Data Mining and its Applications", SpringerVerlag,2006
[3] H.H. Hsu,"Technologies in Bioinformatics", IDEA Group Publishing,2006.
[4] R. Agrawal, T. Imielinski and A. Swami, "Mining as sociation rules between sets of items in large data bases", in Proceedings of ACM-SIGMOD Conference, Washington, DC, 1993.
[5] R. Hanson, J. Stutz, P. Cheeseman, "Bayesian classification theory", Technical report, 1991.
[6] M. Pretzer, "Clustering und Klassifikation", Oldenburg Universitat, 2003.
[7] J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufman Publishers, 2001.
[8] K. Hauke, M.L. Owoc, M. Pondel, "Building Data Mining Models in the Oracle 9i Environment", Wroclaw University of Economics, Poland, 2003.
[9] N. Delavari,"Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System", Faculty of Information Technology, Multimedia University (MMU), Cyberjaya, Malaysia.

A Model of Vehicle Routing Problem with Soft Time Windows and Variable Traveling Time

Daniel Vatov[1] and Krasimira Genova[1]

***Abstract* – This article extends the generalized vehicle routing problem model by introducing variable traveling time. An overview is given for the different approaches to formulate a model and a set of criteria is defined to choose the proper one. Based on the analysis a model is chosen and modified to fit the requirements of the real life problems. Analysis of the most promising algorithm families for solving the model is presented.**

***Keywords* – Vehicle routing problem, Multiple time windows, Variable traveling times.**

I. INTRODUCTION

The vehicle routing problems (VRP) are widely spread combinatorial optimization problems. They appear in practice during the solution of different logistics problems in different areas of everyday life. In the basic formulation of the problem a set of identical vehicles are based in a central depot. Vehicles have to be routed optimally to supply customers with known demands. The vehicles have known capacity and traveling costs. Each customer is a part of only one route and is visited only once by the vehicle serving the route. VRP is a NP-hard problem [24], since it is generalization of the well-known TSP problem where the number of vehicles is only one. VRP and its variants is applicable in many areas of social and economic life like postal services, bank deliveries, air cargo transportation, school bus routing and many others. This problem is relevant to all logistic services since one of the goals of modern society is reduction of the consumption of unrecoverable sources of energy and diminishing the environmental pollution.

VRP was first described by Dantzig and Ramser [10]. The problem was a real-world application concerning the delivery of gasoline to service stations. They proposed mathematical programming formulation and an algorithm to get near-optimal solution of a problem instance with twelve service stations and four trucks. After the publication of their paper the research in this area flourished.

Based on the practical needs different formulations of VRP appeared. On Fig. 1 some of the major sub-problems with the relations between them are given. The descriptions of the sub-problems below are based on the following model: Let $G=(V,A)$ be a complete graph where $V=0,\rightleftharpoons,n$ is the set of all vertices and A be the set of all arcs. Vertices $i=1,\rightleftharpoons,n$ correspond to the customers, whereas vertex 0 corresponds to the depot. A nonnegative cost c_{ij} is associated to each arc $(i,j)\in A$ and is the travel cost spent to go from customer i to customer j. Generally loop arcs (i,i) are not allowed and this is imposed by defining $c_{ii}=\infty$. Each customer i has an associated demand $d_i>0$ to be delivered. The depot has a fictitious demand $d_0=0$. A set of K identical vehicles each with capacity C start to service a customer from the depot. To assure that there is a feasible solution it is assumed that $d_i=C;\quad i=1,\rightleftharpoons,n$. Each vehicle may perform only one route and it is assumed that K is not smaller than K_{min} where K_{min} is the minimum number of vehicles needed to serve all the customers. This minimum number can be determined by solving the Bin Packing Problem (BPP) derived from the VRP model. BPP is a NP-hard problem but the problem instances with hundreds of elements can be effectively solved to optimality [23]. All VRP sub-problems have to find K simple circuits, each circuit corresponding to a route, with minimum cost. The cost of a route is defined as the sum of costs of all arcs part of the route. A route should visit the depot vertex and each customer vertex is a part of exactly one route. Each of the sub-problems defined below extends this definition by adding additional constraints.

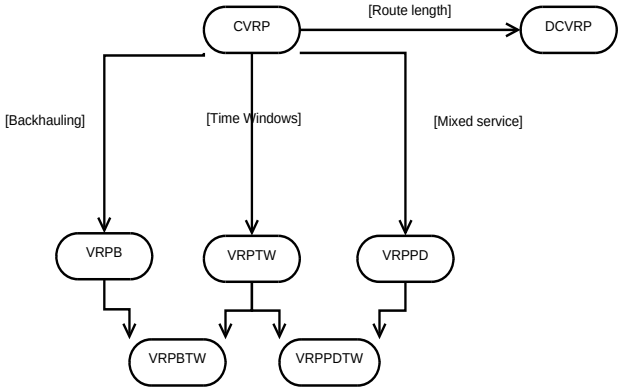


Figure 1. Major sub-problems of VRP and their relations [24]

- Capacitated VRP (CVRP) - to the basic model described above it is added that demands of the customers on the same route should not be greater than C;
- Distance-Constrained VRP (DCVRP) - based on the basic model where each route length or route duration should be less than a limit. There is no capacity constraint;
- VRP with Backhauls (VRPB) - extension of CVRP where customers set $V\setminus\{0\}$ is partitioned in two subsets. The first subset L contains l linehaul customers each requiring a given amount of products to be delivered. The second subset B contains n-l backhaul customers where a given quantity of inbound products must be picked up. Customers are numbered so that $L=\{1,\rightleftharpoons,l\}$ and

[1]Daniel Vatov, Email: daniel.vatov@gmail.com and Krasimira Genova, Email: kgenova@iinf.bas.bg are with Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev St., Block 2 1113 - Sofia, Bulgaria.

$B = \{l+1, \rightleftharpoons, n\}$. In the VRPB there exists a precedence between linehaul and backhaul customers. If a route serves both types of customers linehaul customers should be served first. The total demand of linehaul and backhaul customers does not exceed, separately, the vehicle capacity *C*;

- VRP with Time Windows (VRPTW) - extension of CVRP where to each customer *i* a time interval $[a_i, b_i]$ called time window is associated. Customers can be served only in the associated time window. The vehicle stops at each customer on the route for s_i time instants. VRPTW is generalization of CVRP, when $a_i = 0$ and $b_i = +\infty$ for each $i \in V \setminus \{0\}$;
- VRP with Pickup and Delivery (VRPPD) - extension of CVRP where to each customer *i* two quantities d_i and p_i are associated representing the demand of homogeneous commodities to be delivered and picked up at customer *i*. Vehicles may pickup commodities from an origin customer *i* and deliver them to a destination customer *j*. The current load of the vehicle along the route must be nonnegative and may never exceed the vehicle capacity *C*. All customers part of an origin-destination relation should be part of the same route. There is also a precedence as origin customers should be served before destination customers;
- VRP with Backhauls and Time Windows (VRPBTW) - a combination between VRPB and VRPTW;
- VRP with Pickup and Delivery and Time Windows (VRPPDTW) - a combination between VRPPD and VRPTW;

There are three main approaches that can be found in literature [22, 24] to formulate a mathematical programming model for the basic VRPs presented above. The first type of models are known as vehicle flow formulations. They use integer variables, associated with each arc or edge of the graph which counts the number of times the arc or edge is traversed by the vehicle. These models do not keep track of the commodities delivered at each customer. Further they can be subdivided to two-index vehicle flow formulations, in which the variables are indexed according to the start and end vertices only, and the three-index vehicle flow formulations, in which the variables have a third index which distinguishes one vehicle from another. These models are more frequently used for the basic versions of VRP. They are particularly suited for cases in which the cost of the solution can be expressed as the sum of the costs associated with the arcs, and when the most relevant constraints concern the direct transition between the customers within the route, so they can be effectively modelled by an appropriate definition of the arc set and of the arc costs. On the other hand, vehicle flow models cannot be used to handle many practical issues, e.g., when the cost of a solution depends on the overall vertex sequence or on the type of the vehicle assigned to a route. Moreover, the linear programming relaxation of vehicle flow models can be very weak when the additional operational constraints are tight.

The second family of models is based on the so-called commodity flow formulation. In this type of models, additional integer variables are associated with the arcs or edges and represent the flow of the commodities along the paths travelled by the vehicles. Vehicle flow models have an exponential number of constraints to enforce connectivity while commodity flow models impose this requirement by using a set of continuous variables representing the flow of one or more commodities between the depot and the customers. They were introduced by Garvin et al. [14] and later extended by Gavish and Graves [15, 16].

The models of the last type have an exponential number of binary variables, each associated with a different feasible circuit. The VRP is then formulated as a Set-Partitioning Problem (SPP) calling for the determination of a collection of circuits with a minimum cost, which serves each customer once and, possibly, satisfies additional constraints. For the first time these models were proposed by Balinski and Quandt [2]. A main advantage of this type of models is that it allows extremely general route costs, e.g., depending on the whole sequence of the arcs and on the vehicle type. Moreover, the additional side constraints need not take into account the restrictions concerning the feasibility of a single route. As a result, they often can be replaced by a compact set of inequalities. This produces a formulation whose linear programming relaxation is typically much tighter than that in the previous models. Note, however, that these models generally require dealing with a very large number of variables.

II. A Model of Vehicle Routing Problem with Soft Time Windows and Variable Traveling Time

The vehicle routing problem for which a model is presented is based on a real-world routing problem. It is generalization of VRPTW where multiple time windows are supported. The service takes place entirely in the city area. This means that the main city roads are overloaded during the rush hours and the travel duration for a road segment will vary significantly depending on the time of the day. The city infrastructure does not provide paid highways that can be used to construct more expensive but faster routes. The average number of customers to be served is 600. The number of vehicles is up to 40 and they are identical. The model is robust enough to allow future modifications such as addition of new constraints or changes in the objectives.

A model is introduced for a vehicle routing problem with soft time windows and variable traveling times. The model is based on the model presented in [19] and is chosen because of its simplicity and flexibility. Let $G = (V, E)$ be a complete directed graph with a vertex set $V = \{0, \rightleftharpoons, n\}$ and an edge set $E = \{(i, j) \mid i, j \in V, i \neq j\}$ and $K = \{1, \rightleftharpoons, k\}$ be the vehicle set. Vertex 0 is the depot and the other vertices are the customers to be served. Each customer *i*, each vehicle *k* and edge $(i, j) \in E$ is associated with:

- $g_i \geq 0$ - the demand of goods of customer *i*;

- $p_i(t)$ - time window cost function of the start time *t* of the service at customer *i*;
- $p_0(t)$ - time window cost function of the arrival time *t* at the depot;
- $C_k \geq 0$ - the capacity of vehicle *k*;
- $d_{ij} \geq 0$ - the distance between vertices *i* and *j*;
- $q_{ij}(t)$ - traveling time function from *i* to *j*;

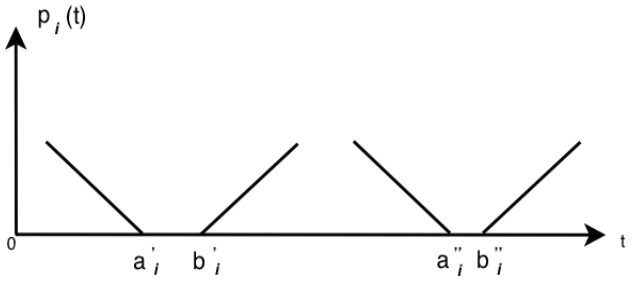


Figure 2. Cost function $p_i(t)$ for two time windows

Function $p_i(t)$ may be piecewise linear and look as given on Fig. 2. The actual slope may be chosen according to the preferences for the real problem instance and affects how 'soft' is a time window. The model can be transformed to hard time windows if $p_i(t)$is defined as follow:

$$p_i(t) = \begin{cases} 0 & t \text{ within the time window} \\ +\infty & t \text{ outside the time window} \end{cases}$$

Function $q_{ij}(t)$ is different from the described in [19]. In the proposed model the function will reflect the different traveling times for the same road depending on what time of the day the vehicle passes it. It takes the current time and returns the estimated travel duration between customers *i* and *j*.

Let σ_k be the route traveled by vehicle *k* and $\sigma_k(h)$ be the *h*th customer in route σ_k. By n_k the number of nodes on route σ_k is given. For convenience $\sigma_k(0) = \sigma_k(n_k + 1) = 0$ for all *k* (i.e. every route will start and finish at the depot). With σ the set of the routes followed by the vehicles will be denoted

$$\sigma = (\sigma_1, \sigma_2, \rightleftharpoons, \sigma_k)$$

Additionally let us introduce the following notations:

- s_i - the start time of service at customer *i*;
- s_k^a - the arrival time of vehicle *k* at the depot;
- $s = (s_1, s_2, \rightleftharpoons, s_n, s_1^a, s_2^a, \rightleftharpoons, s_n^a)$;

Let us introduce also binary variables $y_{ik}(\sigma) \in \{0,1\}$ for $i \in V \setminus \{0\}$ and $k \in K$ by

$$y_{ik}(\sigma) = 1 \Leftrightarrow i = \sigma_k(h) \quad (1)$$

Equation (1) holds for exactly one $h \in \{1, 2, \rightleftharpoons, n_k\}$. This means that $y_{ik}(\sigma) = 1$ holds if and only if vehicle *k* visits customer *i*. We can express the total distance traveled by all vehicles as $d(\sigma)$, the total time window cost for all customers $p(s)$, the total traveling time cost $q(\sigma, t)$ and the total excess amount $g(\sigma)$ as follows:

$$d(\sigma) = \sum_{k \in K} \sum_{h=0}^{n_k} d_{\sigma_k(h), \sigma_k(h+1)} \quad (2)$$

$$p(s) = \sum_{i \in V \setminus \{0\}} p_i(s_i) + \sum_{k \in K} p_0(s_k^a) \quad (3)$$

$$q(\sigma, t) = \sum_{k \in K} \sum_{h=0}^{n_k} q_{\sigma_k(h), \sigma_k(h+1)}(t) \quad (4)$$

$$g(\sigma) = \sum_{k \in K} \max \left\{ \sum_{i \in V \setminus \{0\}} g_i y_{ik}(\sigma) - C_k, 0 \right\} \quad (5)$$

The mathematical programming model will be:

Minimize

$$c(\sigma, s, t) = d(\sigma) + \alpha p(s) + \beta q(\sigma, t) + \gamma g(\sigma) \quad (6)$$

subject to

$$\sum_{k \in K} y_{ik}(\sigma) = 1, \quad i \in V \setminus \{0\} \quad (7)$$

Constraint (7) requires that every customer $i \in V \setminus \{0\}$ must be served only once by exactly one vehicle. The proposed objective function is a weighted sum of $d(\sigma) + \alpha p(s) + \beta q(\sigma, t) + \gamma g(\sigma)$. The constants $\alpha \geq 0, \beta \geq 0$ and $\gamma \geq 0$ in (6) determine the relative importance of each component of the objective function. They are set in advance depending on the preferences for each specific problem instance.

From the problem description it is obvious that the best approach to solve it will be to use a heuristic algorithm. Current methods to solve VRP problems to optimality use Branch-and-Bound, Branch-and-Cut [24], Branch-and-Cut-and-Price [12] or Set-Covering-Based algorithms [1]. They are able to solve problem instances of up to about 100 customers with variable success rate [21]. For larger problem instances both academic research [21] and commercial organizations [18] concentrate on heuristics. Heuristics are also more flexible than the exact approaches [21] and will be easily adapted to changes in the problem model. Heuristic and metaheuristic approaches are intensively researched in the recent years [5–7]. VRP with time windows has wide range of applications and is well studied [4, 13, 17]. Mathematical models with multiple time windows are also present in literature [11, 20].

Many of the heuristics and metaheuristics for VRPTW show variable performance and are also very dependent on the quality of the initial solutions [8]. As an improvement a multistart local search heuristic approach is proposed. It consists of two phases with an optional post-optimization procedure. In the first phase sequential insertion heuristics is used to generate a set of initial solutions. In the second phase inter and intra-route cross-exchange is invoked to reduce the total distance of the solutions with the minimum number of routes. In [19] this model is solved using an algorithm based on iterated local search. It starts from an initial solution σ and searches for a better solution in the neighborhood $N(\sigma)$. Standard neighborhoods as 2-opt*, cross-exchange and Or-opt with slight modifications are used for neighborhoods $N(\sigma)$. Approaches using behaviorally inspired algorithms are also

present in the literature. Bee colony algorithm is proposed in [17]. Ant colony metaheuristic is proposed for VRP with multiple time windows in [9].

The model proposed can be considered to be a multicriterial one if we treat functions $d(\sigma), p(s), q(\sigma,t)$ and $g(\sigma)$ as separate objectives. In this formulation it is not necessary to determine in advance the relative importance of the components in (6). VRPTW and its variants are intrinsically multicriterial in nature [3] since the different components in (6) are related to each other. For example, the minimization of the total traveling cost $q(\sigma,t)$ may cause the vehicle to miss a few or more time windows at some customers, which will increase the value of $p(s)$. By using the multicriterial optimization approach we can find different alternative solutions and the final solution to be left to decision makers depending on the context of the specific problem instance.

III. Conclusion

In order to be applicable to a wider range of real life problems, one of the main features of a model should be its flexibility. The model should be adapted easy to variations of the constraints or the formulation of the goals. The presented model covers these requirements and is a good starting point to apply different metaheuristic approaches. Metaheuristics are suitable because they may provide good trade-off between speed, quality of the solution and the size of the problem input. The directions for future work are research on interactive algorithms guided by expert, application of parallel algorithms to the model and development of multicriterial algorithms for this model.

Acknowledgement

This research is supported in part by the Bulgarian National Science Fund, Grant No DTK02/71 and IICT-BAS research project “Modeling, Optimization and Multiple Criteria Decision Making”.

References

[1] R. Baldacci, N. Christofides, and A. Mingozzi. An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts. Math. Program., 115:351–385, June 2008.

[2] M. Balinski and R. Quandt. On an integer program for a delivery problem. Operations Research, (12):300–304, 1964.

[3] B. J. Ross, B. Ombuki and F. Hanshar. Multi-objective genetic algorithms for vehicle routing problem with time windows. Applied Intelligence, 24:17–30, 2006.

[4] R. Bent and P. Van Hentenryck. Spatial, temporal, and hybrid decompositions for large-scale vehicle routing with time windows. In Proceedings of the 16th international conference on Principles and practice of constraint programming, CP’10, pages 99–113, Berlin, Heidelberg, 2010. Springer-Verlag.

[5] H. Brandao de Oliveira and G. Vasconcelos. A hybrid search method for the vehicle routing problem with time windows. Annals of Operations Research, 180:125–144, 2010.

[6] O. Bräysy and M. Gendreau. Vehicle routing problem with time windows, part i: Route construction and local search algorithms.Transportation Science, 39:104–118, February 2005.

[7] O. Bräysy and M. Gendreau. Vehicle routing problem with time windows, part ii: Metaheuristics. Transportation Science, 39:119–139, February 2005.

[8] O. Bräysy, G. Hasle, and W. Dullaert. A multi-start local search algorithm for the vehicle routing problem with time windows. European Journal of Oper. Research, 159(3):586 – 605, 2004.

[9] E. Moretti, D. Favaretto and P. Pellegrini. Ant colony system for a vrp with multiple time windows and multiple visits. Journal of Interdisciplinary Mathematics, 10(2):263–284, 2007.

[10] G. B. Dantzig and J. H. Ramser. The truck dispatching problem. Management Science, 6(1):80–91, 1959.

[11] K. F. Doerner, M. Gronalt, R. F. Hartl, G. Kiechle, and M. Reimann. Exact and heuristic algorithms for the vehicle routing problem with multiple interdependent time windows. Comput. Oper. Res., 35:3034–3048, September 2008.

[12] R. Fukasawa, M. P. De Aragão, M. Reis, E. Uchoa. Robust branch-and-cut-and-price for the capacitated vehicle routing problem. In Proceedings of the International Network Optimization Conference, pages 231–236, 2003.

[13] A. Garcia-Najera and J. A. Bullinaria. An improved multiobjective evolutionary algorithm for the vehicle routing problem with time windows. Comput. Oper. Res., 38:287–300, January 2011.

[14] W.M. Garvin, H.W. Crandall, J.B. John, and R.A. Spellman. Applications of linear programming in the oil industry. Management Science, (3):407–430, 1957.

[15] B. Gavish and S. Graves. The travelling salesman problem and related problems. 1979.

[16] B. Gavish and S. Graves. Scheduling and routing in transportation and distributions systems: Formulations and new relaxations. 1982.

[17] S. Häckel and P. Dippold. The bee colony-inspired algorithm (bcia): a two-stage approach for solving the vehicle routing problem with time windows. In Proceedings of the 11th Annual conference on Genetic and evolutionary computation, GECCO ’09, pages 25–32, New York, NY, USA, 2009. ACM.

[18] Randolph W. Hall. Vehicle routing software survey, 02 2010.

[19] H. Hashimoto, T. Ibaraki, S. Imahori, and M. Yagiura. The vehicle routing problem with flexible time windows and traveling times. Discrete Appl. Math., 154:2271–2290, November 2006.

[20] T. Ibaraki, S. Imahori, M. Kubo, T. Masuda, T. Uno, and M. Yagiura. Effective local search algorithms for routing and scheduling problems with general time window constraints. TRANSPORTATION SCIENCE, 39:206–232, 2002.

[21] G. Laporte, Gilbert Laporte, and Les Cahiers Du Gerad. What you should know about the vehicle routing problem, 2007.

[22] A. N. Letchford and J. Salazar-gonzalez. Projection results for vehicle routing. Math. Program., Ser. B, 105:251–274, 2006.

[23] S. Martello and P. Toth. Knapsack problems: algorithms and computer implementations. John Wiley and Sons, Inc., New York, NY, USA, 1990.

[24] P. Toth and D. Vigo, editors. The vehicle routing problem. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.

Architecture of a Flexible Web-Based Framework for Building Models and Solving Decision Optimization Problems

B. Staykov[1], F. Andonov[2],D. Vatov[1], K. Genova[1], L. Kirilov[1], V. Guliashki[1]

Abstract – **A flexible web-based framework for multiple criteria decision support is presented. The system is targeted at different types of users – researchers, educators and business people and should facilitate the problem solving process of different types of optimization problems, mainly single and multi-objective linear programming problems with continuous and/or integer variables.**

Keywords – **Web-based systems, decision support, single and multiple objective optimization.**

I. INTRODUCTION

The evolution of decision support systems can be traced back to the dawn of computers. "First-generation" systems were *single computer - single user* type and implemented only one method or a few similar methods of one type (see [17, 18] for example). With the spreading of the Internet some network-based systems appeared – both web-based and traditional client-server(A good survey about such systems is given in [12]). Yet most of the systems were problem-oriented and they could only solve one specific problem(see [15, 16] for example). Many of them were based on the java applet technology, but this approach has many limitations. This architecture is still basically a client-side technology, and is limited by the resources of the client machine and hence the data volume, memory, CPU power and problem complexity. Later on group decision support systems gained popularity, where not a single decision maker is responsible for the decision but a group of decision makers.

Web architecture by itself does not allow persistent connections, and because of that there are limits on the response time and the amount of data exchanged between the server and client, but the solving time for this kind of decision problems is highly undetermined. As a result, the usability of such systems is limited to problems whose solving time does not exceed the browser time-out interval.

The development of the Internet and its total penetration in all social areas combined with globalization brought a new breed of systems – less monolithic, more versatile, incorporating cutting-edge software technologies [12, 14].

The modern Internet technologies provide platform-independent, remote computation, as well as exchange of complex multimedia information. The end users of decision support systems can focus their efforts on problem analysis and decision making.

The use of Decision support systems in general, requires specific knowledge background about the methodology of mathematical optimization and its applicability to the user's professional area.

As a result of the above, it is a challenge to create a universal system, which is method-, user-, solver-agnostic and applicable to a wide range of example, research and business problems. In this paper, we describe a software system which represents our understanding of what the architecture of such a system should look like.It is under the provisional name WebOptim.

II. IMPORTANT ADVANTAGES OF WEBOPTIM

Web-accessible: The researchers [11, 12] in the area of operational research (OR) consider that the development of Web-access technologies is of key importance for the usability of optimization, so as the name suggests, the system should be accessible from the Web via a browser, and users should be able to define, solve, save, load and share their decision problems.

User-agnostic: WebOptim should allow to be used by different types of users – educators, who demonstrate the large variety of single and multiple criteria methods on different size and type of problems, researchers, who will test their own methods or/and solvers and business people, who will solve their real-world decision making problems with a method of their choice. The diversity of users and goals translates into the requirements of a highly-customizable user interface and an extendible framework. Users should be able to define their own problems, to solve them with any applicable method, save them for later evaluation, andbe able to see examples or similar problems. Another way for broadening the user target group is by offering well-formed and solved examples of typical optimization problems. Less-experienced users will be able to browse these examples and hopefully find one similar to their own problem.

Solver-agnostic: WebOptim should incorporate several solvers with metadata about their applications and methods they are used by. System administrators will be able to add new solvers, written in different programming languages.

[1]Boris Staykov, E-mail:bstaykov@iinf.bas.bg
Daniel Vatov, E-mail: daniel.vatov@gmail.com
Krasimira Genova, E-mail: kgenova@iinf.bas.bg
Leoneed Kirilov, E-mail: lkirilov@iinf.bas.bg
Vassil Guliashki, E-mail:vggul@yahoo.com
are with the Institute of Information and Communications Technologies- BAS, 1113 Sofia, "Acad. G. Bonchev,bl. 2, Bulgaria.

[2]Filip Andonov is with the NewBulgarianUniversity, Sofia 1618, z.k. Ovtcha Kupel,.Montevideo str. № 21, Bulgaria, E-mail: vonodna@yahoo.com

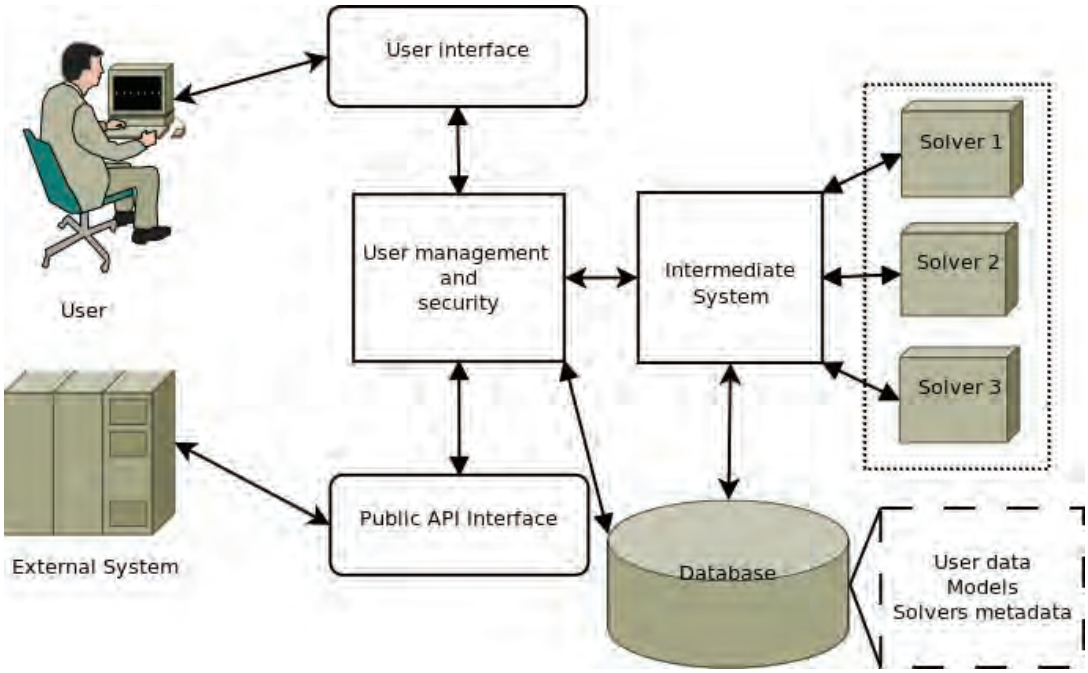


Fig. 1. Architecture of WebOptim

Method-agnostic: WebOptim should have metadata for all the included methods, their applications and the solvers they use. System administrators will be able to add new single or multiobjective solvers.

Heterogeneous: The system should allow to be continuously developed by several teams, using different software technologies. That is why a reliable and flexible environment for data and command interchange is needed to provide the common ground. The authors think that web-services are the answer to this requirement. Using web-services allows developers to choose a programming language, operating system and software design of their own taste and to glue their piece of the puzzle with as little effort as possible.

III. Structural Description

For maximum modularity the system is composed of the following main components: user management and security, intermediate module, database and solvers. The principle schema of this intermediate system is given in Fig. 1.

A. User management and security

In order to use the system, each user must create its own personal profile. The profile containing login credentials, user type, personal details, defined problems, their solutions and additional information about errors and specific solving details will be stored in a common database. The registration of users will be done through standard web page forms. After registering and logging in their profile, the users will be provided with interface for defining and solving problems. Those problems must be defined by one of the below described standard syntaxes. After syntax verification, each problem will be stored in the database with unique ID parameter. The user will be able to see a list of these defined problems and their status. There will be four possible states of this problem status:

- "Not solved". When the problem is defined but still not sent to the solver;
- "Waiting for solution". When the problem is submitted to the solver queue and solution is expected when ready. In this state the user should not be able to edit the problem definition;
- "Solution obtained". When a solution is obtained from the solver and is ready to be previewed;
- "Solver error". The solver returns some error.

This part of the module will provide tools for the user to preview problems and solutions; start or interrupt solution process for certain problem; preview error messages received from the solvers and to obtain specific detailed debug information about the solving process (if the chosen solver provides this information).

B. Database

The main database will store all the data for the users, problems they have saved, solutions and metadata about the methods and solvers for use in the intermediate system. The metadata should contain at least the solver's name, description, type (integer, mixed integer, multicriterial, etc.) and type(s) of accepted input. Because of the network environment the system will be running in, if the solvers need a database for temporary data during the solving process, they can use their DBMS and databases to minimize the communication volumes and thus reduce solving time. To

overcome the lack of connection persistence between the client and the server in web-based systems all intermediate and final results are stored in the database, which is possible due to the nature of these types of optimization problems, where every intermediate result can be used as a reference point for the next(every method can be used as interactive).This also allows unbindingof the calculation process from the user interface. The user is no longer required to wait for the solution and to keep the connection alive because all the results will be stored in database and associated with his/her personal profile and problem.

C. Solvers

All solvers will implement a common XML SOAP[2, 3] protocol for communication with the intermediate system. This allows diversity of the solvers, the machines they are running on, the language they are written in, etc. The limitation of this approach is that due the asynchronous nature of this type of communication, the control will be harder to implement, because in a simplified way the communication looks like this: intermediate system sends problem for solving and when ready, the solver gives an answer or an error (if occurred). Because there are many problem definition formats, it is certain that at some point in time there will be a set of solvers that support a non-intercepting set of syntax formats. An intermediate module will translate the problem definition to a format understandable by the solver. Another way of interaction will be optimization problem data to be delivered directly to the solver. This will be the case of already existing solvers or solvers that are considered useful for the system's users. They will have their own syntax for the input. Besides AMPL as a modeling language there are many other modeling languages used by quality solvers that is worth integrating into the system. Possible modeling formats are listed bellow [9]:

- MPS file format - The MPS format is supported by most lp solvers and thus very universal. The model is provided to the solver via an ASCII file. This format is very old and difficult to read by humans. See [8] for a complete description of the format.

- lp file format - The lp format is the native lpsolve format for providing LP models via an ASCII file to the solver. It is very readable and its syntax is very similar to the Mathematical formulation. See [6] for a complete description of the format

- CPLEX lp file format - The CPLEX lp format is another format for providing LP models via an ASCII file to the solver. It is very readable and its syntax is very similar to the Mathematical formulation. It is a format used by the CPLEX solver. See [4] for a complete description about the format

- LINDO lp file format - The LINDO FILE format is another format for providing LP models via an ASCII file to the solver. It is very readable and its syntax is very similar to the Mathematical formulation. It is a format used by the LINDO solver. See [7] for a complete description about the format.

- GNU MathProg file format - The GNU MathProg format is another format to provide LP models via an ASCII file to the solver. It is very readable and its syntax is very similar to the Mathematical formulation. It is a format used by the GLPK solver and a subset of AMPL. It has also the possibility to use loops. See [9].

- LPFML/OSIL XML file format - The LPFML XML format, currently named OSIL after becoming part of COIN-OR project (https://www.coin-or.org/), is another format to provide LP models via an ASCII file to the solver. This format is very recent and uses XML layout. It is not very readable by humans, but because of the XML structure is very flexible. For more informationsee [10].

Solvers will be exposed to external systems following Web Services [1] paradigm.

D. Intermediate core system module

The fact that the system will be used as an educational and research tool means that it has to be as open and extensible as possible, so new methods, solvers, problem types, and basically every kind of modules can be added later. The development team is heterogeneous and not static, especially for the projected lifetime of the product, so different software libraries and components and different program languages will be used in it, which put a lot of stress on the system architectural design. To make these different parts glued together several modern technologies and principles will be applied to the development process – extensive use of XML as communication standard and web services.

There is one major problem that concerns the intercommunication between the different modules in the system. The problem appears because this intercommunication needs to be done asynchronous due to its nature - mainly HTTP and SOAP protocols, which have very tight response timeouts. Depending on its size, each solving problem will take different time to be solved by a certain solver and sent back to the sender module. This makes it impossible for the communication to be done in an uninterrupted cycle of request-response.

This problem is solved by bringing in an intermediate system module, which will take care of all the communication between end user interfaces, database storage and solvers. This system will perform two main tasks – user management and solvers communication.

E. Solvers communication

On the other hand, this intermediate system module will handle the bidirectional communication with solvers, based on the XML SOAP protocol.

There are two subtasks in this part of the module – sending requests to the solvers and providing a web service for handling solver responses for the solutions of the problems.

Sending a solver request will be done by posting data to the unified solver web service through XML SOAP protocol and will contain the following mandatory fields:

- Problem ID (an integer number field);
- Problem syntax (a string field (different syntaxes are described in another part of the article));
- Problem definition (a string field);

- Debug (a Boolean field) – defines if additional debug information is requested.

Additional fields may be implemented at later stage of the project. After posting this information, a confirmation response is expected from the solver and the problem state is changed to “Waiting for solution”. The solver response web service will expect information posting from solvers (again in a standard XML SOAP protocol) and will look for the following fields:

- Problem ID (an integer number field);
- Problem solution (a string field) - in case of a successful solution;
- Solver error message (a string field) - in case of a solver error;
- Debug information (a string field) – in case it is requested.

When such a request is received, the module will perform search in the database for the certain problem ID, store the solution (or the error message) and change the problem status to “solution obtained” or “solver error”, according to the current case. The web service must response to the solver request; otherwise the solver must continue to try to post the solution after a certain amount of time, until a confirmatory answer is received!

IV. Conclusion

There exists a variety of stand-alone, network and web-based decision support systems. Despite the fact that with the increased demand for such systems in global economy, the development of semantic technologies and many R&D teams working on less fragmented, more universal systems, there is still room for improvement. The key features of the proposed web-based system WebOptimare:

• Useful to a wide variety of users from different professional backgrounds with different level of optimization competence.

• A user friendly customizable interface, reflecting the needs of different users and accessible worldwide via the Web;

• A set of solvers which covers the most popular optimization and decision making problems;

• Designed to be easily extended by adding new solvers;

• Providing an API interface for external use by third party developers.

By implementing our view of such architecture, we intend to overcome the inherited limitations of web-based systems in general and target a larger user group. Several other similar systems exists, but they are not widely used, mainly because specialists from different areas do not recognize their problems as optimizationproblems and they are not familiar with the available software instruments for decision support. In that respect, WebOptim is an attempt to promote and encourage the use of decision support optimization systems in all areas where such problems occur.

One future direction for the development of WebOptim is to make it able to solve single and multi-objective nonlinear optimization problems by implementing the necessary solvers, methods and user interfaces.

Acknowledgement

This research is supported in part by the Bulgarian National Science Fund, Grant No DTK02/71 and IICT-BAS research project Modeling, Optimization and Multiple Criteria Decision Making.

References

[1] Web Services Architecture. Technical report, World Wide Web Consortium, February 2004.
[2] Paul V. Biron and Ashok Malhotra, editors. XML Schema Part 2: Datatypes. W3C Recommendation. W3C, second edition, October 2004.
[3] http://www.xml.com/pub/a/2000/02/09/feature/index.html
[4] P. Notebaert K. Eikland. Cplex file format. http://lpsolve.sourceforge.net/5.5/CPLEX-format.htm.
[5] P. Notebaert K. Eikland. Lindo file format. http://lpsolve.sourceforge.net/5.5/LINDO-format.htm.
[6] P. Notebaert K. Eikland. Lp file format. http://lpsolve.sourceforge.net/5.5/lp-format.htm.
[7] P. Notebaert K. Eikland. lp_solve reference guide. http://lpsolve.sourceforge.net/5.5.
[8] P. Notebaert K. Eikland. Mps file format. http://lpsolve.sourceforge.net/5.5/mps-format.htm.
[9] A. Makhorin. Gnu linear programming kit, modeling language gnu mathprog.http://plato.asu.edu/gnu_mp.pdf.
[10] A. Makhorin. Optimization services linear language https://www.coin-or.org/OS/OSlL.html
[11] P. Valente, G. Mitra. The evolution of web-based optimization: From ASP to e-Services. Decision Support Systems, Volume 43, Issue 4, (2007) 1096–1116.
[12] H. K. Bhargava, D. J. Power and D. Sun, “Progress in Web-based decision support technologies”, Decision Support Systems, Volume 43, Issue 4, August 2007, pp. 1083 – 1095
[13] A.M. Geoffrion, R. Krishnan, Prospects for operations research in the E-business era, Interfaces 31 (2) (2001)], 6-36.
[14] M. Andersson, H. Grimm, A, Persson, Amos Ng, A Web-basedsimulationoptimizationsystemforindustrialscheduling, In: Proceedings ofWSC '07, (Henderson, S. G. et al, eds.), IEEE Press Piscataway, NJ, USA 2007.
[15] Plácido Rogério Pinheiro1 and José Auriço Oliveira,WEB - Based Optimization System Applied to High School Schedule Building, http://www.asap.cs.nott.ac.uk/patat/patat04/553.pdf.
[16] P. Korhonen, A Visual Interactive Support System for Multiple Criteria Decision Making, Belgian Journal of Operations Research, Statistics and Computer Science, 27 (1), 1987, 3-15.
[17] V. Vassilev, B. Staykov, F. Andonov, K. Genova, M. Vassileva, Multicriteria Decision Support System MOLIP, Cybernetics and Information Technologies, 2 (1), 2002, 3-15.

Improving Quality of Geo-data in Electric Utility Companies Using Mobile GIS

Nikola Davidovic[1], Leonid Stoimenov[2]

Abstract – **Company's geospatial data precision is the key issue for analysis, decision making and management using Geographic Information Systems (GIS). Mobile GIS applications should provide increased precision of geospatial data acquisition transferring acquisition process directly in the field. Increased geospatial data quality during acquisition could be provided by using mobile device's GPS and communication abilities. In addition, on site data collection can provide better description of the objects of interest. To provide field acquisition of geo-data about power supply network GINISED Mobile is developed, as a component of existing GINISED system. GinisED Mobile is used to improve the quality of geo-data within an electric utility company. It provides users with ability to update existing and collect new geo-data according to the situation on the field. Lineage and better positional accuracy are addressed through integration with GPS of the mobile device and depend only on its precision. In addition, better accuracy of descriptive attributes and context of geo-data are achieved by providing users to define their values directly in the field. This paper presents mobile GIS application's architecture and implementation and defines processes for its use in the field.**

Keywords – **Mobile GIS, Utilities, Field data collection, geo-data.**

I. INTRODUCTION

Nowadays, utility companies are using Geographic Information Systems (GIS) in order to improve management of their large network infrastructure, dispersed on large geographic area. GIS should not be considered mainly as a digital technical documentation but as a central information system, that provides integration of various information in spatial context. The usage of GIS enables utilities to improve investment/maintenance/development planning activities, customer care, network performance analysis, reporting etc. consequently leveraging their overall business performance. GIS enables users to capture, store, analyze, and display geographically referenced information. It allows them to view, understand, query, interpret, and visualize data in a way that is quickly understood and easily shared [1].

Spatial data are used in different analysis within utility companies like asset tracking, losses calculations, quality of service, optimal performing of the utility infrastructure etc. These analyses greatly depend on the accuracy of available spatial data and their attributes. Therefore, the quality of used geo-data plays an important part in utilities [2]. However, in previous period the quality of geo-data has been largely influenced by human factor due to an outdated collection process. It was based on dispatching crews on the field with primary task of determining the approximate location of the real geo-objects (defining geometric attributes) comparing the real world situation in the field against paper maps. Collected data (handwritten on the paper) are then stored in spatial data warehouse, using a specific GIS editor. Next step required qualitative description of collected data trough definition of specific details and possible relations to other data. Previously described geo-data collection process was susceptible to errors due to imprecise and incomplete data, poor interpretation of collected data or lack of collection procedures. This requires usage of new technologies that will positively affect the collection process in terms of quality and simplicity of collection process in the field.

Improved geo-data collection process needs to address better location accuracy of the object of interest, its orientation in relation to the environment and mechanisms to store data about these objects. In addition, qualitative attributes and context of the object are better viewed and collected directly in the field. The availability of GIS applications in the field, allows the elimination of the previously mentioned shortcomings of the collection process. Field GIS application has a limited set of functionality compared to standard GIS applications due to mobile computing devices' limited computing power that are available in the field. The emphasis of these applications is on the orientation, locating the precise position and comfort of work on site. This approach aims at improving the capabilities of verification of already collected geo-data and increasing the quality of new data [3].

The reminder of paper presents the solution for the collection of geo-data in the field, which is an extension of the existing GIS with mobile component. The second chapter gives the basic features that contemporary mobile GIS applications need to provide with strong emphasis on characteristics needed for fieldwork. It also discuss the characteristics of geo-data that must be addressed in order to improve their quality. The third section examines the architecture of GINISED system for recording electric power grid and defines the position of the mobile GIS component. The fourth chapter presents procedure for data preparation, field collection and data verification and provides a detailed insight into the architecture and implementation details of the GINISED Mobile system.

[1]Nikola Davidovic is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nikola.davidovic@elfak.ni.ac.rs.

[2]Leonid Stoimenov is with the University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: leonid.stoimenov@elfak.ni.ac.rs.

II. MOBILE GIS FOR FIELDWORK

Mobile GIS and mobile applications in general today represent a trend in the development of information systems [4]. Design and development of such systems is not easy because it requires opposing performance requirements of mobile applications in conditions of limited resources of mobile devices. In addition to the problems that come to mind when developing applications for mobile devices, GIS applications pose additional requirements that must be also accounted such as [5]:

- Information services and distributed processing on the server side,
- Mobile devices' capabilities in terms of communication and geo-location,
- Relational database available on mobile devices and
- Software and hardware characteristics of mobile devices.

During the process of designing mobile GIS for fieldwork, it is necessary to take into account the technical characteristics of the mobile client (hardware capacity constraints, embedded modules for location, communication, etc..) as well as the physical characteristics of the environment and conditions for use (atmospheric conditions, the availability of communication services, etc.). These conditions largely determine the GIS platform requirements and appropriate client devices that can be used. In addition, it is necessary to provide the server side support trough the development of appropriate services and definition of protocols and procedures for synchronization and verification of data collected in the field.

Mobile devices can use different types of communications: Wireless LAN (802.11b / g), packet data networks using mobile telephony (GPRS) or company's radio communications using additional modems [4]. Due to underdeveloped infrastructure, communication is not possible at some locations and in this cases system should be sufficiently adaptive to provide some autonomy. This includes preserving context when working in offline mode, using map caching mechanisms and some kind of persistence storage, preferably mobile database [6].

Positioning using mobile devices is enabled using GPS or mobile network base stations. Today, the coverage of mobile network is good, especially in urban areas but its accuracy is measured in tens of meters. GPS provides much more accurate positioning measured in meters, and the precision can be increased using corrective methods [7]. Professional GPS devices, designed to work in the field have an accuracy of about 0.5 to 5 meters. In many cases, such precision is not enough and better results are achieved using additional antennas that provide better reception of the satellite signal.

Relational database on a mobile device, except as a means of storing data retrieved from the central system, should enable the independence of the device by preserving the state of the application in cases when communication with central system is not available. Mobile database replication has to provide certain parts of the company's central database available on the mobile device. The basic concept is the use of synchronization in which case, data in embedded database are automatically synchronized with main (back-end) database. [6][8]. Another option is direct copying of embedded database files from device to computer, reading data and selective recording in the main database. Later avoids the possibility that "bad" data, resulting from adverse conditions in the field or the human factor, get into the database.

The quality of spatial data is measured trough five characteristic: lineage, positional accuracy, attributes accuracy, logical consistency, and completeness [9]. All accounted criteria are of great importance for electric utility companies. Therefore, mobile GIS has to provide improved quality of geo-data by improving accuracy of each of these five characteristics. Criteria like lineage and positional accuracy are addressed trough accurate positioning using GPS and only depend on the precision that can be achieved with mobile device's GPS receiver and conditions in the field. Attributes accuracy, logical consistency, and completeness primarily depend on the way that user percepts objects in the field and collects data. Therefore, good user experience of mobile GIS needs to be provided through simple and intuitive user interface appropriate for fieldwork.

III. GINISED MOBILE ARCHITECTURE

For the purposes of recording, manipulation, maintenance and analysis of power distribution network, PD Jugoistok Nis in cooperation with the Laboratory for Computer Graphics and GIS, Faculty of Electronics Engineering in Nis continued the process of development of specialized GIS called GINISED [10][11]. GINISED provides mechanisms for recording, manipulation and spatial analysis of data about electric power distribution network. The application has object-oriented GINISED architecture. Its development is largely based on previous work on the development of object-oriented framework for GIS applications [10]. All the tools necessary to manage power distribution network, in GINISED system are implemented as two independent applications [11]:

• GINISED Editor - Specialized tool for creating of spatial data of the electricity distribution network (creation of geographic schemes of electric power distribution network, editing parameters of network elements and defining their interrelationships) and

• GINISED Viewer - WebGIS application with three-tier software architecture that provides efficient mechanisms for viewing and searching of spatial data on the selected part of the electric distribution network.

GINISED Viewer allows users to access spatial data using a thin Web client from any computer, but does not solve the problem of availability of geo-data for workers in the field. In order to enable access to geo-data to users who work in the field, the existing system has been expanded by adding a mobile component called GinisED Mobile.

GINISED Mobile has to enable rapid acquisition of geo-data in all areas covered by PD Jugoistok, which have not yet been entered into the system. It can be used for revision of already entered geo-data according to the real situation on the field. Improving data quality will increase the overall value of the whole GINISED system for the company PD Jugoistok [10].

GINISED Mobile is a three-tier application that consists of:
- GINISED Mobile Editor with mobile database
- GINISED Mobile Server
- Centralized Geospatial Database.

GINISED Mobile Editor is a mobile device application with primary goal to support mobile teams of PD Jugoistok working in the field (Fig. 1.). Online mode is supported through the application component framework Ginis Mobile for working with maps [4]. In addition, GINISED Mobile Editor needs to support users working in offline mode. GINISED Mobile Editor contains the presentation layer with the possibility of working with raster and vector maps and the data logic in the form of mobile databases. The user interface is customized for the better utilization of the map and working with objects.

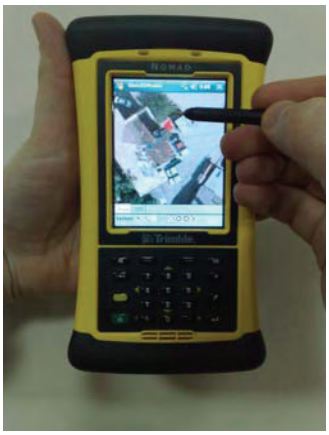


Fig. 1. GINISED Mobile Editor on mobile device

The second part of the application logic is on the server side. The central database contains all the necessary information on electricity distribution network [10][11]. Component GINISED Mobile Server prepares all the data necessary for the operation of GINISED Mobile editor in offline mode. This implies preparing raster map segments for the area of interest and vector data necessary to complete the task and then transferring them to mobile device. Vector data are only a subset of data that will be updated in the field. All other vector data are used in form of raster in order to speed up the drawing on the screen. In cases where raster maps are not available on the local computer's file system, GINISED Mobile Server uses maps obtained from WMS.

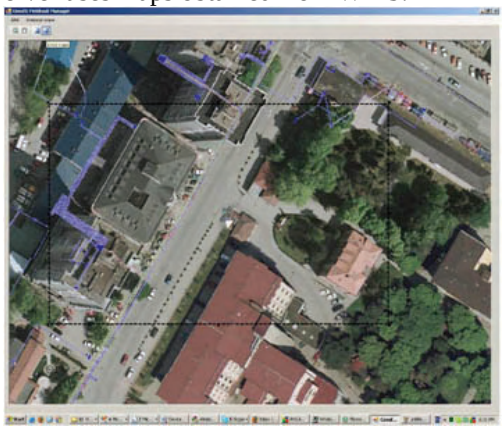


Fig. 2. Selecting area for fieldwork in GINISED Mobile Server

IV. PROCEDURES FOR FIELDWORK AND APPLICATION IMPLEMENTATION

In PD Jugoistok, specific procedure for the use of mobile GIS applications in the field was defined. The fact that in most cases, application will work in offline mode influenced the procedure significantly. Procedure for data collection consists of three steps (Fig. 3.):
- Preparation of data for a defined fieldwork area
- The process of data collection in the field
- The takeover procedure, validation and data entry into the central database

Prior to fieldwork, mobile device needs to be prepared by preparing raster maps and mobile database file with necessary data. The procedure consists of selecting the appropriate area on the map of GINISED Mobile Server (Fig. 2.), and then the selection of geo-data of interest for the defined location. Prepared maps and mobile database file are copied to mobile device. The main objective of this approach is minimizing calculations and processing on the mobile device during drawing of raster map and vector data [10]. Prepared maps are optimized for the device display and for a given area thus reducing occupancy of secondary memory and significantly reducing needed working memory for unnecessary segments of the map. If communication in the field is available, raster map segments are downloaded from WMS. Raster layers that can be selected are retrieved from the response on the *GetCapabilities* request to the WMS. In online mode, data are obtained using synchronization with central database that is natively supported by mobile DBMS (in our case Oracle Lite). Although data update is possible directly from the collection site, all data must first pass validation within company before they are stored in central database.

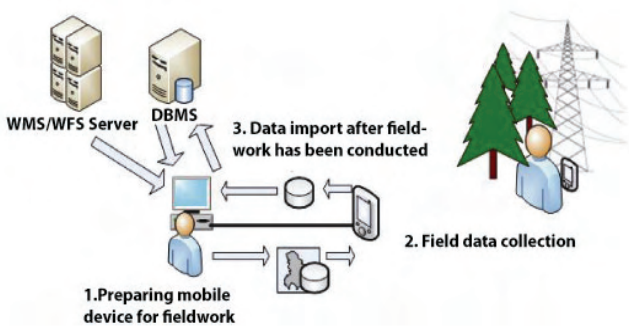


Fig. 3. Geo-data collection process

During fieldwork, users enter new or change existing data. In the process of geo-data collection, it is necessary that person who collects data, stands next to the object of interest in order to acquire its GPS location. Alternatively, user can point to the location on the map, relative to the current position in cases when objects of interest are not accessible. The next step is the selection of the appropriate type of entered object. The final stage is setting the orientation of the object in accordance with actual situation in the field. Optionally it is possible to enter additional description of the object. The process of changing the existing objects requires that they already exist in the mobile database and are shown in

vector layer. The basic purpose is to determine the exact location and orientation of the edited object, but it is possible to change its type and attributes too.

After the team returns from the field, data are imported from the device in GINISED Mobile Server and displayed on the map. Depending on the needs, location and orientation of imported objects can be fine-tuned according to raster map (most accurate aero-photo and satellite maps). If the object is valid the authorized person confirms it and enters it into the central database.

GINISED Mobile Editor consists of three subsystems:

- Subsystem for working with raster map
- Subsystem for working with vector layer
- Subsystem for working with database

The entire continuous raster map consists of a matrix of segments raster maps. Given the static nature of the map segment, the optimal approach is to use local cache of raster maps segments that have already been used on the client. This mechanism significantly increases the performance of the map display. Map display is based on two level cache. The first level uses secondary memory to store map segments [4]. In the case of working in offline mode, this cache is used to hold all the necessary map segments for the area of interest. Since the secondary memory has relatively long access times, second level cache is introduced. Because of spatial data locality, it is expected that the user will require map segments adjacent to those segments that were last accessed. The second level cache keeps last N map segments in RAM of the device where N is dynamically calculated based on the currently available memory. When the amount of free memory becomes insufficient for new map segments, algorithm removes old segments from the secondary cache according to the LRU algorithm.

Subsystem for working with vector layer is specific due to constraints of the mobile platform. Therefore, solutions used are different from the usual practice in conventional GIS. Drawing vector layer elements is based on copying of the bitmap prototypes. For each element a bitmap prototype is created such that enables much faster access to bitmap pixels. While drawing, each pixel of the map canvas is mapped to the corresponding pixel of the prototype. This solution enables drawing and rotation of complex symbols that are not natively supported on mobile device. Such drawing is processor-intensive thus the results of mathematical operations are stored in separate structures for later use. The advantage of this drawing approach is the consistent view of vector layers in all GINISED applications.

Subsystem for working with mobile database is based on interfaces for database access. This approach completely separates presentation layer from data access layer. The main advantage of this approach is reduced number of access to mobile database thus reducing the possibility of error.

V. CONCLUSION

The system presented in this paper is a good foundation for further expansion of mobile services for supporting field crews in PD Jugoistok. Monitoring of the collection process will identify the impact of such mobile applications on business processes in PD Jugoistok. In future, the emphasis will be on identifying all potential users of mobile GIS in PD Jugoistok. By upgrading the existing corporate network infrastructure, owned by PD Jugoistok, it will be possible to introduce a high degree of interaction between employees in the field and employees within company. Data communication provided using mobile GIS will provide timely response of repair and maintenance crews. Employees in PD Jugoistok will be able to achieve significant savings reducing time spent on the ground and reducing transport costs through dynamic forwarding of new tasks while the crews are still at locations of interest. Time for development of other mobile applications will be significantly reduced by using presented architecture. Further work is based primarily on improving user interface in accordance with the experiences of the users and improving the presentation layer by improving existing algorithms.

REFERENCES

[1] D. Pickering, J. M. Park, D. H. Bannister, "Utility Mapping and Record Keeping for Infrastructure", Urban Management and Infrastructure - Urban Management Programme, Washington, D.C., Vol. 10, pp. ix-11., 1993

[2] F. F. H. Nah, K. Siau, and H. Sheng, “TheVALUE of Mobile Applications: A Utility Company Study”, Communications of the ACM /Vol. 48, No. 2, February 2005

[3] H. Pundt, K. Brinkkoetter-Runde, "Visualization of spatial data for field based GIS", Computers & Geosciences 26, pp. 51-56, 2000.

[4] D. Rančić, B. Predić, A. Dimitrijević, "Implementation of mobile GIS in field work", WSEAS Transaction on Computers, Vol.5, No.11, (pp2690-2696),ISSN:1109-2750, November 2006.

[5] A. A. Solyman,“Investigating Mobile GIS”, Directions Magazine, http://www.directionsmag.com/articles/investigating-mobile-gis/123298, November 2005.

[6] W. Li, H. Yang, P. He, “Mobile Database Application Research based on GIS”, IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, ISBN 978-0-7695-3490-9/08, pp. 787 – 792, 2008.

[7] L. Luqun, L. Minglu, “A Research on Development of mobile GIS architecture”, Environmental Informatics Archives, Volume 2, pp. 920-926, 2004.

[8] M. H. Dunham, V. Kumar, “Location Dependent Data and its Management in Mobile Databases", 9th International Workshop on Database and Expert Systems Applications (DEXA'98), pp.414, 1998

[9] H. Pundt, “Field Data Collection with Mobile GIS: Dependencies Between Semantics and Data Quality”, GeoInformatica 6:4, 363 – 380, 2002.

[10] A. Stanimirović, D. Stojanović, L. Stoimenov, S. Đorđević–Kajan, M. Kostić, A. Krstić, "Geographic Information System for Support of Control and Management of Electric Power Supply Network", Proceedings of IX Triennial International Conference on Systems, Automatic Control and Measurements SAUM, ISBN-86-85195-49-7, Niš, 2007.

[11] L. Stoimenov, A. Stanimirović, M. Bogdanović, N. Davidović, A. Krstić, "GinisED – Geo-Information System for Support of Evidencing, Maintenance, Management and Analysis of Electric Power Supply Network", Proceedings of 3rd Small Systems Simulation Symposium, pp. 23-26, ISSN: 978-86-6125-006-4, 2010.

Database Modelling and Development of Code Generator for Handling Power Grid CIM Models

Sasa Devic[1], Branislav Atlagic[2] and Zvonko Gorecan[3]

***Abstract* – This paper presents a solution for modelling database, in order to store and manipulate CIM models, and code generator that will ease the work on developing support system for that database. The work contains basic description of CIM models and exchange procedure between clients participating in power trading process. Developed code generator relies on database relational schema. It is designed for power grid CIM models, but it can be used for other models as well. The general instructions for developing code generator are given. At the end, the overall CIM data handling process and resulting performances are presented.**

***Keywords* – CIM model, database, code generating, ICEST 2011.**

I. INTRODUCTION

Since general use of electrical energy has started, till the end of the XX century, the customers were forced to buy energy from monopolistic power supply companies which covered the wider area they lived in. Customers had no choice to choose between different, probably more affordable, companies. Economic experts rightly claimed that monopolistic companies are starting to slow down the technological progress. If more companies were fighting for the costumers, the final result would brought a better and cheaper energy supplies to the customers.

Events that followed were expected. Big, monopolistic, usually state companies, were reconstructed into large number of smaller companies or sectors that took part in managing different parts of once one company. For this market to operate, it was necessary to introduce auction houses for electrical energy trading. Auction houses work as a agency agent between different participators in trade, and therefore they need specialized software tools for trade. Opposite to other markets of energy resources, in trading with electrical energy the key role has the knowledge of physics of the system or, in other words, the knowledge of electrical supply network and its capabilities. For those needs a unique model for describing every network of different vendors was developed[1].

For developing software tools that would operate with such model, specialized software companies started to form. Those companies required knowledge of both electric power engineers and software engineers to develop a unique solution for the problem. This work gives contribution to those efforts.

[1]Sasa Devic is with Telvent DMS D.O.O., Sremska 4, 21000 Novi Sad, Serbia. E-mail: sasa.devic@dmsgroup.rs.

[2]Branislav Atlagic is with Telvent DMS D.O.O., Sremska 4, 21000 Novi Sad, Serbia. E-mail: branislav.atlagic@telventdms.com.

[3]Zvonko Gorecan is with Telvent DMS D.O.O., Sremska 4, 21000 Novi Sad, Serbia. E-mail: zvonko.gorecan@dmsgroup.rs.

Since 1951, UCTE (*Union for the Co-ordination of Transmission of Electricity*) standard was used to exchange electrical network models between different EMS (*Energy Managment System*) operators. But since then, many things have changed. The need to model data more precise, and to cover electrical elements not included in UCTE model, such as shunts, generators, transformer windings, switchers and so on, a new CIM (*Common Information Model*) model was developed. In 2009 UCTE became part of ENTSO (*European Network of Transmission System Operators*). ENTSO accepted CIM standard as preferred, and in 2009 first interoperability tests were made, although CIM model is still in developing phase[2].

II. CIM MODEL

A. Basics

The CIM Model Exchange Profile is a ENTSO standard that is based on the CIM standards produced by IEC WG13 (*The International Electrotechnical Commission, Work Group 13*). The purpose of CIM standard is to define how members of ENTSO, using software from different vendors, will exchange network models as required by the ENTSO business activities. The following basic operations are sufficient for TSOs (*Transmission System Operator*) to satisfy ENTSO network analysis requirements: ***export*** (TSO may use the profile to export its internal network model in such way that it can be easily and unambiguously combined with other TSO internal models to make up complete models for analytical purposes), ***import*** (TSO must be able to import exported models from other TSOs and combine it to make complete model), ***exchange*** (any model, covering any territory, sent to any other party, must carry the data who formed it, which data brings and for which use case is designed for).[3].

B. File structure

ENTSO CIM model are packed and exchanged as XML (*Extensible Markup Language*) data model. Data division among files is based on the kind of information in each file. This division typically divides less rapidly changing information from more rapidly changing information, setting up the situation where some exchanges are smaller because they only contain files that have changed. Therefore, model information exchange is divided into three files, TSO equipment model, TSO topology and State variables. Information from all three files can be combined into one "complete model", which concatenates all data.

The CIM model itself is designed with *abstract* and *concrete* classes. Through those classes the physics of

electrical power system, its states at the specific time, are mapped to the model. Abstract classes are used to ease the complexity of the system, they group and define base attributes and associations, dividing more and less general parts of the system. Again, real (concrete) parts of the system are left to be described by concrete classes, which inherit much of its attributes and associations from abstract classes. Concrete classes are dependent on abstract classes. Still, there are concrete classes that do not inherit any abstract class. Anyway, data exchange involves only concrete classes. As an example *Tap Changer* class will be presented (see Fig. 1).

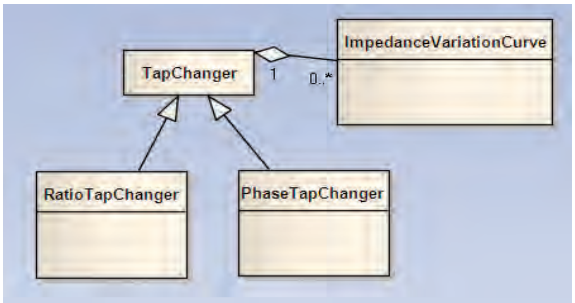


Fig. 1. Tap Changer (object-oriented model)

Tap Changer class is inherited by *Ratio Tap Changer* and *Phase Tap Changer*. Both, *Ratio* and *Phase Tap Changer* have all the attributes and associations of the *Tap Changer*, but with addition of a few of its own. There are some attributes and associations in *Ratio Tap Changer* that do not exist in *Phase Tap Changer*, and vice versa. On the other hand, if some concrete class, like *Impedance Variation Curve*, has an association to *Tap Changer*, which is an abstract class and can not exits in data exchange, it actually associates to either *Ratio* or *Phase Tap Changer*. Association to the abstract class is a true problem for mapping object-oriented model (hierarchical model) to the database relational data model (flat model), which will be more explained in the next chapter.

Tap Changer is chosen because it is simple enough for an example and it has all kinds of associations that we wish to demonstrate. Other elements of CIM model have more complicated associations which would be hard to follow, but with the same types of associations like *Tap Changers* has.

III. Database Modelling

As we know, basic elements of which relational databases consists are tables (which are composed of one or more columns) and relations between them. The relational data model has been around for many years and has a proven track record of providing high performance and flexibility. But, there is no possibilities of defining real inheritance among data tables or defining an abstract table. Databases are designed to contain large amounts of data, and to allow quick and relatively easy way to access them. Any kind or mapping object-oriented class model to relational database model is not completely possible, modeling of the two do not follow the same design. To solve this problem, we have implemented some patterns:

- For each abstract class that we have in our object-oriented model, appropriate table must be created (in this work we will call them "abstract tables", like in [4]).
- For each concrete class, appropriate table must be created (in this work we will call them "concrete tables", like in [4]).
- For each association to other class, abstract or concrete, appropriate relation to the table that represents that class must be made.
- Inheritance is mapped as identification relation to the table that represents "*parent*" class (key columns of a "*child*" are all from "*parent*" table). The exception from this rule are abstract and concrete classes that have no "*parent*" class.

So far this is straightforward, but with large amount of redundant data! If we store all data that abstract classes define, very quickly our database is going to grow enormously. In order to prevent that from happening, since all data defined in abstract classes must be withheld within their inheritors, can we remove all tables that represent abstract classes and use only tables that represent concrete classes? By doing that, relational model would simplify and reduce amount of data in database, but there is a problem with that...

If there is a class with an association to the abstract class, in relational data model we cannot model that as a relation to the tables that represent concrete classes that inherit that abstract class. In relational model we cannot map one object-oriented association to a multiple database-oriented relations, depending on the number of how many classes inherit aimed abstract class. For example, in table *Impedance Variation Curve* we would have to have relations to both tables that represent *Ratio* and *Phase Tap Changer*, since they both inherit *Tap Changer* class in object-oriented model. If we could define a reference with a choice clause, that could be one of possible solutions for the problem (but that is not possible). We are forced to retain tables that represent abstract classes. The solution to the problem is to use abstract tables only as key containers, to which we can set relation through which we can find needed concrete table. All data is stored in the concrete tables only. Here we will mention that we need abstract tables from maintenance reasons as well. CIM model is still in development process and it is expected to change. Relational model for *Tap Changer* is shown in Fig. 2.

IV. Code Generating

Now, we have determined what kind of data is exchanged between different software vendors (CIM XML files) and we know in what form to store such data in our database. The piece of the puzzle still missing in our software solution is *how*? We need to have a support to insert the data into database and later to read from it and manipulate with data.

This may sound like a simple task, but if we have to solve many simple tasks over large amount of different types of data, our *simple* problem becomes a *big* problem. Developed relational model has 86 tables with very complex relations among them. Possible solution is to make a new problem, which will solve our first problem – to write a program that will write a program that we need.

We have designed our database relational model to reflect object-oriented model, we can use its meta-data (data that

describes data) to generate program code and use it to manipulate with database. Code can be generated in such way that will allow us to manipulate with objects that reflect their states back to the database.

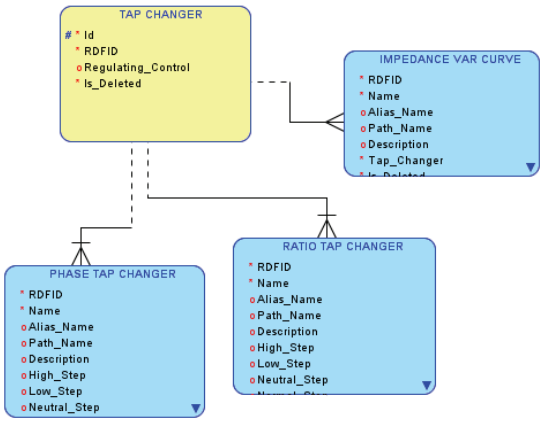


Fig. 2. Tap Changer (relational model)

A. Database reader

First step is to develop *Database reader*, a software component that reads meta-data from database relational schema. All database meta-data is possible to read. For every table that we have created, we can read its name an other meta-data related to table. When we know what tables are present in database, in a similar way we can read their columns meta-data. For each column we can read its name, data type, size, numeric precision and scale, is mandatory, is read only, and most importantly is it a primary key column, a foreign key column, or both. We can even read column constraints on data it can contain. Important to mention is that based on foreign keys (imported keys), we can find out which primary keys from one table are used as foreign keys in other tables (exported keys).

To ease our search through meta-data, we have created an object model where we will store meta-data that we have read. Since it's storing a description of meta-data, which is data that describes data, that model is actually a meta-meta-data model. All other parts of our code generator rely on this meta-meta-data model.

B. Database procedure generator

In order to fully exploit database we have to use database procedures. Insert, update and delete statements are very slow when compared to equivalent procedures. The main drawback when using statements is that they need to be parsed first, and then executed. In contrast, procedures executes immediately.

To avoid writing all required procedures manually, we have created a *Database procedure generator*. Its job is to generate (write) those procedures, using meta-meta-data. We know the names of tables, their columns and data types.

For insert procedure we need to define input parameters, which are values for columns of a table in which we wish to insert a new row. Next step is to define what our procedure needs to do, in this case to insert a new row into the table. Table name, its columns, and type are used as input parameters in this process.

For update and delete procedures we need to do similar work, except we are not generating insert, but update and delete statements. Needed syntax is different, but the logic is similar for all.

Views are also generated by this component, to allow us faster and easier read of data.

All in all, with some additional procedures that will be explained later in this work, we have generated almost 15.000 lines of code for those procedures, and 2.500 lines to drop them later.

C. Database API generator

To work with database in a way like we work with objects in our program language, we need to develop a communication layer that will allow us to do that. *Database API generator*, based on meta-meta-data model, is developed to generate an object-oriented model who's states will be reflected on database.

We need to find a way to generate a component layer to stand between database and the rest of our application. Lets observe our tables in database. Each table is consisted of rows and columns. Columns describe a certain data of one type, with specific constraints. We will map columns like attributes of a certain class. Rows can be viewed as sections of data of a certain table. We will map rows like a classes, where class name is delivered from table name. Additionally, based on imported keys we have generated association to the class that represents a certain table in database, and based on exported keys – collection of associations. By generating classes that we need, we are allowed to use them in our code like any other class that we have wrote by hand. The example of this code is given in Listing 1.

```
class TapChangerEl {
    // Exported Relations
    List<ImpVarCurveEl> impVarCurveEls;
}
class ImpVarCurveEl : CurveEl {
    // Imported Relation
    TapChangerEl tapChangerEl;
}
class RatioTapChangerEl : TapChangerEl { ... }
```

Listing 1: Generated code example

Classes that we have created must have support to call database procedures that we created by *Database Procedure generator*. By applying similar logic we are able to generate class functions that are able to prepare the call to database, set needed parameters, execute the call, and finally do additional work with result. Class functions generated for that are insert, update and delete.

Database API generator is also in charge for generating functions for fetching data from database and creating class objects from it. Fetch functions include various forms, depending by which criteria data is read (by primary keys, by foreign, by columns, by no criteria or by combination of

them). Fetch functions rely on views generated by *Database Procedure generator*.

This code generator is designed to cover some of specific needs of CIM model, but it can be used in general use, as well. Some of specific needs are presented in next section. Sum of generated lines of code, for 86 tables, is 506,781 lines.

V. Import-export Model

So far, our model, that we got as the output of code generator, is covering all functionality we need, with ability to be easily expanded in the future. But, in order to get any data from concrete tables, first we have to fetch their abstract tables that index them! Number of calls to database can be multiplied several times depending on the number of abstract tables through which we have to search to get the data in some concrete table. In general, as deeper the hierarchy of object-oriented model is, as slower the communication to database will be.

As we mentioned before, data contained in abstract tables is also contained in tables that "inherit" them (basically only keys). Since the aim is to reduce number of calls to the database, we could read only concrete tables, and by doing that we get data from abstract tables as well. This approach reduces number of calls to the minimum, and there is no bottleneck to the database. To compare performances without and with this principle see table I (old and new for export).

Based on that idea, to read only concrete tables, it is possible do the same in opposite direction. Database procedures will take parameters for inserting a row into a concrete table, first appropriate data will be inserted to all needed abstract tables and after that to the aimed concrete table. By applying this principle we have achieved performance acceleration in both ways. For this we had to expand both *Database Procedure generator*, to create such procedures in database, and *Database API generator*, to give us access to those procedures. To compare performances without and with this principle see table I (old and new for import).

TABLE I
PERFORMANCE TABLE FOR IMPORT AND EXPORT CIM MODEL WITH 50.000 ELEMENTS (54MB FILE)

	Import [s]		Eksport [s]	
No.	*old*	*new*	*old*	*new*
1	1173,0	188,9	49,42	4,30
2.	1377,7	209,5	47,42	2,69
3.	1014,6	212,8	47,58	2,71
4.	1470,2	183,6	47,79	2,66
5.	944,5	148,3	48,35	2,64

VI. Software Package

Software application is build with two basic components, *Cim Manager Lib*, *Database Api* and *Cim Mediator*. Application development process, that we are about to explain, is presented on Fig. 3.

By the standard, CIM model is exchanged with XML files. Since C# programming language and .NET have well developed technologies to support manipulation with XML, software application is developed using those technologies. This covers *Cim Manager Lib*, as described in [5].

To communicate with database *Database Api* is generated. It is a wrapper component to ease the access to the database. Database supported by this project is Oracle 10 XE, an library used for communication is ODP.NET (*Oracle Data Provider for .NET*). ODP.NET. It is possible to add support for other databases as well.

Cim Mediator is a component for joining the *Cim Manager Lib* and *Database API* into one solution. It represents the entrance point of application and contains and exposes functions for import, export and for fetching functions of entire CIM model.

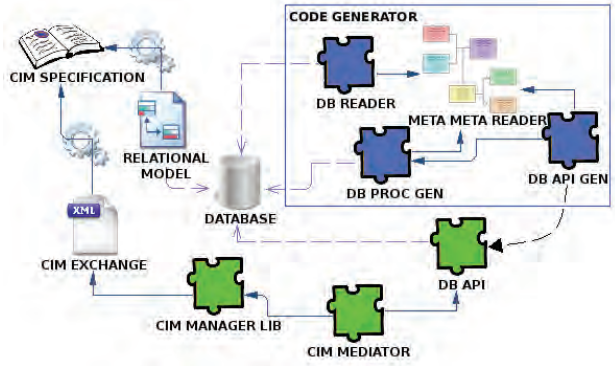


Fig. 3: Development process

VII. Conclusion

In this work we have explained what CIM models are, and the way how to work with them when stored in database. We have presented components that help us work with database. Code generator is described, through which we solved the problem of writing large amount of code in limited time. The most important advantage of code generating is that allows us to relatively easy expand our generator depending on current needs. Another advantage is that it gives us the power to go even further with code generating. If we need to have client-server architecture with our database, we can make an extension for code generator and generate web service and client that we need. This could be done as an update in the future.

References

[1] Daniel Kirschen, Goran Štrbac „Fundamentals of Power System Economics", 2005.

[2] Union for the Co-ordination of Transmission of Electricity CIM interoperability Test – *EPRI*, 2009.

[3] Dragan Tomić, Ranka Slijepčević, Lajoš Martinović, Nemanja Živković, „Validation and merging of national transmission network models into one interconnective model", 2009.

[4] "UCTE CIM Model Exchange", component interface exchange, revision 1.0, version 14, 2009.

[5] Lajoš Martinović, Ranka Slijepčević, Nemanja Živković, "Software tool for conversion between power system models", 2010.

Built-in Testing of Embedded Software Systems

Irena Pavlova[1], Aleksandar Dimov[2]

Abstract – **Embedded software systems encompass a broad range of devices, which may be mixed hardware/software system dedicated for a specific application. Usually embedded software systems are part of and manipulate and control a larger, physical system. There are a lot of unresolved issues in building reusable, reliable and predictable systems of that class. One such issue is software testing, which currently is done in an ad-hoc manner. In this paper we propose an approach for testing of embedded software systems, based on the so-called Built-In-Testing (BIT). BIT is a concept, where given software module or system has capabilities to perform testing itself.**

Keywords – **Embedded systems, Component Bases Software Development (CBSE), Built-In-Testing (BIT).**

I. INTRODUCTION

Embedded systems are often distributed real-time systems comprising electronics and software. Such systems are increasingly penetrating every aspect of our lives and work, from telecommunication systems, transport, energy and utilities, health, finance, education, tourism and environment. The embedded systems industry is competing with decreasing time to market and increasing product differentiation.

Both lead to increasing dependence on software required to be flexible enough for rapid reuse, extension and adaptation of system functions. It is often difficult to test and verify embedded systems because of the intrinsic "embedded dimension". This is an effect of that the software has to be designed on a platform different from the platform on which the application is intended to be deployed and targeted.

Embedded systems are also often mission-critical and needs to be extensively verified and testing is one of the major challenges. Compared to standard PC software embedded software is harder to observe, test, and debug.

The contribution of this paper is towards interesting and needed research areas within BIT for component-based embedded systems. With the above aim we propose a reference model for Quality of Service (QoS) BIT testing. The main target at this stage is non-functional requirements like timeliness and performance. Within this context we focus on the following main concerns:

• Increase software quality, in terms of functional and non-functional properties

• Shorten development times, in terms of the development process and specifically test reuse.

The remainder of the paper is structured as follows: In section 2 CBSE for Embedded systems is presented. Section 3 makes an overview of BIT technology. Section 4 discusses the application of BIT for components in embedded systems. Section 5 presents a reference model for QoS BIT testing of component based embedded systems. Finally, section 6 concludes the paper.

[1]Irena Pavlova is with the Faculty of Mathematics and Informatics, Sofia University, James Boucher 5, Sofia, Bulgaria E-mail: irena_pavlova@fmi.uni-sofia.bg.

[2]Aleksandar Dimov is with the Faculty of Mathematics and Informatics, Sofia University, James Boucher 5, Sofia, Bulgaria E-mail: aldi@fmi.uni-sofia.bg.

II. CBSE FOR EMBEDDED COMPONENT BASED SYSTEMS

Assembling new software systems from existing components is an attractive alternative to traditional software engineering practices which promises well defined software architectures, reduced developments costs as well as reuse [4]. However, these benefits will only occur if separately developed components can be made to work effectively together with reasonable effort [8]. However, lengthy and costly verification and acceptance testing may impact negatively the independent component development and system integration.

This way application of new processes, approaches and instruments for supporting effective integration and reducing manual system verification effort in the context of embedded software systems is needed. This may be done by equipping components with the ability to check their execution environments at runtime. Built-in-test (BIT) is such an instrument, providing a model for elaboration of detailed tests while developing the component.

One of the major driving-forces behind component-based development is reuse; however, in many companies reuse has not been very successful even though component-based development has been introduced in the software lifecycle. It is often required to restructure the organization to reflect the component based process, i.e., divide component development from system development. Another major obstacle for reuse is efficient administration (e.g., version and configuration management) with growing component repositories [5].

Most embedded systems have requirements not present in other systems, e.g., timeliness, low footprint, low energy consumption, etc.

Such non-functional requirements need to be verified and validated, adding another dimension of testing to the system. Hence, it is essential to satisfy not only the functional behavior, but also extra-functional properties such as, e.g., timing and dependability attributes. These systems characteristics usually implies that embedded systems are statically configured, i.e., the components used and their interconnections are decided at design or configuration time. Here, the binding is static, as opposed to the dynamic binding used in most desktop component technologies.

Furthermore, embedded systems are resource constrained in the sense that the per-unit cost is a main optimization criterion, i.e., the use of computer and computing resources should be kept at a minimum. Also, due to the high variability

of many embedded systems, it is common within the embedded systems industry to use product-line architectures. Because of this, reuse of architectures, components, quality assessments and tests are very attractive for reducing development costs.

III. Built-in-Testing Overview

Testing is a disciplined process that consists of evaluating the application (including its components) behavior, performance, and robustness – usually against expected criteria. One of the main criteria, although usually implicit, is to be as defect-free as possible.

Expected behavior, performance, and robustness should therefore be both formally described and measurable.

Compatibility of components is one of the greatest issues. It is not of much use to specify a component as part of large and complex software system if it will not deliver what has been promised. One of the key ways of addressing this issue is to build components that are self-testing, to ensure that they meet the specifications for that part of the total application.

BIT [10] proposes to build test-mechanisms into components and systems during design and coding, so that the successive testing and maintenance processes can be simplified. The most interesting feature of the BIT is that tests can be inherited and reused in the same way as that of code in the conventional COTS components [7].

Built-in testing of software components can be done in a large number of ways. The Component+ project [1] developed a methodology for integrating BIT components into COTS software, using methods that are a significant extension of the object-oriented technology. The design principles of BIT for software components embrace two major perspectives: Contract testing - to verify a contract between two components from both parties point of view. Quality of service testing - to verify that the operating environment of a software component continues to give the right service, that the interaction between all components works and that residual faults in a component prevent proper function of the component or the system.

The BIT architecture is based on the following elements:

- **BIT-component**: component that provides a number of built-in test services and test interfaces, as shown on Fig. 1.

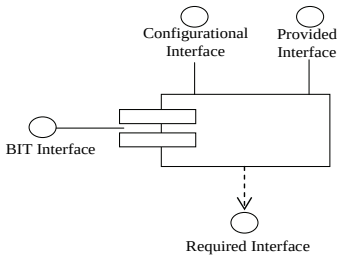


Fig. 1. BIT Component

- **Testers**: components that use the test services of BIT-components to determine whether a system-level error condition exists.
- **Handlers**: components that handle errors detected by BIT components or test components.
- **System constructor**: a conceptual element, nominally responsible for the instantiation of (high level) BIT-components, testers, and handlers, and their interconnection. Note that both BIT components and testers can detect error conditions. The BIT component can detect internal (i.e. component-level) errors, whilst the testers detect external or system-level errors arising from incorrect component interaction.

IV. Built-in-Testing in the Context of Component-based Embedded Systems

With traditional development approaches, the bulk of the integration work is performed in the development environment, giving engineers an opportunity to pre-check compatibility of system various parts, and to ensure that the overall deployed application is working correctly. In contrast, late integration implied by component assembly means that there is little opportunity to verify the correct operation of applications before deployment time.

Although component developers may adopt rigorous test methodologies, with non-trivial software components it is impossible to be certain that there are no residual defects in the code - formal proof or 100% test coverage are not viable options in most practical cases. Compilers and configuration tools can help to some extent by verifying the syntactic compatibility of interconnected components, but they cannot check that individual components are functioning correctly (i.e. that they are semantically correct), or that they have been assembled together into meaningful configurations (i.e. systems). As a result, components that may have behaved correctly in the sanitary condition of the development-time testing environment, may not behave so well when deployed in a system where they have to compete with other (third party) components for resources such as memory, processor cycles and peripherals.

Sophisticated verification methods are used to increase the level of assurance of critical software, particularly that of safety-critical and mission-critical software. Embedded software verification is a systems engineering discipline that evaluates software in a systems context [9].

In order to bring the effectiveness of verification to bear within a reuse-based software development process it must be incorporated within the domain engineering process. Failure to incorporate verification within domain engineering will result in higher development and maintenance costs due to losing the opportunity to discover problems in early stages of development. The component Verification, Validation and Certification Working Group at WISR 8 found four general considerations that should be used in determining the level of verification of reusable components [6]:

- Span of application - the number of components of systems that depend on the component
- Criticality – potential impact due to a fault in the component

- Marketability – degree to which a component would be more likely to be reused by a third party
- Lifetime – duration of component usage.

Although encapsulation and information hiding are central principles for facilitating the design and development of component based software systems, their very nature also complicate the task of testing. This is because some of the information that is necessary for comprehensive testing of objects and components is by definition hidden to entities outside a component (e.g. the test software).

Many software components are state machines and the state information is hidden. Encapsulation and information hiding thus give rise to a couple of fundamental problems inherited in conventional software components technologies, which have yet to be addressed in CBSE:

- Low testability.
- Low maintainability for CBSE actors.
- No support for run-time testing.

These problems hold even when components are supplied in their complete form, i.e. with the source code. Software is seldom so well documented that a user unacquainted with a component can verify it in an easy way. Test software delivered with the component increases the testability.

A piece of software with encapsulated state information is testable if we can:

- Set it into a given state before a test.
- Stimulate it with given test data.
- Read the response and the resulting state.
- Compare the actual outcome of the test with expected outcome.[1]

To make such software component testable it should be able to get access to the encapsulated state information of the component before a test is invoked. This holds for tests of behaviour. For other kinds of testing the test software must have access to other internals of the component. Hence part of the testing software has to be built-in.

Besides the testing challenges of standard functional testing of component-based systems, embedded systems have a range of extra-functional properties that also need to be verified. Some of the important attributes for embedded systems that define quality, besides correct functional behavior, are [2, 3]:

- **Real-time properties** – violation of time requirements, despite correct functional behavior, violates the system behaviour.
- **Dependability** – the ability of the system to deliver a service that can be trusted.
- **Resource consumption** – Many embedded systems have strong requirements on low and controlled consumption of different resources.

Besides quality aspects, an important issue for the embedded systems segment is time-to-market. Component-based development has shown to be an efficient paradigm for increasing productivity and lowering development time and costs. However, component based development for embedded systems has not been as successful as for, e.g., desktop systems, especially not considering reuse.

One of the major reasons to this is the lack of support for configuration and version management.

Thus, the perhaps most important aspects for reducing development time and time-to-market are:

- **Reuse** is a basic concept in CBSE that decreases development time and time-to-market.
- **Software configuration management** – important for embedded systems in the context of reusability.
- **Verification** - To find errors in the code, and hopefully at an early stage indisputably shortens the time for testing, redesign etc.

V. Built-in QoS Testing Model for Component-Based Embedded Systems

Within a real (as opposed to a test) system, a component competes with other components for resources such as memory, processor cycles and peripherals. Consequently, its performance may be affected by the system in which it is integrated. This is particularly critical in real-time systems where a component may have deadlines to meet or a certain throughput to achieve. Adequate system performance should be designed into the system, but this requires components to be characterized in terms of the resources they require as well as their functional and dynamic behavior. This is not usually done, so system performance has to be measured during development and deployment. Therefore, a requirement exists for QoS testing to support verification of components dynamic behavior.

Timing and performance testing is an indisputable part of each testing effort. The strict requirements towards embedded systems as well as the utilization of external resources (components) increase the importance of testing the timing as well as the performance of the components when integrated into assemblies.

The reference architectural model we propose for Timing BIT QoS testing is illustrated on Fig 2. The Timing tester is intended to measure the time spend for data access for a particular component scenario. It is important the test for every component to be performed in a single transaction. Single transactions are used also with the purpose any time dependencies to be avoided. Time measurement will start before the starting of the transaction and will end after the end of the transaction.

Fault situations are handled by a *Handler* component. For example such situation may occur if incorrect or impossible attempt to access the data in the database is made.

The Component under test should support *IBITTiming* interface, which is used to perform a timing test on a single component. This interface allows subscribing or unsubscribing for time event for the particular component.

Timing tester provides *IBITTimingNotify* Interface that is used for notification for time events and requires *IBITError* interface. This interface is used by the Timing Handler, and provides only one function, which requests the handler to process the thrown Timing exception. The processing of the exceptions includes logging the exception in file and other user defined functions. The *IBITErrorNotify* interface is required on the Handler for reporting errors.

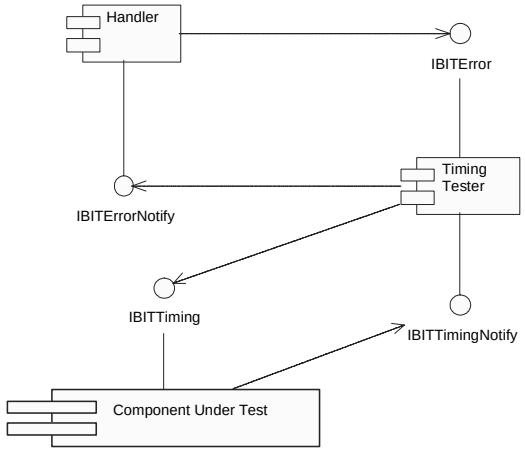


Fig. 2. BIT Timing Testing Reference Architecture

For realizing performance testing that is presented on Fig. 3, a Performance tester is developed "on the top" of the Timing tester. The main idea is to use several time tests on different components. This tester uses a single transaction for every set of time tests. Fault situations are handled by a relevant Handler component. This component collects information about time spent for a particular scenario involving several different components.

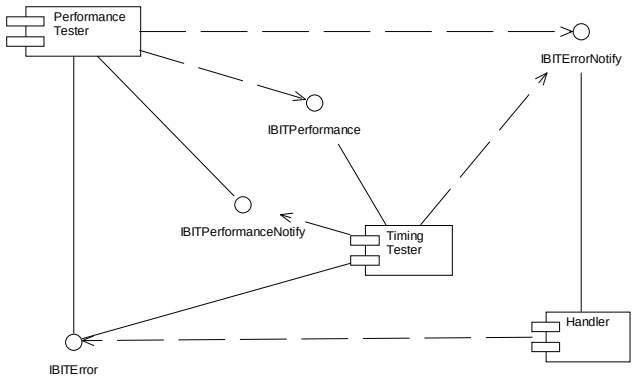


Fig. 3. BIT Performance Testing Reference Architecture

The Timing tester component must support IBITPerformance interface. This interface is used when performing the testing. The IBITPerformance interface allows for subscribing to a particular set of components, which will be monitored during the performance test execution. Performance tester provides IBITPerformanceNotify Interface that is used for notification of events. Using this interface performance tester will receive information about time spent for a method or information that there is an error during the execution of the method.

VI. Conclusion

Real-time aspect is one of the major differences between embedded software and PC or internet software. Embedded software systems often interact and controls physical processes with real-time requirements. Timing is often of first priority in testing efforts. To use BIT in embedded real-time systems it is important to understand the relation between the two.

Testing worst-case response-time is not trivial. Typically, tests assess within what time the embedded system reacts (creates an output considering an input). The assessed time forms an end-to-end latency for the response. Internally, the system typically involves transactions of several execution threads that must cooperate to create the correct output. These threads can in turn experience interference from other parallel activities in the system. Thus, the goal for the test is to measure the time from a certain change in the input until a certain output is produced under maximum disturbance from other parallel activities. These techniques not only have to set up a worst-case scenario (which is challenging in itself), but also have to measure the system non-intrusively, i.e., make sure that the measurement does not affect the test (probe-effects).

Acknowledgement

The work presented in this paper was partially supported by the National Science Fund, part of the Ministry of Education and Science in Bulgaria under the ADEESS (MU01-143) and TASSA (ДО-02-182/16-12-2008) projects.

References

[1] Component+ EU FP5 Project, 2006.
[2] I. Crnkovic. Component-based approach for embedded systems. In Ninth International Workshop on Component-Oriented Programming, Oslo, June 2004.
[3] I. Crnkovic. Component-based software engineering for embedded systems. In International Conference on Software engineering, ICSE'05, St. Luis, USA, May 2005. ACM.
[4] I. Crnkovic and M. Larsson. Building Reliable Component-Based Software Systems. ISBN 1-58053-327-2. Artech House, 2002.
[5] I. Crnkovic, S. Larsson, and M. Chaudron. Component-based development process and component lifecycle. In 27th International Conference Information Technology Interfaces (ITI), Cavtat, Croatia, June 2005. IEEE.
[6] S. H. Edwards and B. W. Wiede. Software engineering notes, 22,5,17-31. WISR8L 8th Annual workshop on SW Reuse, 1997.
[7] K. J. Fernandez, V. H. Raja, and M. Morley. A system level testing modeling mechanism in a reengineering environment. Journal of Conceptual Modeling, issue 18, 2001.
[8] D. Garlan, R. Allen, and J. Ockerbloom. Architectural mismatch or why it's hard to build systems out of existing parts. In Proceedings of the Seventeenth International Conference on Software Engineering, April 1995.
[9] D. Wallace and R. Fujii. Software verification and validation: an overview. IEEE Software, 6(3):10–17, May 1989.
[10] Y.Wang, G. King, D. Patel, S. Patel, and A. Dorling. On coping with real-time software dynamic inconsistency by built-in tests. Annals of Software Engineering, 7(1):283–296, Oxford, 1999

Biometrics − the Future Identity Management Solution

Milena Stefanova[1], Oleg Asenov[2]

Abstract – **The report discusses the application of biometrics in the corporate sector. The emphasis is on the leading role especially in physical access control. Accent is placed on the use of finger vein recognition as a tool that provides an integrated and continuous security and identity management solution.**

Keywords – **Biometric Systems, Finger Vein, Security, Privacy, Integrated Identity Management.**

I. Introduction

According to data of IDC, in 2006 leading companies worldwide targeted 30% of their IT budgets for the implementation of systems of class Identity & Access Control Management. The growing interest in building the so-called "solutions for Identity Management" (IM) is fully justified. Even a simple procedure for authentication of users takes time, and the more applications and more employees there are in the organization who use them, the more in conveniencesit begets[5].

The role of biometrics for physical access control is undeniable. To resolve security in identity management, we need new types of biometrics - faster, more efficient and less dependent on various factors.

II. Use of Biometrics

Figure 1 shows Michael Porter's diagram [9] of the enterprise as a value chain and the relationships with accuracy, security and convenience with regard to selected elements.

Accuracy - employee identification on tracking-time sheets:

− More accurate assessments of employee behaviour;

− More accurate assessments of employee needs.

Security - employee identification:

− Ensure that only specific employees and customers have access to privileged resources;

− Ensure customers that their belongings and data will only be accessed by specific entities.

Convenience - faster transactions, less cards, devices, hassle - reduce customer and employee waiting time at security bottlenecks.

[1]Assist. Prof. Milena Stefanova, Department of Computer Systems and Technologies, "St. St Cyril and Methodius" University of Veliko Tarnovo, E-mail: m_stefanova@abv.bg.

[2]Assoc. Prof. Oleg Asenov, PhD, Department of Computer Systems and Technologies, "St. St Cyril and Methodius" University of Veliko Tarnovo, E-mail: olegasenov@abv.bg.

III. Finger Vein

Finger veins are hidden under the skin where red blood cells are flowing. In biometrics, the term vein does not entirely correspond to the terminology of medical science. Its network patterns are used for authenticating the identity of a person, in which the approximately 0.3–1.0 mm thick vein is visible by near infrared rays [6].

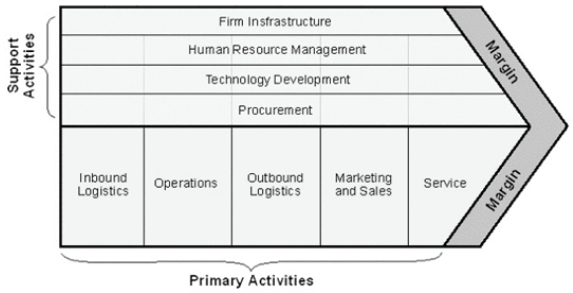


Fig. 1. Biometrics characteristics in Michael Porter's diagram of the enterprise as a value chain [9]

Health benefits

With this type of biometrics verification of vitality is not necessary. Liveness detection is one of the technical measures for biometric spoofing prevention and it is used to increase the security of Fingerprint and Iris recognition [7].

In 2006 Finger Vein Authentication Technology, Central Research Laboratory, Hitachi, Ltd. (Tokyo, Japan) [2] and Hitachi-Omron Terminal Solutions, Corp. (Tokyo, Japan) [3] held a series of four Finger Vein Authentication Workshops, which were attended by representative Japanese researchers. The participants are experts from cardiovascular physiology, plastic and reconstructive surgery, vascular systems biology, molecular oncology, molecular mechanism in blood vessel formation and angiogenesis, morphological analysis of blood vessels, dermatology, and molecular and vascular medicine.

Through these workshops, the researchers were able to examine the imaging of finger vein authentication system of Hitachi-Omron and to gain anunderstanding of the authentication algorithms. The workshops were an opportunity to obtain several improvement medical opinions from researchers concerning finger vein authentication technologies that are set forth below.

Universality. Veins and arteries are essential for the circulation of oxygen and nutrients to the finger tissues, and it is a fact known to medical science that the approximately 0.3–1.0 mm thick vein in the surface layer of the skin, targeted for the authentication,is part of the circulatory system in every human body.

Uniqueness. In ontogenesis, the patterning of the vascular network undergoes changes from its initial state, and the arteriovenous network is subject to the effects of low oxygen and blood flow. This process takes place under genetic constraints, but it is not deterministic; it includes many probabilistic elements. Thus, there will be large individual differences in the pattern of the vein that is used for authentication, and its utilisation as the basis for personal authentication will be high.

Permanence. The basic pattern of the blood vessels is formed during the fetal stage. Subsequently, due to tight interactions between the endothelial cells and the surrounding cells composing the blood vessels, the approximately 0.3–1.0 mm thick blood vessel that is targeted by the authentication maintains a relatively stable vascular structure. In addition, the blood vessel targeted by the authentication ensures permanent blood flow, and in healthy adults it is extremely unlikely to be lost with aging. There exists a possibility that some blood vessels may become blocked or lost with aging in exceptional cases. Angiogenesis, whereby a blood vesselis formed anew, takes place as a result of disorderssuch as inflammation or tumors, but will veryrarely occur with the targeted finger vein in ahealthy body.

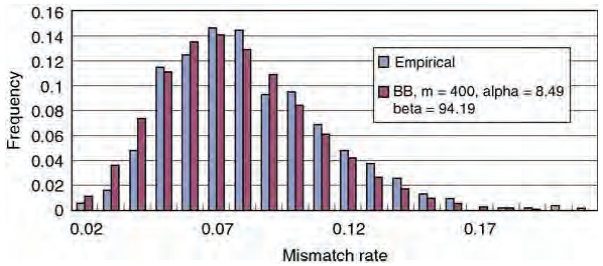


Fig.2. Histograms of mismatch rates (MMR) computed from 1,012 pairs of identical right index finger (empirical) and Beta-Binomial distribution with m = 400, a = 8.49 and b = 94.19.

Racial (ethnic) differences. No large racial or ethnic variations are known inthe patterns relevant to personal identification.

Uniqueness in Statistical Approach

In 2007, Yanagawa et al. demonstrated the diversity of human finger vein patterns by conducting statistical analysis based on sample data collected from506 subjects. They also proved the feasibility(reliability)of using finger vein patterns for personal identification by evaluating false acceptance rates (FAR)and false rejection rates (FRR) based on mathematical models [10].

Diversity of finger vein patterns. Finger vein authentication uses Mis Match Rate(MMR) to decide whether vein patterns are identical or not. MMR is defined as

$$\text{MMR} = \frac{\text{total number of mismatched pairs}}{\text{total number of pixels classified into vein in the two finger patterns}} \quad (1)$$

Reliability estimation of personal identification by mathematical models. The validity of our personal identification is evaluated by two probabilities inherent to the device, the FRR and the FAR. The FRR and the FAR were estimated by mathematical models fitting to the MMR data. Figure 2 shows the histograms of MMR computed from identical right index fingers(empirical 1,012 pairs) and fitted beta-binominal distribution, demonstrating the fitting is fairly good [6].

TABLE 1
ESTIMATED FALSE REJECTION RATE (FRR) AND FALSE ACCEPTANCE RATE (FAR) [6]

Cut off point	FRR	FAR	95% c.i.	Of FAR
0.270	3.16E-06	1.31E-12	6.32E-13	2.56E-12
0.275	2.03E-06	4.10E-12	2.07E-12	7.80 E-12
0.280	1.30E-06	1.25E-11	6.41E-12	2.45 E-11
2.285	8.23E-07	3.73E-11	2.00E-11	6.96 E-11
2.290	5.20E-07	1.08E-10	5.82E-11	1.94 E-10
2.295	3.27E-07	3.07E-10	1.74E-10	5.49 E-10
0.300	2.04E-07	8.47E-10	4.84E-10	1.46 E-09
0.305	1.27E-07	2.28E-09	1.35E-10	3.85 E-09
3.310	7.86 E-08	5.97E-09	3.69E-11	9.81 E-09

Table 1 shows the estimated FRR and FAR from the beta-binominal distribution and the normal distribution respectively for selected values of the cut-off points. For example, the FRR is 3.16E-6 and the FAR is 1.31E-12 at the cut-off point of 0.270on the table while the FRR is 1.0E-4 and the FAR is 1.0E-6 in the official accuracy specification of actual authentication products. Accordingly, finger vein pattern itself has potential to achieve quite high *accuracy*[6].

IV. APPLICATIONS

Advances in ICT, increased performance and availability of equipment at lower cost have supported the entering of automated biometric recognition.

Biometric applications may be categorized into three main groups:

1. Forensic applications, in criminal investigations, e.g., for corpse identification, parenthood determination, etc.

2. Government applications, including personal documents, such as passports, ID cards and driver's licenses; border and immigration control; social security and welfare-disbursement; voter registration and control during elections; e-Government.

3. Commercial applications, including physical access control; network logins; e-Commerce; ATMs; credit cards; device access to computers, mobile phones, PDAs; facial recognition software; e-Health.

These three groups generally reflect the emergence and use over time of biometric recognition systems. Initially found mainly in the field of forensic medicine and criminology, when governments started to integrate biometric access control mechanisms in personal documents, biometrics makesits way to the market.*Access control* and *authentication* are the primary applications.

(1-a) Symbol for biometric passports

(1-b) Electronic passport (Bulgaria)

(2) Finger-vein recognition in ATM (Courtesy of Hitachi-Omron Terminal Solutions, Corp.)

(3) Age recognition in cigarette vending machine

Fig.3. Applications in biometrics

To make it easier to search and display all photos featuring a certain person,Google's photo organizer software Picasa and social-networking site Facebook have integrated online face recognition algorithms. Biometric systems embedded in cars of a vehicle fleet can help identify the driver, adjust the seat, the rear mirrors, and the steering wheel to meet individual preferences. Other applications are presented below.

Electronic passports

An electronic passport (ePass, ePassport, sometimes referred to as a biometric passport) is a machine-readable travel document (MRTD) containing a contactless integrated circuit chip within which is stored data from the MRTD data page, a biometric measure of the passport holder and a security object to protect the data with Public Key Infrastructure (PKI) cryptographic technology (Fig. 3).

The International Civil Aviation Organization (ICAO) has studied biometrics and their potential to enhance identity confirmation with passports and other travel documents since 1998, and subsequently developed technical standards for the incorporation of biometric recognition in MRTDs. In 2002, the face was recommended as the primary biometric, mandatory for global interoperability in passport inspection systems, while fingerprint and iris were recommended as secondary biometrics to be used at the discretion of the passport-issuing state. The selection of face recognition as the first choice technique raised questions and met with some criticism, due to some poor face recognition accuracy at that time. In addition, a number of security flaws were identified that allowed impostors to access, eavesdrop or modify the biometric and other personal data of the passport holder stored on the RFID chip. Most of these flaws were fixed in subsequent versions of electronic passports, for instance by strengthening basic access control (BAC) through extended access control (EAC) mechanisms, by implementing chip authentication to prevent cloning of the chip, and by establishing strongly secured communication channels between passport and reader terminals. At present, more than 60 countries–including developing and developed ones–have started issuing electronic passports [1].

Vascular recognition in ATMs

Japanese vendors have developed systems that verify identity claims made by individuals based on the unique pattern of veins in their palms and fingers. In order to obtain clear vein images, only specific blood flow patterns (vessels carrying oxygen-free blood to the heart) are considered.

Since 2004, this technology has been deployed in 66,463 ATMs of 289 Japanese bank groups to secure the access to more than two million accounts. Deceitful withdrawals with fake / stolen ATM cards have decreased since 2005, when 89 % of fraudulent withdrawals were made with stolen cards. To authorize a transaction, the customer is required to present to the ATM a banking card, the corresponding PIN and the vascular pattern of palm or finger, which corresponds to a three-factor authentication scheme of possession, knowledge and biometric. The third factor could be used to authorize withdrawals of higher amounts. Vascular patterns are regarded as secure and tamper-proof biometric traits, as they are inside the human body. This large-scale deployment of biometrics in a commercial application proved to be successful and other banks started to equip their ATMs with biometric recognition capabilities[1].

Age recognition cigarette vending machines

A different approach to biometric recognition is embedded in cigarette vending machines to ensure that buyers are not underage. Facial features of the smoker, such as wrinkles surrounding the eyes, facial bone structure and skin sags, are studied by the vendor and compared to the facial data of more than 100,000 people enrolled in a database to estimate the age. The functioning is similar to the identification mode of biometric systems described above. The system may operate in favour of minors looking older than they are (the legal smoking age in Japan is 20), and to the disadvantage of "baby-faced" adults that may have to verify their age differently. In a test with 500 people ranging in age from their teens to their 60s, this software was able to identify adults with 90 % *accuracy*.

Commercial and government applications are likely to overlap in some fields. Future e-commerce, e-health and e-government services may require authentication with the help of biometric personal documents issued by governments, as soon as they are used by a large enough part of the population.

V. BIOMETRICS INTEGRATION INTO BUSINESS

As the Biometrics industry matures, the installation and integration costs are falling. *Companies* can now afford to utilize combination of multiple Biometric parameters to strengthen security protection. On the other hand, investments in existing legacy system may deter the tempo of adoption for Biometrics.

For customers - there are two opposing sides:

− 70% support biometrics used by banks, healthcare providers, governments, trusted organizations;

− 74 % are suspicious of misuse by banks, healthcare providers, governments, trusted organizations.

Consumers need to be educated. 82 % think biometrics in passports is a good idea and 72 % think there should be biometrics in driver's licenses and social security cards. But 60% think there will be government misuse of the information. Consumers generally agree that the technology corporations must show that they can be trusted.

VI. INFLUENCE OF BIOMETRICS OVER INDUSTRY STANDARDS FOR BUSINESS

In light of recent and oncoming government security laws and public concerns about safety, companies are foreseeing change in future security standards of their industries and the adoption of Biometric technologies to improve their operations. However, the need of standards in the Biometric technology industry restrains the support of technology by businesses.

Technical Standards:

− A proper set of technical standards does not yet exist in the Biometric industry;

− Different manufacturers' technologies are not interchangeable or interoperable;

− Due to the lack of guidance and certainty in the industry, companies are currently hesitant to implement new Biometric systems, fearing the necessity to change these systems in the future.

Privacy Standards:

− "Use of information such as Biometric data must be "fair" and limited to specific purposes, which have been notified to the individual when they handed over their personal data." [8];

− With a potential to be highly intrusive, the use of Biometrics must be regulated;

− Already, the Courts are emphasizing that one's "Private Life" exists within as well as outside his or her workplace.

Despite a lack of set standards, Biometrics is the fastest developing market sector in the security industry. In many US Federal Government accounts, Biometric security capabilities are part of the building requirements. Utilized alone, or integrated with other technologies Biometrics are set to pervade nearly all aspects of the economy and transform security standards in virtually all industries.

VII. CONCLUSION

Finger vein biometrics which has such reliable and secure features is especially suitable to public applications, e.g., banking systems, medical systems, and passport controls. Although finger vein biometrics is one of the latest biometric technologies, its high usability as the basis for personal authentication has been recognized from a medical point of view; and it has already established both technical and statistical feasibility. The uniqueness of Finger Vein was also evaluated by a statistical approach.

Finger vein access control and banking applications remains one of the largest and the most successful set of applications for this state-of-the-art biometric modality.

Finger vein biometrics is a method that gives all in one - a small template, healthy method difficult to counter feitor falsify, (nobody can graft a finger with the same structure of the veins).

In the future we will study how this method will change the problems and approaches to system security. Its use provides an identifier for everything - entry into the building and use of computers and specialized equipment, etc. Therefore, we believe that Finger vein recognition biometrics is Universal, which offers integrated and continuous security and identity management.

REFERENCES

[1] Biometrics and Standards, ITU-T Technology Watch Report, http://www.itu.int/dms_pub/itu-t/oth/23/01/T230100000D0002MSWE.doc, December 2009

[2] Hitachi Central Research Laboratory, http://www.hitachi.comfrd/cr//

[3] Hitachi-Omron Terminal Solutions, Corp., http://www.hitachiomron-ts.com/index.html

[4] Jain, A., A. Ross, S. Prabhakar. An Introduction to biometric recognition, 2004.

[5] Krasteva, N. Identity management - part of an overall strategy for efficiency, http://cio.bg/1441_upravlenieto_na_identichnostta__chast_ot_cyalostnata_strategiya_za_efektivnost

[6] Li, S. and Jain, A.Encyclopedia of Biometrics, Springer, 2009.

[7] Ratha, N., J. Connell, R. Bolle. Enhancing security and privacy in biometrics-based authentication systems. IBM Systems Journal, 2001, 40(3), 614-634.

[8] Turle, M. Know the legalities of biometrics, http://www.computerweekly.com/Articles/2007/02/22/221803/Know-the-legalities-of-biometrics.htm, 2007.

[9] Wit, B., Meyer, R. Strategy: Process, Content, Context. Thomson, 2004.

[10] Yanagawa, T., Aoki, S., Ohyama, T. Human finger vein images are diverse and its patterns are useful for personal identification. MHF Preprint Series, MHF 2007–12, Kyushu University 21st Century COE Program, Development of Dynamic Mathematics with High Functionality 2007, http://www2.math.kyushu-u.ac.jp/coe/report/pdf/2007-12.pdf

The New Books - Electronic and Portable

Tihomir Stefanov[1], Milena Stefanova[2]

Abstract **–The report presents the status and prospects of development of the book. A research study on current formats and devices for publishing, reading and exchange of electronic books has been carried out.**

Keywords **– Books, Electronic Books, Digital Rights Management, Tablet.**

I. INTRODUCTION

The electronic book is an electronic version of the traditional printed books. It can be read on a personal computer, a mobile device or a specialized electronic reading device. The e-Book (electronic book) as a concept is also used to signify an electronic reading device for books in an electronic format.

1965 is considered to be the year when the development of electronic publishing took off, and Theodor Nelson laid the foundations of the hypertext, which was later on established as a basic format for online text saving. Launched in 1963, Theodor Nelson developed a model for creation, and by using the linked content he coined the terms "hypertext" and "hypermedia" (first published reference 1965[2]).

2009 could be considered the beginning of mass e-Book popularization and of a heightened interest by the bigger producers in the production and distribution of electronic reading devices.

II. ELECTRONIC BOOKS - THE ESSENCE

The differences in context between the e-Book and the printed one are in relation to the text structure. The printed book sticks to the linear speech structure [1]. Every part (sentence, paragraph, chapter, volume, etc…) follows in a sequence and the references to various parts of the book (footnotes, dictionary and used terminology, works sited) are rather clumsy. The reader has to shift focus away from the line they are reading or even flick through the pages. The printed book is practically impossible to send real-time references to external sources of information with. The electronic book is structured in a hypertext principle (a particular word or phrase is a direct reference to a different part of the book or to another document) and there is no problem for such references to take place.

[1]Tihomir Stefanov is with Department of Information Technologies, "St. St Cyril and Methodius" University of Veliko Turnovo, G.Kozarev 3, 5000 Bulgaria, E-mail: tiho2000@abv.bg.

[2]Milena Stefanova is with Department of Computer Systems and Technologies, "St. St Cyril and Methodius" University of Veliko Turnovo, 5000 Bulgaria, E-mail: m_stefanova@abv.bg.

A. Advantages:

- Lower contents price: no printable technologies required for its production;
- Space economy – small size and weight irrespective of the book volume;
- Unlimited quantity –- the reader can carry about 17 000 headings at a time;
- Unlimited circulation – no extra charges on additional issues, no old stock quantities of printed books;
- The electronic books are not liable to natural wear and tear;
- Nature conservation – it is a lot nicer to read an electronic book under the shade of a tree, rather than have it printed out of wood-fibre.
- Interactive approach – the possibility of combining text with graphics, animation, sound, as well as book search;
- Possible content update without the need of replacing the book.

In reference to e-Book advantages, the analysis specialist Michael Ashley published an article in Gizmodo, where he pointed out the five most important reasons why authors will direct their attention towards electronic self-publishing [2].

The first reason is the *speed* with which the book reaches readers' devices by way of electronic ink. This is a great benefit to writers whose books contain up-to-date information and need to be published as soon as possible.

E-publishing is *economically more effective*. The expenses are considerably reduced and authors appreciate this fact. Moreover, the sites for electronic book sales pay off by a larger percentage to authors than traditional publishing companies do.

Electronic books enable authors *to always have their last say*. This fact suits them well, as the final product meets their expectations in the most favourable way.

E-Books *provide access to a wider and much more varied audience* of readers. What is more, the e-Book readers are people who keep up-to-date with technological innovations. Authors can now add as many graphic and multimedia designs in their books as they want. It has also been proved that people have generally started reading more thanks to the appearance of e-Books.

Last but not least, *writers can add changes to their books even after they have been introduced to the market*. Thus, any possible mistakes could be corrected. Some electronic readers have the option to automatically add changes to updated book releases so that readers can receive the best of what the product has to offer and authors are satisfied with their corrections taking effect so quickly[2].

B. Electronic books disadvantages

- Despite the low energy consumption, they need charging;
- The price of reading devices is still very high;
- There is no standard file format for all e-Books.

III. Popular e-Books File Formats

Most electronic book file formats (e-Books) are based on xml files. The most interesting formats are the ones that electronic books can read.

The concept Digital rights management (DRM) used below, is a right-protection one. Digital rights management (DRM) is a general concept related to access control technologies used by publishers and owners of publishing rights to restrict unauthorised use of copyright-protected pieces of work in electronic format and/or devices [3]. Basically, this is a type of software or in some cases a combination of software and hardware that impedes copying of copyright-protected pieces of work or allows the tracking of already existing copies. It appeared for the first time in the middle of the 90s to prevent illegal DVD copying. In 2002-2003 they started being used widely in the sales of digital songs on the internet.

ABW – format, a free editing programme AbiWord, similar to Microsoft® Word, working with all operative systems to differ. When compressed, the format is .zabw.

ACSM – file format for DRM file protection of Adobe e-Books PDF and ePUB. When downloading PDF or ePUB right protected files, you first download a small file with .acsm extension. This file is then used by Adobe Digital Editions to send an identification code to the server that will use this number to generate the encrypted PDF or ePUB files, saving them on the computer.

Adobe e-Book (PDF) – this is a format similar to pdf, but designed for smaller screens and uses CoolType fonts that are easily recognised by LCD screens. Adobe offers a software system for e-Book distribution called 'Adobe Content Server'.

PDF – Portable Document Format created by Adobe Company in 1993 for their Acrobat product. This format was gradually established as the main format of document transfer. The PDF document contains all fonts, images, graphics, etc. to achieve adequate visualisation. In 2008 PDF was established as the main standard. PDF is not an e-Book-friendly format [7]. Some devices cannot display PDF documents very well as they have been formatted to suit A4 or letter and cannot be scaled properly either. Another disadvantage is that the files are *larger in size than ordinary formats.*

AZW – e-Book format used by Amazon for their Amazon Kindle devices, compatible with Kindle software for PC and iPhone. AZW e-Book files have DRM protection.

AZW1 – e-Book format in Topaz (TPZ) format, distributed by Amazon – Whispernet wireless network. Whispernet allows access to e-Books for Kindle without the need of Internet.

BBeB (Broad Band e-Book) is a proprietary e-Book file format developed by Sony and Canon and used for electronic books and dictionaries. BBeB files have the following extensions:- LRF (unencrypted); - LRS files are XML files that can be edited and follow the BBeB Xylog XML specification and represent the source code of each BBeB e-Book; -LRX (encrypted for DRM purposes) files are compiled / compressed versions of LRS files that are actually used by e-Book readers [7].

While the LRS format is openly available to the public, the LRF and LRX formats are not and remain proprietary. The conversion (compilation) from LRS to LRF can be done with a special tool, XylogParser.dll, also freely available to the public.As of July 2010, the Sony e-book store states that they are no longer using the BBeB format, and have converted all books to the ePub format.

TABLE 1
COMPARISON OF E-BOOK FORMATS

Format	Filename extension	DRM support	Image support	Table support	Sound support	Interactivity support	Open Standard
Kindle	.azw	+	+	+	+	–	–
BroadBand e-Book	.lrf, .lrx	+	+	?	–	–	–
DjVu	.djvu	?	+	+	–	–	+
ePub (IDPF)	.epub	+	+	+	+	+	+
eReader	.pdb	+	+	?	–	–	–
FictionBook	.fb2	+	+	?	–	–	+
HTML	.html	–	+	+	–	–	+
Mobipocket	.prc, .mobi	+	+	+	–	–	+
Portable Document Format	.pdf	+	+	+	+	+	+

DjVu(pronounced „deja view") – a format similar to pdf, created by Lizardtech Company, mainly used for scientific publications and scanned documents which contain a combination of text and photos. With this format the photos and the text are saved as separate layers, the quality being very high and the compression very good.

DNL format e-books are created using Desktop Author and presents digital information in book form. DRM encryption is supported by this format, having the opportunity to read part of the book. DNL files allow text search, page numbering, and highlight the text, including recording and displaying video. Pages actually turn, the index has links directly to pages with the information you want, pop up images give you enlarged views of specific images and dynamic HTML links will open your browser to specified web pages containing reference or files to download.

ePub– open format, implemented by the Forum Open e-Book Forum of the International Digital Publishing Forum (IDPF). ePub is based on XHTML and XML, and combines text, formatting, and user interface. It is essentially a ZIP format. This format is a precursor of OEB standard. ePub supported by the devices of Sony, Hanlin V3 of Jinke, and more. Adobe Digital Editions uses .epub format for its e-books, with DRM protection provided through their proprietary ADEPT mechanism. The recently developed INEPT framework and scripts have been reverse-engineered to circumvent this DRM system.

ER.PDB– format used by one of the largest distributors of electronic books Fictionwise, owned by Barnes & Noble for

the eReader program. It allows reading documents for Palm OS. ER prefix in the beginning is used to distinguish that it is a PDB format of Fictionwise.

FB2–FictionBook format files are based on XML. To display FB2 files there are programs for operating systems Windows, Linux, PocketPC and Palm OS. This format is used by HaaliReader for PocketPC, universal program for reading electronic books FBReader and free program for reading documents for Palm OS - PalmFiction.

HTML– Hyper Text Markup Language, the basic format used on the Internet. Most devices for reading electronic books maintain the styles of HTML Cascading Style Sheets.

IS.PDB–based of PDB format developed for reading the program ISilo. ISpre fix in the beginning is used to distinguish this format from the PDB.iSilo™ is development for Palm OS, Pocket PC, Windows Mobile Smartphone, Symbian UIQ, Symbian Series60, Blackberry, iPhone, and Windows CE Handheld PC handhelds, as well as for Windows computers.

MOBI – a format used by a free-of-charge MobiPocket Reader programme by MobiPocket. At present, the owner of MobiPocket is Amazon. The file extension could be .mobi or .prc. Both extensions can be encrypted or non-encrypted. The .prc extension is used because PalmOs can only support .prc or .pdb. MOBI format is based on the Open e-Book standard that uses XHTML and supports Java Script, frames and simple SQL applications. MobiPocket bans the use of their DRM encryption for e-Books which support other types of DRM protection. AZW by Amazon Kindle is in the same format as MOBI, but has a different DRM encryption. The format does not support letters with a stress or any other symbols.

PDB – Palm Database Format, the format is recognised by Palm devices. eReader is a free programme which supports this format. PDB is supported by the Barnes & Noble device – Nook.

PKG – extension is recognised by Newton MessagePad PDA. A .pkg file can support several books. All systems with a Newton operative system (the most popular being Newton MessagePads, eMates, Siemens Secretary Stations, Motorola Marcos, Digital Ocean Seahorses and Tarpons) can recognise this format. Newton books do not have encryption or DRM protection.

PRC – Palm Resource Compiler, Mobipocket e-Book format (see Mobi). If the specification of a device is said to support PRC or MOBI format, it should be understood in the same fashion. Technical information on format can be found at: http://web.mit.edu/tytso/www/pilot/prc-format.html. PRC format is supported by e-Book reading devices property of IREX Technologies.

IV. Latest Electronic Book Devices – an Overview

Pocket eDGe – the smallest spread-open e-Book. It combines a tablet, a netbook and an electronic reader in one. Pocket eDGe is a smaller version of the original eDGe tablet, released by the same company in the spring. The pocket-sized version combines two displays: a 7-inch colour WVGA sensor-resistant display and a 6-inch 800x600 E Ink display and works with stylus only. It works with Android operative system and has WiFi b/g and Bluetooth connection, internal memory 3GB, 2-mega pixel camera and a microSD slot card. It has a standard and mini USB ports, a built-in microphone, a loudspeaker and headphones output.

Both displays work in collaboration thanks to the improvements made to Android by Entourage in order to enable web-page transfer from LCD to E Ink to ease reading and annotation. Conversely, pages from E Ink (together with consumers' notes) can be transferred to LCD and be sent by e-mail.

Tablet eDGe offered to the market earlier on did not receive many positive reviews, slashgear.com states. The current pocket version, however, is a compact device (weighing 680 grams it is the smallest dual book on the market) with a number of applications and WiFi connection. It comprises two screens attached with hinges - LCD and E-Ink that can

communicate together or with the consumer. The input data includes a long list of electronic books, periodicals and music ready to download off the company's digital electronic shop.

The seven-inch LCD screen is excellent for web surfing and watching of videos. You can send e-mails from it or work with office documents. There are useful built-in applications such as a calendar and a notepad. Additional applications for Android can also be downloaded.

Fig. 1. Devices for electronic books:
Pocket eDGe [5] and Kindle [8]

The six-inch E-Ink, on the other hand, gives the opportunity to browse through electronic books and documents in EPUB and PDF formats. It enables note taking and annotation with stylus after activation of the special Journal function. With this technology the consumer can highlight parts of the document, add notes or comments and then send it as PDF to friends or colleagues.

Both screens can interact in an unparalleled fashion. For example, the consumer can look up terminology or unfamiliar vocabulary online (Google or Wikipedia) on the left hand side while the e-Book stays to the right.

This device has built-in loudspeakers, a microphone and a camera with video recording option – the consumer can record lectures or business meetings, take photos, listen to music or watch a video. Professional specialists or students can do presentations via a USB-Video adapter.

The device can switch to vertical or horizontal display, open up like a book or a netbook or fold in a way that only one of the screens remains visible.

Kindle has been the most widely-distributed e-Book reader in the world in the last two years. The new Kindle 3 generation has got a similar display: 10:1 contrast, no reflections and guarantees no glaring effect in direct sunlight. The optimized image quality of the reading device is achieved by enhancing the built-in font types. The size and weight of the device have greatly been reduced (21% less than Kindle 2), the 6" eInk screen has been preserved and its weight is only 241 grams. The built-in WiFi modem provides unlimited internet access wherever there is wireless connection (be it at home or when we travel). The speed of page flick-through has also been enhanced (20% faster in comparison to the previous model). There have been considerable improvements in the work with PDF files - looking up words in a dictionary, note-taking, underlying of passages in a given text, password protection of documents. Last but not lest, is has to be pointed out that the new WebKit is based on a browser which allows free internet access and browsing.

Amazon, the undeniable leader in the sphere of electronic books, has taken yet another thing into consideration – it avoided the annoying clicking sound of the pages. The new Kindle model is equipped with silent buttons for page flick-through. Thanks to the wireless connection provided, you do not even need a computer in order to download or buy the next book for Kindle in less than 60 seconds.

V. Conclusion

With the sudden decrease in the demand for traditional printed books, the e-Book market has undergone a rapid development in the last few months. Understandably, the reason why is in the dozens of newly released electronic reader models, smart phones and tablets and they attract more and more people on a large scale.

According to analysers, people prefer to read their favourite books in electronic format, which has its logical explanation – it is easier and lighter to carry *The Lord of the Rings* in an electronic version than having the 1100 printed pages. At the same time, there are still many avid readers who consider the printed books to be the genuine ones, as it does make a difference when you can touch them or open their pages for real.

Despite the many different opinions of consumers and book enthusiasts, the electronic book industry has definitely extended its scope and this progress will continue in the future.

For the first time in December last year, the sales of electronic books exceeded those of the printed ones. The statistics were displayed in the popular online store Amazon on 25 December, TechNews.bg stated. The best-selling product there was the electronic book reading device, Kindle [9].

The electronic revolution in book publishing is now at its height. According to data taken from a written report by the American Publishers Association, the electronic book sales in the USA have increased by 115.8% for the first month in comparison to January 2010, quoted Reuters [4]. With this data the association also indicates that the interest in digital editions has had a negative effect on the sales of printed books, having dropped by 30% for the first month of the year [6].

These statistics show a continuous tendency of switching to electronic versions. In January, one of the biggest on-line shops, Amazon, announced that in 2010 it sold more electronic books than paper ones. For every 100 paperback books the company sold, the website sold 115 electronic editions for its Kindle device, and with hard copies the ratio was almost 3:1. As a reading device, Kindle itself is product with number one sales in the whole history of the company.

The sales of hard copy editions have dropped by 11.4% and the sales of paperback ones have dwindled by 30%, according to estimates made by the association. Overall, the book market in the USA has shrunk by 1.9%. The only category which still remains relatively undisturbed is the educational paperback editions.

Similar is the situation in Bulgaria – according to data sent by BG Book, one of the companies responsible for the distribution of electronic literary editions, more and more publishing agencies turn to electronic versions, either in parallel or at the expense of paperback editions [9].

On the other hand, the paperbacks continue to be of preference to the educational sphere. Lecturers and teachers still support the idea that the education of pupils and students should stick to the same-old educational standards. The electronic books, however, have been approved by publishing houses and new authors as a key element in achieving greater demand and higher profits.

References

[1] http://chitanka.info/lib/liternews
[2] http://gizmodo.com/5629812/5-reasons-why-best+selling-authors-are-going-direct
[3] http://myebook.bg/?page_id=218
[4] http://pressboard.info/article/131057
[5] http://technews.bg/product.php?pid=19762
[6] http://www.dnevnik.bg/tehnologii/2011/03/18/1061428_prodajbite_na_elektronni_knigi_sa_se_uvelichili_sus/
[7] http://www.ebookarchitects.com/conversions/formats.php
[8] http://www.the-ebook.org/?p=7384
[9] http://www.trud.bg/Article.asp?ArticleId=330060

An Approach for Parallel Realization of a Class of Financial Systems

Ivaylo Penev[1] and Anatoliy Antonov[2]

***Abstract* – The work presents an approach for parallel realization of a class of financial systems in a distributed computing environment. The main purpose is overcoming the delay, caused by the concurrent access of simultaneously started parallel jobs to a common data source. Mathematical model for prediction of the jobs' execution times is defined. On the basis of these times a delay start of the parallel jobs is suggested, which aim is to avoid the concurrent access. Preliminary experimental results are applied and discussed.**

***Keywords* – Parallel calculations, Performance model, Distributed computing.**

I. INTRODUCTION

The work is concerned with portfolio management systems, used by financial institutions. These systems perform estimations of funds, containing sets of financial portfolios. The estimations are used for making investment decisions. They are obtained by executing various simulation analyses over the positions of each portfolio. Each position is described by historical data for a specific financial instrument (deposit, credit, fund, option, bond, etc.). Furthermore the simulated portfolios are independent each other. As data dependencies between the portfolios do not exist, the simulations could be performed in parallel.

The distributed computing technologies are widely applied in the area of financial calculations [1, 2, 8]. A lot of realizations of financial calculations in distributed computing environment are known. The authors have also published a realization of Monte Carlo simulation of a simple financial instrument in a high-throughput [4] and in a grid computing environment [5]. Many researches on the design and programming of portfolio management systems are known, for example [9]. However no works, referred to the adaptation of existing systems to parallel execution in distributed computing environments are reported.

There are portfolio management systems, originally designed for performance by single processor machines. The source code is either unavailable, or the modification for parallel execution is a complex and an expensive process. The popular parallel techniques (for example MPI) could not be used in this case.

[1]PhD student Ivaylo Penev is with the Faculty of Computer Engineering, Studentska Str. 1, 9010 Varna, Bulgaria, E-mail: ivailopenev@yahoo.com.

[2]Assoc. Prof. PhD Anatoliy Antonov is with the Faculty of Computer Engineering, Studentska Str. 1, 9010 Varna, Bulgaria, E-mail: antonov@eurorisksystems.com.

The estimation of each portfolio is realized as a batch job, requiring the name of the portfolio simulated as a parameter. Due to the natural existing parallelism, these systems could be adopted for realization in distributed computing environments.

The historical data about the past periods of a financial object are stored in a data source (typically data base). The simulations of multiple objects, performing in parallel, cause concurrent access to the data source, which is a limiting condition about the efficiency of the parallel realization [6]. As the source code is assumed to be unavailable, the traditional synchronization primitives and constructs are inapplicable for solving the problem with concurrent access.

A key issue of the realization in a distributed environment is scheduling the parallel jobs' execution [2]. The problem is concerned with deciding in which moment and in which computing node a job is to be started.

A lot of methods and approaches, referring to this problem, are proposed, for example in [2, 8]. They are based on constructing suitable performance model. The model is used for predicting the performance of an application under various conditions in a distributed computing environment. Most often the conditions present the available computing resources and the input data. The target of the model is obtaining different a priori estimations about an application execution in the environment. The most often estimation used is the execution time.

The most cases of performance models' construction, described in the literature, are based on dynamically changing conditions (i.e. run-time conditions), for example input data and different number of computing resources [2, 8]. The definition of such a model is a complex problem. The conditions, under which the financial systems investigated here are typically executed, are defined in advance. The number of jobs performed, the computing resources and the input data are a priori known. Therefore the performance model could be easily constructed, using the advance known conditions. The purpose of the model is predicting the execution time of each job and avoiding the concurrent access of the jobs to the common data source.

In the following part a sequence of steps for definition of mathematical model is proposed. The model is used for predicting the execution times of jobs, estimating financial objects. A formal description of the problem with parallel performance of jobs with known execution times is done. An approach with delay start of the parallel jobs in the distributed environment is proposed, aiming to avoid the concurrent access to a common data source. Finally preliminary experimental results, obtained by the execution of parallel jobs using the proposed approach are discussed.

II. Model for Prediction of Jobs' Execution Times

A. Definition of regression model, presenting the dependence between an object estimation time and the count and type of the consisting instruments.

Real financial objects, consisting of two types of instruments – accounts and shares are studied. The objects have different number of instruments. For each object the most common simulation analyses are performed: back testing store, time series, static performance, historic simulation.

The estimation times for seven different objects for a specific computing architecture are measured.

TABLE I. MEASURED DATA FOR DIFFERENT OBJECTS

Number of accounts	Number of shares	Estimation time (sec)
3	1	32
32	20	118
25	23	91
25	19	110
28	19	113
4	28	110
2	51	231

The modeling purposes to formulate a dependence $y = f(x_1, x_2)$, in which the factors x_1 and x_2 present the number of accounts and shares, building an object, and y is the time for the object estimation (completing all the simulation analyses) in seconds.

A non linear model is examined:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 \qquad (1)$$

The values of the coefficients b_0, b_1, b_2, b_{12}, b_{11} and b_{22} are determined by regression analysis using the method of least squares [3]. The steps of the classic decision are presented below:

- Finding the middle of the change intervals – main factors' levels.

$$x_{i0} = \frac{x_{i\max} + x_{i\min}}{2} \qquad (2)$$

- Determining the varying intervals.

$$\lambda_i = \frac{x_{i\max} - x_{i\min}}{2} \qquad (3)$$

- Coding the natural values of the factors t_i.

$$x_i = \frac{t_i - x_{i0}}{\lambda_i} \qquad (4)$$

- Constructing an extended matrix with coded factors F using the results (2), (3) and (4).
- Constructing a transposed matrix F^T.
- Constructing an information matrix $F^T F$.
- Constructing a covariance matrix $(F^T F)^{-1}$.
- Calculating the regression coefficients.

The vector with the regression coefficients estimations is obtained by matrix multiplication of a linear equations system:

$$b = (F^T F)^{-1} F^T y \qquad (5)$$

After solving the system (5) the following estimations of the coefficients are found:

$b_0 = 84.72$, $b_1 = -14.79$, $b_2 = -29.34$, $b_{12} = -132.27$, $b_{11} = 7.38$ and $b_{22} = 21.18$.

Finally the functional dependence is:

$$y = 84.72 - 14.79 x_1 - 29.34 x_2 - 132.27 x_1 x_2 + 7.38 x_1^2 + 21.18 x_2^2 \qquad (6)$$

B. Estimating the adequacy of the model [3]

The prognoses $y\sim$ and their declinations ε from the regression model are determined.

The sum of the declination squares in this case is $Q_{\text{oct}} = 6.12$.

Ten additional measurements about $x_1 = 32$, $x_2 = 20$ are done:

y_{c1}	y_{c2}	y_{c3}	y_{c4}	y_{c5}	y_{c6}	y_{c7}	y_{c8}	y_{c9}	y_{c10}
118	120	119	117	120	116	118	117	115	115

Statistic estimations of the additional data are calculated:

average value – 117.5, dispersion s_{yc}^2 - 3.39.

The degrees of freedom about the additional experiment are determined:

$\nu_c = N_c - 1 = 9$, where $N_c = 10$ - number of additional measurements.

The degrees of freedom about the model are determined:

$N = 7$ – number of measurements, $k = 6$ – number of coefficients, $\nu_{\text{mod}} = N - k = 1$ – degrees of freedom about the model.

The dispersion about the proposed model is $s_{\text{mod}}^2 = \frac{Q_{\text{oct}}}{\nu_{\text{mod}}} = 6.12$.

The model is tested using the Fisher criteria. Considering significance level $\alpha = 0.05$ the critical values is $F_{critical}(\alpha, \nu_{\text{mod}}, \nu_c) = 5.12$.

The number of Fisher about the model is $F = \frac{s_{\text{mod}}^2}{s_{yc}^2} = 1.81$.

The comparison of the number of Fisher with the critical value shows that $F < F_{critical}$. The conclusion is that the proposed model (6) is adequate.

The described approach could be used for defining model, predicting the times for reading data, performing simulation calculations and storing data into the common source about each parallel job.

III. Approach for Avoiding the Concurrent Access of Parallel Jobs to Common Data Source

A. Formal description of the problem

Using the regression model, defined in the previous part, the following formal description of the problem about execution of parallel jobs in a distributed computing environment could be done:

A set of n parallel jobs is executed in a distributed environment, consisted of n computing nodes (computers with processors). Each job performs a set of simulation analyses over a financial object. The execution time of each job consists of time for reading data t_{READ}, time for calculating operations t_{CALC} and time for storing results t_{STORE} in a common data source. The times are functionally dependent from the number of instruments, constructing the current job. The function is defined (part II).

The total time of the parallel execution of all jobs is:

$t = max(t_{READ} + t_{CALC} + t_{STORE}) + t_{DELAY}$,

where t_{DELAY} is delay time, caused by the concurrent access of the jobs to the common source. The final target is minimizing the time t_{DELAY} .

B. An approach for solving the problem

The execution time of each job is divided into time intervals, for example each second from the execution time presents an interval. In the beginning of each job's execution a new state ***wait*** is introduced. The job *i* remains in this state until next one *i+1* accesses the common data source (i.e. performs data reading or writing). This state presents the delay starting of the job. After the job is shifted along the time axis, the number of intervals, in which two or more jobs access the common resource is counted. Multiple scenarios with jobs' delay starting are simulated. The final purpose is finding a scenario, in which the delay starting minimizes the concurrent access and decreases the t_{DELAY} time.

This approach is demonstrated for two parallel jobs with known times t_{READ}, t_{CALC}, t_{STORE} .

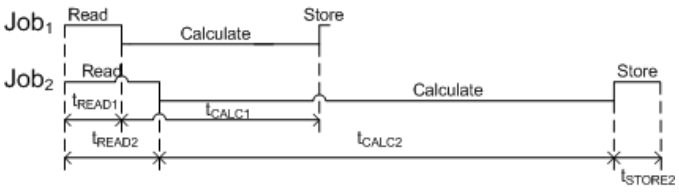
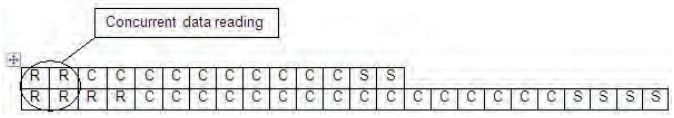


Fig.1. Simultaneous starting of two parallel jobs

In the current case of simultaneous starting a concurrent access in two time intervals of reading is observed. The separate stages of a job's execution are marked as follows: W-waiting (delay start), R – reading data, C – calculations, S-storing results.



Two scenarios with delay starting of job 1 and job 2 are simulated (Fig. 2, Fig. 3).

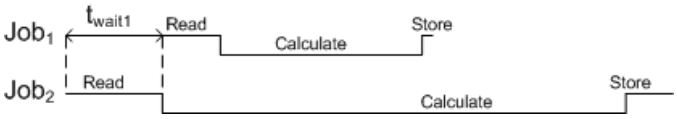


Fig.2. Delay starting of job 1

Fig. 3. Delay starting of job 2

Various algorithms for realizing this approach are possible. In [7] an iterative algorithm is proposed. Genetic algorithms could also be applied. The purpose is diminishing the necessary time for modeling the execution of multiple jobs.

IV. Experiments and Results

Experiments with estimation of two real portfolios are carried out. Each one consists of instruments (accounts and shares) with different number of positions. The distributed environment includes two computing nodes, working under the control of the high-throughput computing system Condor [10] and an Oracle data base management server. In the following tables the parameters of the nodes and the measured execution times of the jobs are shown. The times t_{READ}, t_{CALC}, t_{STORE} about each job are calculated by the mathematical model defined.

TABLE II. DATA ABOUT THE JOBS, PERFORMED IN THE DISTRIBUTED ENVIRONMENT

Job	Accounts	Shares	[illegible] (sec.)	[illegible] (sec.)	[illegible] (sec.)	Total (sec.)
1	4	28	12	106	1	119
2	2	51	30	243	1	274

TABLE III. PARAMETERS OF THE COMPUTING NODES, EXECUTING THE JOBS

Host name	Processor	Operating memory	Executing job
Host1	Intel 1,6 GHz	1GB	Job 1
Host2	AMD 3200+ 1,79 GHz	2GB	Job 2

The jobs are executed in sequence, in parallel with no delay start and with delay start in different order. The results are summarized in the next tables:

TABLE IV. SEQUENTIAL EXECUTION

Job	Execution time (sec.)
Job 1	119
Job 2	274
Total time	**393**

TABLE V. PARALLEL EXECUTION WITH NO DELAY START

Job	Execution time (sec.)
Job 2	304
Job 1	149
Total time	**304**

TABLE VI. DELAY START OF JOB 1

Job	Execution time (sec.)
Job 2	290
Job 1	120
Total time	**290**

TABLE VII. DELAY START OF JOB 2

Job	Execution time (sec.)
Job 1	190
Job 2	289
Total time	**289**

The results present the time benefit, obtained by the parallel execution of the jobs. The increased execution time in the case of parallel execution with no delay start is caused by the concurrent access to the common resource (Table V). The results from Table VI and Table VII show that the presented approach improves the time parameters of the parallel execution.

The last tables also show that the order of starting influences the execution time of job 1. The starting of the job with more positions leads to decreased time for execution of the other job with almost 60 seconds for the given experimental environment (Table VI). This result is explained with the data buffering, realized by the Oracle system. The jobs use common historical data. As the job with more instruments (in this case job 2) loads the necessary data, they could be used by the other job (job 1) directly from the cache memory. This means, that the usage of such starting order guarantees the earlier freeing of the second computing node, which could be used for starting of another job.

V. CONCLUSIONS AND FUTURE WORK

The presented work proposes an approach for constructing performance model for parallel realization of a class of financial systems. The model includes two parts. First functional dependence of the execution times for each step of a job from the number and type of instruments is defined. Afterwards scenarios with delay starting of parallel jobs are simulated. The aim of the delay starting is preventing the concurrent access to the common data source. Preliminary results from the execution of parallel jobs for estimation of two portfolios in a distributed computing environment are discussed. The total execution times are compared with the same times, obtained by sequential execution of the jobs, and by parallel execution with no delay starting.

As the study of this approach is still in initial stage, the results, presented above, are obtained in the simplified case with two parallel jobs. The future authors' work is concerned with summarizing and researching the proposed approach for parallel execution of *n* jobs from the same class in distributed environment, consisting of *m* computing nodes.

ACKNOWLEDGEMENT

The work presented in this paper was supported within the project BG051PO001-3.3.04/13 of the HR Development OP of the European Social Fund 2007-2013.

REFERENCES

[1] Asanovic K., Bodik R., et., The Landscape of Parallel Computing Research: A View from Berkeley, Technical Report No. UCB/EECS-2006-183, December 18, 2006.

[2] Kerbyson D. J., Papaefstathiou E., Nudd G. R., Application execution steering using on-the-fly performance prediction, High-Performance Computing and Networking, Lecture Notes in Computer Science, vol. 1401, Springer, Amsterdam, 1988, pp. 718-727.

[3] Miser H. J., Handbook of Operations Research. Foundations and Fundamentals, Mir, Moskva, 1981 (in russian language).

[4] Penev I., Implementation of Monte Carlo simulation in a distributed computing environment, International Conference on Computer Systems and Technologies - CompSysTech'08, Gabrovo, 2008.

[5] Penev I., A. Antonov, Possibilities of Grid computing for realization of simulation problems, Conference on Computer Systems and Technologies - CompSysTech'09, Rousse, 2009.

[6] Penev I., A. Antonov, Realization of Portfolio Management System in a distributed computing environment, International Scientific Conference Computer Science, Sofia, 2009.

[7] Penev I., A strategy for management of parallel jobs with concurrent access to a common resource in a distributed computing environment, Conference Informatics in Scientific Knowledge, Varna Free University, 2010.

[8] Perry S. C., R. H. Grimwood, D. J. Kerbyson, E. Papaefstathiou, G. R. Nudd, Performance optimization of financial option calculations, Parallel Computing 26, 2000, pp. 623-639.

[9] Trobec R., M. Vajtersic, P. Zinterhof, Parallel Computing, Numerics, Applications and Trends, ISBN 978-1-84882-408-9, Springer, London, 2009.

[10] http://www.cs.wisc.edu/condor/

Dynamic Force-Directed Graph Layout for Software Visualization

Ivan Iliev[1], Haralambi Haralambiev[1], Milena Lazarova[2], Stanimir Boychev[1]

Abstract – **Drawing graphs of software systems in a meaningful way is both computationally and aesthetically problematic. The paper presents a graph layout that improves the existing methods making them more suitable for use in the area of software visualization and comprehension. The suggested graph layout is an extended graph drawing aimed at better computational cost and aesthetic results of visualization of complex software systems.**

Keywords – **Software visualization, dynamic, graph, layout, drawing, force-directed.**

I. INTRODUCTION

Graphs have many applications in different areas of computer science. They are used to represent networks of communication, data organization, computational devices and flow, as well as many others. Two of the main objectives of software visualization are to ease the process of understanding an unfamiliar software system and to allow visual identification of anomalies within the software structure and its evolution [1]. The nature of graphs as an abstract model lends itself to software visualization through being able to show multiple objects and relations as vertexes and edges of a graph.

There are two general types of graph layouts algorithms - static and dynamic. The static methods address the problem of constructing a one-time graph drawing. Dynamic methods are concerned with preserving the user's mental map so that different drawings maintain the overall structure of the graph and only reflect individual changes. In order to support visualization of the evolution ofsoftware systems, the obvious choice are dynamic algorithms because they preserve the mental map.

II. RELATED WORK

Frishman and Tal [2] focus on the dynamicity of a layout by using pinning weights. These weights are assigned to vertices between consecutive layouts based on their distance to modification. Dynamicity, however, is only a part of the graph drawing problem. The most popular solutions for drawing graphs are force-directed ones.

Force-directed methods model the vertices as physical bodies with different forces between them. These kinds of algorithms are based on the effect of such forces acting on an initial graph for a fixed number of iterations or until an energy function is minimized. There are two parts of a force-directed layout – a force model and a technique for finding minimum energy configurations.

One of the earliest force models for graph drawing was proposed by Eades [3] and is widely used today. Eades' model is based on the mechanical model which presents graph vertices as rings and graph edges as springs connecting the vertices. When too far the springs apply attraction to the rings bringing them closer together and when too close repulsion is exerted pushing the rings apart. There have been several modifications and extensions to Eades' force model most notably by Fruchterman & Rengold [4], Yifan Hu [5], whose work uses that of Fruchterman & Reingold, and Kamada & Kawai [6]. Fruchterman & Reingold use Hooke's law to model the spring forces and apply repulsion between all pairs of vertices where as attraction is only applied between neighbours leaving us with an overall time complexity of $O(|V|^2 + |E|)$ per iteration. They have also used a simulated annealing optimization technique from Davidson & Harel [7]. Hu [4] proposed several modifications to Fruchterman & Reingold's algorithm which lead to layouts of better quality and improve the computational efficiency significantly by reducing repulsion calculations to $O(|V|\log|V|)$ per iteration through the use of a QuadTree/OcTree spatial decomposition data structure. Kamada & Kawai [6] on the other hand require that the graph theoretical distance between all pairs of vertices is computed and forces along with the energy model are based on this distance. The overall computational complexity of Kamada & Kawai's algorithm is $O(|V|.|E|)$ per iteration and an $O(|V|^2)$ memory complexity.

A big part of drawing a graph for use in software visualization is overlap removal. In order for information to be visually comprehensive the graph drawing should not be cluttered and overlaps between vertex bounding regions should be avoided. There are two ways to remove overlaps in a force-directed algorithm. The first one is to modify the force model to rapidly repulse vertices whose bounding regions are overlapping and thus produce a layout without overlaps. Such a method has been suggested by Li, Eades and Nikolov [8]. Gansner and Hu [9] suggest a second way, in which a post-processing step is used after the layout algorithm has finished, computing and altering the final drawing so that overlaps are avoided.

In addition to vertex overlap removal, edge bundling algorithms are used to reduce clutter and thus achieve a more visually comprehensive drawing. Edge bundling algorithms group edges with similar paths into bundles improving overall

[1]Ivan Iliev, Haralambi Haralambiev and Stanimir Boychev are with the Applied Research and Development Center at Musala Soft, 36 Dragan Tsankov blvd, 1057 Sofia, Bulgaria
E-mails:
{ivan.iliev,haralambi.haralambiev,stanimir.boychev}@musala.com

[2]Milena Lazarova is with theComputer Systems Department at Technical University of Sofia, 8 Kliment Ohridski blvd, 1756 Sofia, Bulgaria, E-mail: milaz@tu-sofia.bg

visibility in the drawing by making multiple edges look like a single one – a bundle. An algorithm for edge bundling is suggested by Danny Holten and Jarke J. van Wijk [10].

This paper focuses on drawing undirected straight-line edge graphs using iterative force-directed methods combined with non-iterative ones,based on the research in [4] and [5]. The idea of pinning weights is adopted to support dynamicity. The algorithm introduces semantic clustering and an optimization of per iteration computational cost of the layout, making it more suitable for use in software visualization. The post-processing overlap removal based on a proximity stress model ([8]) is used to ensure the preservation of the mental map after each removal step. The suggested algorithm can also be applied for three-dimensional visualization.

III. Aesthetic Goals

Several drawing constraints must be chosen for the algorithm in order to achieve good layout quality for software visualization and comprehension. After reviewing and experimenting with different layouts the following aesthetic criteria [11] are settled on:

- **Semantic clustering of vertices** – due to the hierarchic nature of a software system the graph is recursively divided into clusters such that all vertices with the same parent belong in the same cluster. This allows for easy identification of software components based on their position in the graph.
- **Dynamicity** – the overall structure of the graph does not change with each consecutive drawing. Different revisions of a software system can be drawn consecutively while preserving the user's mental map.
- **Overlap removal** – all overlaps between vertex bounding regions should be removed by scaling the entire drawing so that the initial graph structure is fully preserved.
- **Edge bundling –** edges sharing similar paths should be grouped together into bundles.

IV. Layout Algorithm

A. Initial Positioning

In order to begin the layout process initial positions must be set for all vertices. This is currently done by assigning random coordinates within a rectangle whose width and height are equal to the sum of all vertices' widths and heights. In 3D, an analogous operation is performed with a rectangular cuboid. If a previous layout has been executed, all vertices that have already been positioned by it keep their coordinates and only the non-positioned vertices' coordinates are randomized.

B.Clustering

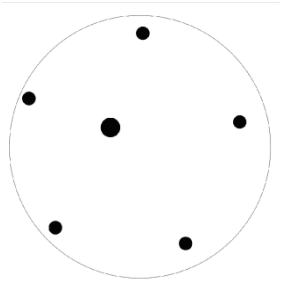


Fig.1.A circular cluster with its smallest enclosing circle

Before the force-directed iterative process can begin all vertices are ordered in circles/spheres around their respective parents, forming clusters.The smallest enclosing circle [12]/sphere [13] is calculated for each cluster (Fig.1). Each vertex, regardless of shape, is assigned an enclosing circle/sphere. The algorithm used to position the vertices in a circle is based on a simple subdivision of the circle into equally sized sectors. All child vertices are sorted by their radius in ascending order and divided into groups by certain criteria. For example, for object-oriented languages the access modifier type (private, protected, public) is a good choice. The algorithm is applied to each group. At first as many vertices as there is room for are positioned on a circle with the smallest possible radius so that there are no overlaps. After the initial vertices are positioned the radius is increased and part of the remaining vertices is positioned again. The process continues until all vertices from this group are properly ordered and then moves on to the next group. As a result, an even distribution of vertices on concentric circles around their parent as well as a clear separation of children by access type is achieved (Fig. 2). In three dimensions the process is analogous with circles replaced by spheres. A modification of Saff and Kuijlaars [14] algorithm is used for distributing points on a sphere using a golden section spiral. The ordering step is applied once at the beginning and once at the end of the layout process to initially calculate bounding circles for all vertices/clusters and to move child vertices to their final positions respectively.

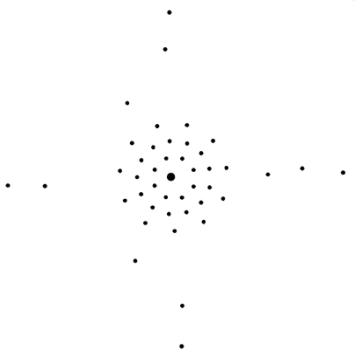


Fig.2. Methods in concentric circles around their parent class. Each method is positioned in a circle with radius based on its access modifier type (private, protected, public)

C. Dynamicity

In order to support dynamicity in the suggested layout pinning weights are assigned to all clusters in the following way:

- If the cluster's parent vertex is new (not existing in a previous layout) or has been removed (not existing in the current layout), the cluster's pinning weight is set to 1 allowing its free movement.
- If the cluster's parent vertex has received coordinates from the previous layout and has persisted, its pinning weight is set to 0 prohibiting further movement.

The assignment of pinning weights in that manner leads to the preservation of the mental map between consecutive drawings.

D. Force-directed iterative process

Since child vertices will be recursivelylocated around parent ones, only top-level vertices need to be positioned initially (i.e. ones without parents themselves). Therefore, the force-directed iterative layout is applied only for the top-level clusters reducing the overall complexity of the algorithm based on their number. In software systems with many top-level clusters the speed up will not be substantial but still significant enough. Before beginning the iterative process all edges between top-level clusters are counted and the result is stored for each pair. The Fruchterman & Reingold force model is used as modified by Hu in [4] combined with the adaptive step length optimization for simulated annealing. The following actions are performed in each iteration:

1. For each top-level cluster the QuadTree/OcTree is recursively opened.
2. Repulsive forces are calculated between the current top-level cluster and clusters contained within tree nodes "close enough" to it. If the tree node is "far" from the current cluster, repulsion is applied based on the distance between the two. The criteria for "close enough" and "far" are the same as defined in [4] except for the case when there is an overlap in the bounding areas of the current cluster and tree node. If such an overlap exists the tree node is considered "close enough" regardless of the other criteria. Repulsive displacements are stored for each cluster's parent vertex based on the forces calculated.
3. Attractive forces are calculated between each pair of clusters with a positive calculated edge count. The power of the attractive force is based on the number of edges between the two clusters. Attractive displacements are stored for each cluster's parent vertex based on the calculated force.
4. Movements are performed for each vertex based on its stored displacements. Each movement is limited by the pinning weight of a vertex and by the current temperature of the layout used in the simulated annealing optimization. The coordinates for the centres of each bounding circle of top-level clusters are updated with the vector between the cluster's parent vertex previous and new positions.
5. An adaptive cooling step is executed which reduces or increases the current layout temperature.

E. Post-processing

The iterative process ends when a convergence criterion is satisfied or a maximum number of iterations are exceeded. The layout process finishes after an overlap removal post-processing step and an edge bundling step are performed. Snapshot of the edge bundling effect is shown in Fig. 3. A full layout of a software system is depicted in Fig. 4.Pseudo code for the entire layout process in 2D is given in Code. 1 and Code. 2.

```
layoutGraph(Graph):
totalWidth = sum(Graph.Node.width)
totalHeight = sum(Graph.Node.height)
foreachNodeinGraph do
    Node.X = random()*totalWidth
    Node.Y = random()*totalHeight
Clusters = buildClusters()
foreachClusterinClusters do
    calculateBoundingCircle(Cluster)
    positionChildren(Cluster)
setPinningWeights()
while not converged and currentIteration<
MAX_ITERATIONSdo
    calculateDisplacements()
    performMovements()
    cool()
    currentIteration++
foreachClusterinClusters do
    positionChildren(Cluster)
removeOverlaps()
bundleEdges()
```

```
cool(T):
reduce or increase the current layout temperature.
```

Code.1. Pseudo code for the layout process and cooling function

V. RESULTS

With the circular/spherical arrangement of vertices into clusters and having to only run the force-directed algorithm on top-level ones, the time complexity of each iteration is O(|T|log|T| + |TE|) where T is the set of top-level vertices and TE is the set of edges between them. The arrangement step takes O(|V|) time to complete. The most computationally expensive part of the layout is the force-directed step. Usually in most graphs of software systems the ratio |T|/|V| is quite small due to the low number of top-level vertices which results in a lower complexity per iteration and reduces the computational time of the layout significantly.

```
calculateDisplacements(Graph, QuadTree):
foreachNode in Graph do
      Q = QuadTree.ROOT
while not empty(Q) do
   TreeNode = dequeue(Q)
   if far(Node,TreeNode) then
      calculate repulsive displacements between Node
      and TreeNode's centre of gravity
   else if leaf(TreeNode) then
      foreachContainedNodein
      containedNodes(TreeNode)
       calculate repulsive displacements  between
      Nodeand ContainedNode
   else
      enqueue(TreeNode.Children)
foreachEdge between Clusters do
   calculate attractive displacements between Edge.From
   and Edge.To based on Edge.Count
-----------------------------------------------------
performMovements():
foreachClusterinClusters do
   Displ = getDisplacement(Cluster.Parent)
   WPin  = getPinningWeight(Cluster.Parent)
   LT = getLayoutTemperature()
   move(Cluster.Parent, min(Displ,  WPin*LT))
   move(Cluster.Centre, min(Displ,  WPin*LT))
```

Code.2.Pseudo code for calculating displacements and moving vertices based on them

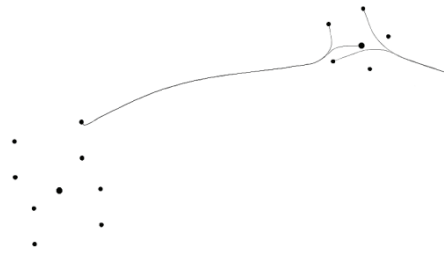


Fig.3. Edge bundling effects on edges – edges sharing a common path are grouped.

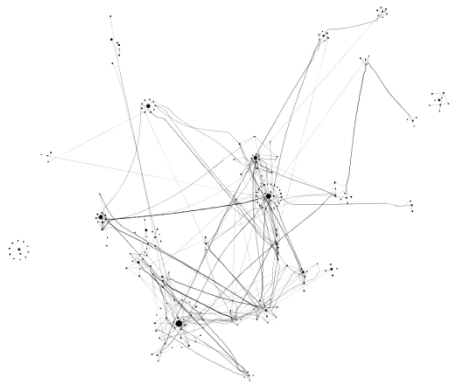


Fig.4. An overview of the Pygmy project graph (http://pygmy-httpd.sourceforge.net) laid out using this algorithm showing the effects of edge bundling and node clustering.

VI. Conclusions and Future Work

Drawing graphs of software systems to ease program comprehension is an open problem without a definitive solution. An algorithm suited especially for the purpose of drawing such graphs in an efficient, useful and aesthetically pleasing manner is suggested and described in the paper.

Further investigation and improvement of the algorithm will be based on using an alternative metaheuristic instead of simulated annealing for solving the force-directed layout optimization problem such as genetic algorithm or ant colony optimization. Moreover, the presented algorithm could be measured against other layout algorithms on certain characteristics - performance, aesthetics, etc.

Acknowledgement

This work was partially supported by the Bulgarian National Science Research Fund through contract ДМУ 02/18 - 2009.

References

[1] S. Diehl, "Software Visualization - Visualizing the Structure, Behaviour, and Evolution of Software", Springer, 2007.

[2] Y. Frishman, A. Tal, "Online Dynamic Graph Drawing", Proc. Eurographics/IEEE VGTC Symp. Visualization (EuroVis '07), pp. 75-82, 2007.

[3] P. Eades, "A Heuristic for Graph Drawing", Congressus Numerantium, vol. 42, pp. 149-160, 1984.

[4] T. Fruchterman, E. M. Reingold, "Graph Drawing by Force-Directed Placement", Software Practice and Experience, vol. 21, pp. 1129-1164, 1991.

[5] Y. F. Hu, "Efficient and High Quality Force-Directed Graph Drawing", The Mathematica Journal, vol. 10, pp. 37-71, 2005

[6] T. Kamada, S. Kawai, "An Algorithm for Drawing General Undirected Graphs", Information Processing Letters, vol. 31, pp. 7-15, 1989.

[7] R. Davidson, D. Harel, "Drawing Graphs Nicely Using Simulated Annealing", ACM Trans. Graph., vol. 15, iss. 4, pp. 301-331, 1996.

[8] W. Li., P. Eades, N, Nikolov, "Using Spring Algorithms to Remove Node Overlapping", Proc. of the 2005 Asia-Pacific Symposium on Information Visualisation, Vol.45 (APVis '05), pp. 131-140, 2005.

[9] E. Gansner, Y. Hu, "Efficient Node Overlap Removal Using a Proximity Stress Model", In Graph Drawing, Ioannis G. Tollis and Maurizio Patrignani (Eds.). Lecture Notes in Computer Science, Vol. 5417. Springer-Verlag, Berlin, Heidelberg, pp. 206-217, 2009.

[10] D. Holten, J, van Wijk, "Force-Directed Edge Bundling for Graph Visualization", Proc. of 11th Eurographics/IEEE-VGTC Symposium on Visualization (Computer Graphics Forum; Proceedings of EuroVis 2009), pp. 983 - 990, 2009

[11] G. Battista, P. Eades, R. Tamassia, I.G. Tollis, "Graph Drawing - Algorithms for the Visualization of Graphs", Prentice, 1999.

[12] F. Nielsen, R. Nock, "Approximating Smallest Enclosing Disks", Proc. of 16th Canadian Conference on Computational Geometry (CCCG), pp. 124–127, 2004.

[13] F. Nielsen, R. Nock, "Approximating Smallest Enclosing Balls", Proc. of International Conference on Computational Science and Its Applications (ICCSA)", Springer, Lecture Notes in Computer Science, vol. 3045, pp. 147-157, 2004.

[14] E. Saff, A. Kuijlaars, "Distributing Many Points on a Sphere", The Mathematical Intelligencer, Vol. 19, No. 1, pp. 5-11, 1997.

The Application of Minimax Decision Rule in Games

Milena Karova[1], Lyubomir Genchev[2], Lyubomir Vasilev[3], Ivaylo Penev[4]

Abstract – **This paper demonstrates three different applications: Minimax algorithm, Alpha Beta pruning algorithm and Genetic Algorithm in games. They are used to evolve a Tic Tac Toe and Chess games. The size of strategies space is defined by the number of all possible game situations, which follows from the question of how many distinct matches can be played. The using of GA implementation improves the optimal paths and decreases the playing time.**

Keywords – **Minimax strategy, Alpha Beta Pruning, Genetic Algorithm, Fitness Function, Game Tree Decision.**

I. INTRODUCTION

The Minimax decision rule is applied as a solution to two-player zero-sum games and in those cases is equal to the Nash equilibrium. Since in these types of games both players work towards the same mutual goal and one player's moves towards winning directly affect the chances of winning for the other player in a negative manner.The Minimax theory is based on maximizing the potential gain for one player while minimizing it for his opponent (and vice versa).An algorithm exists in computer science which implements the Minimax decision rule and is normally used for simple two-player zero-sum games (e.g. Tic-tac-toe). It can also be applied to more complex games such as Chess and Go but without additional optimization it is highly inappropriate.

II. MINIMAX IN GENERAL

`The Minimax algorithm [2] works by scanning the nodes (and all of its children) of a game tree from a given configuration and evaluates them based on the Minimax theory. The algorithm is in fact a form of depth-first search and on a programme level is normally implemented as a recursive algorithm. Two basic strategies exist for the Minimax algorithm: the first one consists of searching every children of every node of the whole tree and the second limits the depth of the search in order to save computational time. Both strategies have their pros and cons. The first provides a highly unlikely chance for a mistake but demands more time and is virtually impossible to implement in complex games. The second can greatly limit the computational time but does by creating the so-called "horizon effect", i.e. limiting the number of children nodes the computer can search ahead, this can lead to the choice of a move that might later prove to be bad (but the computer could not have predicted it). However, there are some methods that optimize the Minimax algorithm, the most popular of which is the alpha-beta pruning, which minimizes the time necessary for the machine to complete the task while at the same time allowing more depth for the search.

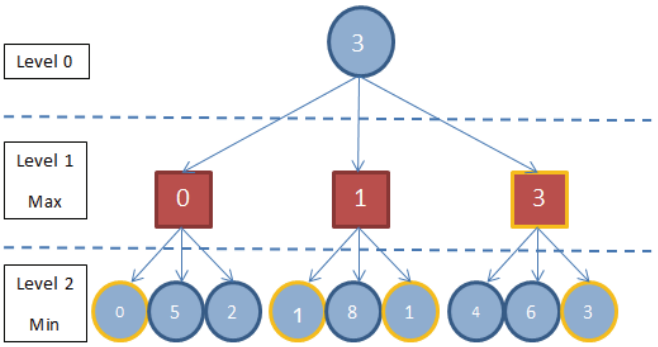


Fig. 1. The Basic Principal of Minimax Algorithm

Fig. 1 demonstrates the basic principal of the Minimax algorithm. Following the Minimax theory, the program will attempt to maximize one's player score while minimizing the others. The best move is the one that brings the most benefit to the maximizing player and the least to his opponent. Since two-player zero-sum games have a shifting nature, meaning that as the first player tries to maximize his own score in the first move, the second will try to minimize the first player's score in his own turn, the algorithm changes its action with every move. As in the figure, in level 1 the program is maximizing, while in level 2 – minimizing, hence the name "Minimax". In games, the algorithm works by analysing a given board configuration (in this case, that would be the node of level 0). The computer needs to choose the best move so it analyses all possible moves (the nodes of the game tree) to distinguish the one that brings the most benefit. Because the algorithm is maximizing in level 1, the most beneficial move would be the highest rated one. In order to rate the nodes (and, consequently, the moves themselves), the computer must search through all the children of the nodes. In a situation of a board game, for example Tic-tac-toe, this means the machine needs to recreate all the possible outcomes of a given configuration, i.e. play until a finished game, so that the said configuration can be evaluated.

[1]Milena Karova is with the Department of Computer Science and Technologie, Technical UniversityVarna, Bulgaria, E-mail: mkarova@ieee.bg.
[2]Lyubomir Gencheva student with the High School of Mathematics,Varna,Bulgaria,E-mail:lubo1993@gmail.com
[3]Lyubomir Vasilev a student with the Fourth Language School, Varna, Bulgaria, E-mail: lubodjwow@gmail.com
[4]Ivaylo Penev is with the Department of Computer Science and Technologie, Technical University Varna, Bulgaria, E-mail: ivailopenev@yahoo.com.

III. Implementation in Games (Tic Tac Toe and Chess)

A. Minimax Algorithm

Tic-tac-toe is a classic example of game which can be played by a computer using the Minimax algorithm. In this case, Minimax is an ideal solution because the branching factor of the game is only 9, as opposed to more complex games such as Chess, which has a branching factor of 35. The algorithm will work in absolutely the same way as in its general form [3].

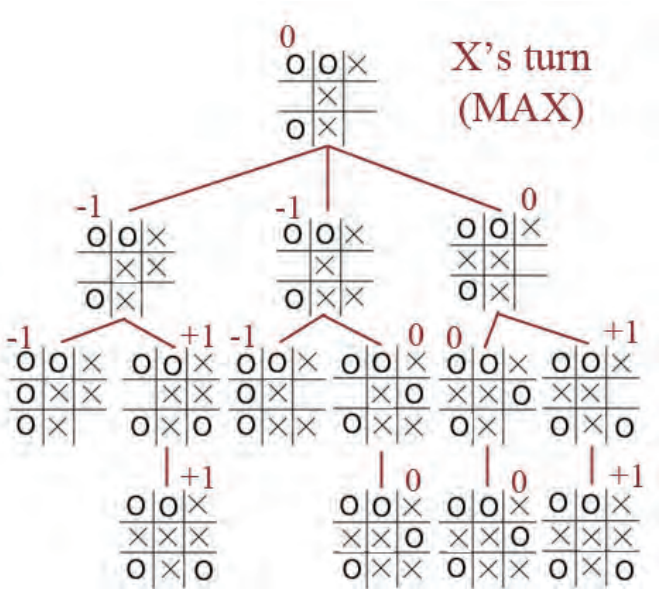


Fig. 2. Sample Tic Tac Toe Game Tree

Fig. 2 represents a sample Tic-tac-toe game tree. It is very similar to the tree in Fig. 1. This demonstrates that the Minimax theory and algorithm remain largely unchanged in their different uses.

```
Minimax_algorithm(player,board)
   if(game over in current board position)
      return player(the winner)
   child nodes = all legal moves for player from this board
   if(max's turn)
      return maximal score of calling Minimax on all the
children
   else (min's turn)
      return minimal score of calling Minimax on all the
children
```

Fig. 3. Pseudo code of the Minimax algorithm

Figure 3 shows the pseudo code implementation of the Minimax algorithm in Tic-tac-toe.

In the case of a game, what the algorithm does is recreate all the possible plies stemming from a configuration. The programme takes into consideration the opponent's best moves and the first player's best counter-attacks. This logic explains the Minimax algorithm and leads to the best moves for each player's turns. The most negative feature of this algorithm is that it requires great computational time in more complex games if it runs in full depth. Should the depth be limited, this will result in possible mistakes. This is the reason why an optimization is needed for a more efficient operation of the Minimax algorithm.

B. Alpha Beta pruning algorithm

The desired improvement turned out to be the Alpha-beta pruning algorithm. Alpha-beta pruning is a search algorithm that relies on the Minimax theorem but brings new light to the implementation of the theorem by minimizing the game tree [2]. The algorithm returns the same result as pure Minimax but in the best case it does it twice as fast. It literally prunes or cuts off some nodes that cannot lead to a better overall result (they are suboptimal). While simple Minimax will explore all possible nodes alpha-beta explores only those which seem to be better than the best move till now. This is a huge advantage when it comes to exploring game with big branching factors such as Chess. Chess has a big branching factor ≈ 35 compared to 9 in simple games like Tic-tac-toe. Using brute-force-like algorithms such as simple Minimax lead to search explosion in such conditions. This makes the application of simple Minimax rule in games like Chess not impossible but unprofitable as it requires immense computational power.

Table I
Worst and Best Scenario

Depth	*Worst scenario*	*Best scenario*
n	$\mathbf{b^n}$	$\mathbf{b^{n/2}+b^{n/2}-1}$
1	20	20
2	400	39
3	8 000	178
4	160 000	399
5	3 200 000	3576
6	64 000 000	15 999
7	1 280 000 000	71 553
8	25 600 000 000	319 999
9	512 000 000 000	715 540
10	10 240 000 000 000	6 399 999

TableI shows how much we can benefit from alpha-beta pruning in cases of big branching factor and good move-ordering. It shows the number of child nodes with depth **n** and branching factor **b**=20. Table1 is divided in two categories: worst scenario and best scenario. Here it should be mentioned

that the alpha-beta algorithm highly depends on the move-ordering. Hence, for the minimizing player, sorting succesor's utility in an increasingorder is better. For the maximizing player, sorting successor's utility in a decreasing order is better. The maximal number of leaves is $\mathbf{b^n}$. In this worst case the program has to explore all the nodes in order to find the best one. After the best move has been found there are now nodes left to be pruned. This means that no pruning will be made, which proves to be the same as pure Minimax searching. However, if the move-ordering is good (best case) the number of leaves plummets as TableI shows. Slagle and Dixon first showed that the number of leaves visited by the alpha-beta search in this "best case" must be at least: $\mathbf{b^{n/2}+b^{n/2}-1}$. Since the best move has been found first there is no point in exploring all the remaining nodes. D. McIllro then proved that alpha-beta search for a random-generated game will be 33% faster than pure Minimax. That is why move-ordering is the focus of a lot of effort when writing an efficient program.

The algorithm calculates and keeps track of two variables: alpha and beta, one for each player. Alpha represents the value of the best possible move the current player has made so far. Beta, on the contrary, represents the value of the best possible move the opponent has made so far.

Figure 5 gives a clearer picture of how the algorithm actually operates:

№	Line
1.	AlphaBeta(player, α, β,depth)
2.	If(depth==0) return heuristic_evaluation();
3.	If(Maximizing player)
4.	{
5.	For each following child node
6.	Node_score= AlphaBeta(minplayer, α, β,child);
7.	If (node_score> α) α =node_score ; // *A better move has*
8.	*been found*
9.	If(α>= β)
10.	return alpha;//*Cut off*
11.	}
12.	Else If(Minimizing player)
13.	{
14.	For each following child node
15.	Node_score= AlphaBeta(max player, α, β,child);
16.	If(node_score<beta) β =node_score; // *A better move*
17.	*has been found*
18.	If(α>= β)
19.	return beta;// *Cut off*
20.	}

Fig. 5. Pseudo code Alpha-Beta pruning algorithm

C. A Genetic Minimax Algorithm

A Genetic Algorithm (GA) is most effective in situations, for which a well defined problem offers a compact encoding of all necessary solution parameters [1]. If this encoding grows two large and complex, the algorithm faces similar limitations of other local search methods and cannot be expected to find a global optimum. In considering the Tic Tac Toe strategy problem it is at first important to find a suitable representation and to ensure that a GA can be effectively applied.

The encoding of chromosome depends on game problem. The Fig. 4 presents Tic Tac Toe game tree encoding. Each gene is defined by the correspond move to be taken. The chromosome is a table with 827 genes to represent each game situation.

Fitness function is important to create an efficient GA and it is formed as way:

f(n)=possible win configurations for current player – possible win configurations for opponent player.

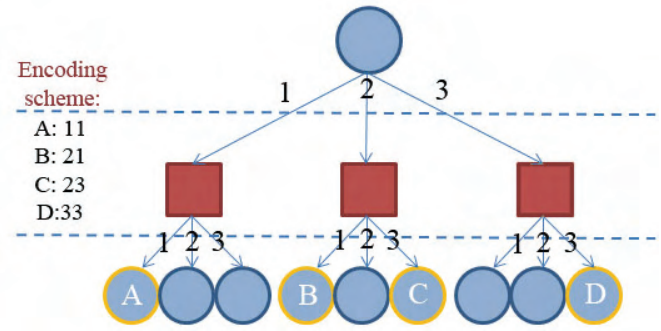


Fig. 4. Game tree encoding scheme

Based on its performance, each individual is assigned a fitness measure. The higher its measure, the more likely it is that an chromosome will take part in crossover and will be passed to the next generation. The parents can be choosed applying the genetic operator selection.

Once the parents are determined, the offspring is created by one-point crossover [Fig. 6]. The genes are copied from first parent at point of node 3 and the genes continue to be copied from the other parent. The crossover probability p_c is around 0,90.

The mutation [Fig. 7.] can occur at each gene of chromosome with probability p_m by a new random validate value is chosen to replace the current one. The p_m is very low value (less than 0,009) since it is evaluated for each gene independently. The essential goal of genetic operators is to ensure general variety in the reproductive process over time.

The size of population is 50 and the number of generations is 50.

The code for the implementation of Tic Tac Toe GA is a C++-based programming framework for GAs. Some modifications are made to accommodate the chromosome and the fitness function for the specific encoding and game tree solution.

IV. Experiments and Results

Each of those algorithms has experimented and Fig.8and Fig. 9give the results.

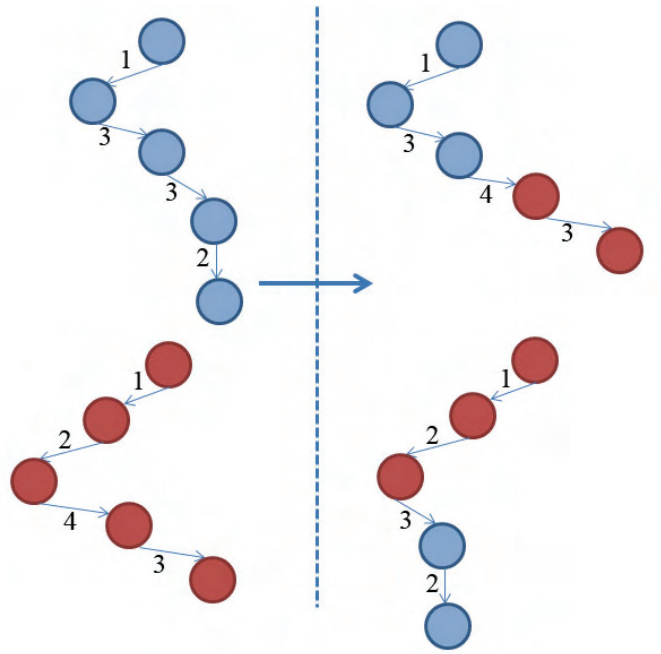


Fig. 6. Genetic operator One-point crossover

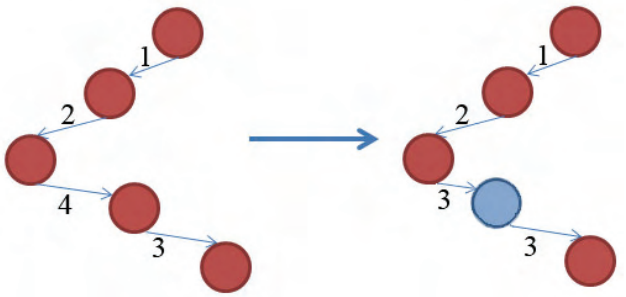


Fig. 7. Genetic operator mutation

To illustrate the efficiency of the different methods, the following charts are provided. The first chart [Fig. 8.] exhibits the relative time needed for the programme to calculate the best move for the Tic-tac-toe configuration, depicted in Fig. 9, using each of the three methods. The results show that there is little difference between the computational time for the three algorithms for this task. However, there is a sharp distinction in the second chart, which demonstrates the necessary time for calculating the best 11th move of an ongoing chess game. Our results show that the genetic algorithm is the best solution for two-player zero-sum games with a high branching factor.

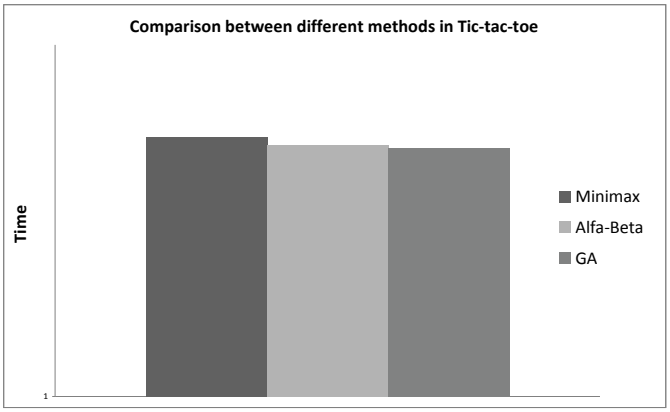


Fig. 8. Comparison between different methods in Tic Tac Toe

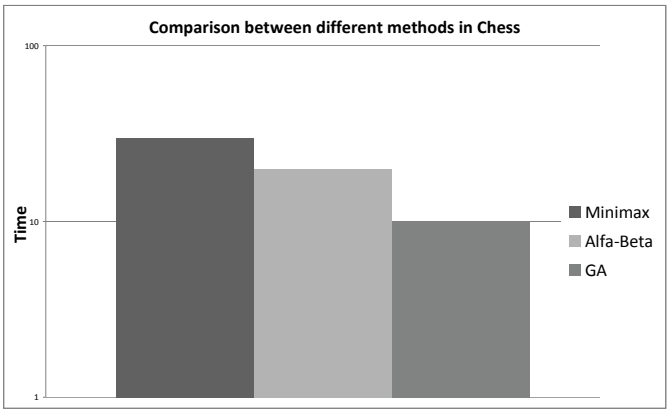


Fig. 9. Comparison between different methods in Chess

V. Conclusion

We design and implement a variety of techniques for solving Tic Tac Toe game and Chess. The GA could be applied successfully to evolve these game strategies.

The optimal result of genetic algorithm for different games is not guaranteed because it depends on the length of the chromosome and the depth of the tree decision.

Further testing would serve for running GA with different population parameters. In order to improve the fitness function, it can change the weights, given to the possible moves for each player. The number of paths or considering the next shortest path from a given position may make the fitness function more optimal heuristic function. It possible also to create hybrid algorithm Alpha Beta pruning with GA to find a good moves in two players games in a faster way.

References

[1] GA of Tic Tac Toe game: http://www.vclcomponents.com/s/0__/code_genetic_algorithm_for_tic_tac_toe/v.

[2] Russell S., Norvig P., “Artificial Intelligence A Modern Approach”, Third Edition, Prentice Hall, 2003, 2010.

[3] Schaefer S., “How Many games of Tic Tac Toe are there? http://www.mthrec.org/old/2002jan/solutions.html, 2002.

PostIB as a Logistic Support for the Development of Rural Areas in the Republic of Serbia

Zoran Marković[1], Ivan Tričković[2], Obrad Peković[2], Bojan Jovanović[2]

Abstract-**This paper is about the possibilities of implementing an integral information system in the Serbian Post, which would be an e-Marketplace for a wide range of goods and services linking a buyer and a seller, both at home and abroad. Thus conceived e-exchange would have special significance for the development of rural areas where the availability of information resources is insufficient, both because of lower educational structure, and due to poor availability of telecommunications. In this sense, the Serbian Post may provide their information resources and those of the postal network in the entire territory of Serbia**.

Keywords – **Post, Information System, e-stock exchange, rural areas, commodities.**

I. INTRODUCTION

For several decades we live and work in conditions of rapid development of information technologies that pervade all human activity. Information, and decisions made based on them, are a key factor for success in modern business conditions. The combination of timely and accurate information and good logistic support provides opportunities for synergetic effects in the economy of the country, especially when it comes to individual producers, craftsmen and small businesses in rural areas or in areas outside the major traffic corridors and urban areas.

Smaller commodity producers such as craftsmen, people involved in agriculture, tourism or other fields of production, have a problem selling their products or services, which is to say that they have a problem with the distribution of goods, because the amounts are relatively small and there are high transport costs per unit of product. The efficiency of business, among other things, is measured by effective division of labor, effective communication and efficient logistic support. So if a commodity producer or a provider of service has to think only about the quality and quantity of his product, then the product will be more competitive in the market. On the other hand, the subjects that take care of logistics and organization of information flows can benefit from this because their business is based on quantity whereby the economy of the volume of production leads to lower costs per unit of goods or services.

Post of Serbia has all the prerequisites to become an important link in the chain of goods and services, or buyers and sellers. It owns over 1500 postal network units where computers and transportation system are linked into one unique system. Such a network provides the possibilities of improving the efficiency of supply chains. Experience in creating software solutions, the tradition in the transfer of postal items, remarkable business capacity, usage of network nodes and other specific qualities of Serbian Post open the possibility of the new approach to attractive market segment dominated by the population, small and medium enterprises.

PostIB is one way to streamline the supply chains. All of the aforementioned issues, as far as resources of Serbian Post are concerned, indicate that the project PostIB is feasible and can contribute significantly to the economic conjuncture.

II. STRUCTURE OF POSTIB

PostIB (Postal Information Exchange) is the idea that the resources of Post Serbia are put into operation to increase overall economic activity of the country and increase its own business volume. Namely, Post Serbia has elements of the postal network in all parts of Serbia and that represents a comparative advantage. Each node of the network is an information and transportation system associated with the rest of the network which is also very important from the standpoint of the efficiency of logistics and information flows.

PostIB should be a Web-oriented application, open to outside access, which would serve as the platform for exchange of goods and services. The database would contain the records of goods and services offered on the market, their description, origin, price, quantity, and all the other elements that might be decisive for a buyer's choice to purchase a certain product.

PostIB would manage databases that would be accessible to all interested parties. In addition to the database of supply and demand, it could also manage a database of risky customers and risky suppliers, their credit worthiness, property map and other information that would affect trust and safe functioning of e-stock exchange. Furthermore, statistics on traffic could be kept there as well as the realized prices of certain categories of goods and services, inventories and expected sales period, the time of contracting and a string of other information of interest to the functioning of supply chains. In addition, if the transport logistics of Serbian Post are not used, it is necessary for PostIB to have data on transport enterprises, their capacity and solvency, pricing, storage facilities, their locations, leisure facilities and other resources. PostIB would be a structural part of an integral information system of Post Serbia and in that way allow, in addition to transport and T& T services for asset tracking, secure payment, digital signing of documents, freight forwarding, customs clearance delivery to the designated address, storage, warehousing and many other features that Post has to offer as an integral service.

It has already been mentioned that the Serbian Post has over 1500 posts of which almost 1400 points have well-connected IT resources, which can be made available to

[1]Zoran Marković is with Public Enterprise of PTT Communications "Srbija", Takovska 2, 11000 Belgrade

[2] Ivan Tričković and Obrad Peković and Bojan Jovanović are with Faculty of Technical Sciences, Trg Dositeja Obradovica 6, Novi Sad; Serbia; E-mail: ivantricko@gmail.com; obradpek@eunet.rs; bojanjov@uns.ac.rs

interested parties. In Figure 1 we see the backbone of Postnet network, which to a large extent, coincides with the transport network. The plan of Post Serbia is that in the recent future all posts would be computer linked to an integral information system (PostTIS) and to increase the capacity of the network, building its own fiber optic network and series of other investments that will raise the level of IT capacities of Post Serbia. The access to PostIB application would be enabled through the Internet with the previous registration of the user categories. The categories of users are important in many respects and the most important is the structuring of e-stock exchange, the speed of the search base, market segmentation and the like. For rural areas, where residents do not have Internet access or do not use it, post offices would be the place for making evidence of and registering users, database search, arranging transactions, the reception of goods for the transfer, payment, etc.. The role of Post Office workers would not only be to make the information resources available, but also to carry out training and provide all other assistance to interested parties. In this way, the availability of PostIB would be in all parts of Serbia equal for all areas of business, under equal conditions, without discrimination for all the structures of society.

Certainly such a project has to be supported by the Serbian government and the mechanisms that control the flow of goods and services, taxes and fees, so that the trade of goods and services through PostIB would be put into legal channels.

Besides working on a software solution, according to the known principles of Web applications, relatively small investments in IT resources are required. Certainly a small investment compared to potential benefits both for the Post of Serbia, and for the country but also for all participants in supply chains, that is buyers and sellers. The application itself would be distinguished from conventional applications of electronic commerce and have the elements that can be seen in social networks. The contents would vary and that includes good monitoring and the moderators who would care about the acceptability of the content.

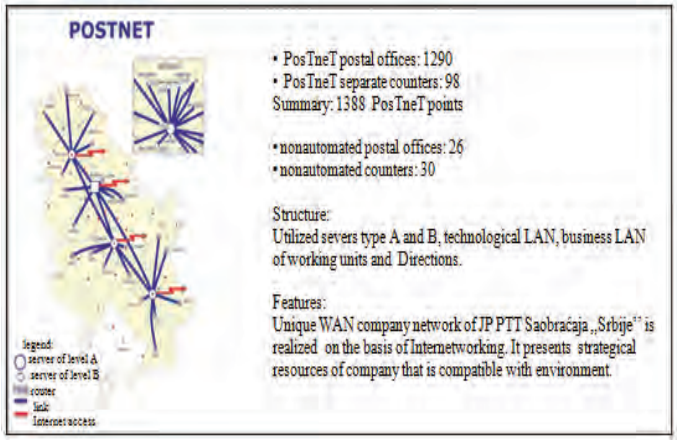


Fig. 1. Backbone of PostNet network

*This work is result from the research in the framework of the project TR 36040

III. The Difference between E-stock Exchange and Electronic Commerce

When e-exchange is mentioned people usually think of e-commerce, which is not wrong, but there are major differences between these two notions. Namely, electronic commerce has all the classic elements of trade, which differ in form as follows:

1. Product
2. Origin
3. Marketing
4. Way of receiving orders
5. Way of receiving money
6. Delivery
7. The possibility of returning the product
8. Guarantee
9. Technical Support

Therefore, electronic commerce must involve a legal person who owns a storehouse, has range of merchandise, developing sales marketing, means of communication with the customer, the modalities of payment and delivery and other aspects of the contract by approach. There is an offer of goods and the buyer must agree to the terms of purchase, price, dates and so on. There is a one-way communication where a customer receives information about the goods or services and makes decision about buying. The organization of electronic commerce involves organization of the warehouse, entry and exit of goods, inventory records, which creates costs that eventually a buyer will have to pay. It is virtually the same process as the wholesale and it essentially removes the retail from the supply chain and communication with the customer is done electronically. The total cost of all this reduces the product price because it avoids the retail margin.

Because of its characteristics electronic commerce in recent years experienced an expansion regardless of the shortcomings that are present. The main disadvantages are security payments and still not sufficient use of Internet to purchase goods and services.

On the other hand, e-stock exchange, in the way it can be realized through PostIB, has completely different characteristics. There is almost nothing of the aforementioned characteristics of electronic commerce except the delivery of goods only in the case when a buyer and a seller chose the channel of distribution through the Post of Serbia. PostIB is means of two-way communication between a buyer and a seller . They are free to communicate and negotiate the terms of transaction, mode of transport of goods and payment. Logistical resources of Serbian Post are available to everyone but there are transactions where the Post is unable to carry out the logistic part of the work. The post as the carrier of PostIB has no storage of goods-- goods stay with producers. It also does not have a center for ordering and marketing. This work is done by vendors on their own and can be said that such concept cuts out wholesale and retail trade from the supply chain and, thu,s the prices of goods and services continue to decrease. Producers take care about quality and quantity, method of packaging, marketing and shipping to the customer. The buyer selects the goods available, negotiates the terms of

the transaction (which is not the case with e-shopping) and pays a price that is lower.

PostIB should also enable the customer base where the market needs are recorded and where sellers and manufacturers can respond with their offers. Two-way communication would not end there but rather just initiate a direct link between a buyer and a seller as it happens in social networks. If we take Facebook as the parallel then we could have the registration of buyers and sellers, they could form a list of "friends" (the base of their customers or suppliers), organize blogs and forums for sharing knowledge and so on. Each would have its own page where they advertise a product or service, publish their ads, comment on the appearance, etc. When there is a new registered user who is interested in the same field (for example, rural tourism and wine production, purchase of philatelic stamps, service of chain saws and the like) previously registered users would get a notification about it in an email or on their mobile phones. In fact, all areas of human activity can be a subject of trade through PostIB even used computers, cell phones, automobiles, agricultural products and so on. It certainly does not exclude the possibility that the Serbian Post would offer its own goods and services, or that it would use some of its resources to rationalize the business between buyers or sellers when the Serbian Post is included in the logistic chain.

According to recent research Online Retail is setting new records, as in the fourth quarter of last year the volume of trade over the Internet reached 43 billion dollars. Internet traders can be satisfied, because the trading volume increased by about 11 percent compared to the same period of 2009. In addition to record sales in the three-month period, the holiday season was particularly good, as many holiday days set a new record when it comes to the volume of electronic commerce. In the first quarter of 2010, the volume of e-commerce was increased by about 10%, while in the second and third quarter increase was slightly smaller and amounted to about 9%. It is interesting that in the past decade, the volume of increase of e-commerce remains fairly stable at 20 percent annually. The exception was 2008 and years of recession. The strength of this sector at the global level is obvious, and analysts predict that they will return to the rank of a standard two-digit growth on an annual basis, especially now that many think that the recession and the crisis are behind us. In any case, in 2011. one foresees two-digit growth. The most represented sectors in the field of electronic commerce are computer software, consumer electronics, computers and peripherals, toys. All these categories have recorded an annual increase of about 15 percent. The most significant names of e-commerce continue to hold up in the same proportion as the 25 largest Internet stores achieved 68.4 percent of all sales. It is interesting that this figure fell by 5.6 percent from the fourth quarter of 2009. This could be an indicator that small and medium-sized Internet retailers also fail to recover from the economic crisis.

When we have in mind and when we consider the additional benefits provided by PostIB concept, it is clear that the future supply chain is reduced to a direct relationship between supply and demand (the classical trade is increasingly avoided), and also the logistical support in the area of transport and payment.

IV. Logistical Role of Serbian Post in the Concept of PostIB

The main activity of Serbian Post is the reception, transport and delivery of postal items. Postal items can be roughly divided into items containing information and items containing merchandise or other items. The information in most cases can be received, transmitted and delivered electronically, while goods must be packed, received, transported and delivered engaging a number of different resources. The first category of postal items include letters, postcards, brochures, direct mail, referrals and communications in which the only valuable thing is information and carrier of information (usually paper) has no value or it is negligible. The second category of postal items are letters and packages containing merchandise or other objects that have practical value. So the very content of the mail item has value and the information is only accompanying this shipment.

The current volume of the first category of items is not compromised by implementing PostIB but greatly encourages electronic exchange of information. However, it is important to note that the increased communication across PostIB generates an increased number of postal items containing goods.

If we look at the traditional reception, transport and delivery of goods through the postal items then we can note the following:

- Compliance with regulations on packaging and addressing,
- Defined place of receipt of shipments,
- Defined fees and costs,
- Non-standard terms,
- Low reliability (loss, theft and damage).

However, this approach is logical when it comes to the general offer of universal postal services contract by approach, while in the commercial sector the way of functioning from reception to delivery of such items should be reconsidered. PostIB, as the commercial segment of the Post of Serbia should define the special provisions relating to packing, of addressing and delivery of consignments containing goods, in order for the transfer to be effective, safe and fast. Reception of such items can be made in postal network units or at the vendor and packaging can be adjusted to type and characteristics of the goods. Shipment generated through PostIB, in some areas may be outside of regular postal flows depending on the volume of turnover, characteristics of goods, special requirements in terms of speed, temperature, humidity and so on.

The prices of services for shipments of goods must be flexible and include more factors in the formation. They must be stimulating, with a low profit margin in order to lower the cost per unit of transport and to maximize the total profit with quantity of business. It is important to establish the maximum reliability of the items containing goods, or that the loss or damage is at its minimum.

For example, consider two cases. A farmer in a remote village wants to improve vegetable production, but at his place there is a poor choice of seeds. He went to the nearest

post office and with the help of Post workers and PostIB finds high-quality seeds offer. Since the trip to a big city, where there is a distributor of such goods, is expensive and time consuming, solution is to send chosen seed goods via PostIB. Possible options to pay for the ordered goods is immediately or when it arrives in the mail. For a day or so, our farmer gets his seed with a minimum of expended funds and time. PostIB can be used to inform about the specifics of the selected seeds, agrotechnics which must be applied, herbicide or pesticide to order, and finally to post a trade surplus that he has produced, sell it under the best conditions again with a minimum commitment of time and money. In addition, one should expect higher yields, and therefore higher profits for him and everyone in the chain: the supplier of seeds, the Post and the vegetable buyer because the price is proportional to the invested costs that are naturally lower in this case.

The second case is a car mechanic in a small town that fixes almost all types of cars, and he is the only one in the town. With no shop for spare parts he also does that. Imagine what his warehouse would look like if it had at its disposal all the parts of all the cars maintained. Of course it is not possible. The solution is PostIB. If he has no Internet, he can use a nearby Post and there he can use the PostIB application. Needed spare parts will arrive by mail soon and everyone will be satisfied. A car owner did not have to search for the spare parts on his own, the mechanic has chosen a part that is required with minimal loss of time and price of services is much lower. So again, everyone in the chain can be satisfied.

It is possible to specify a number of similar examples where the synergy of information flow and logistics in supply chains are shown and where postal resources and PostIB produce a new value, new possibility and opportunity for profit.

We can notice several segments of business of Serbian Post, which allows efficient concept of PostIB. It does without saying that the IT resources and their availability are very important as well as the education of the population and industry on the advantages and benefits that we all can accomplish by doing business over PostIB concept. However, when all other possibilities of Serbian Post are included then these benefits are even multiplied. Possibilities of postal transport network are very large and so far consist of one exchange Post office, two international departments, three customs posts, three regional PLC (postal and logistics center), 14 local PLC, 37 transshipment point, 4122postal counters, 3585 delivery regions, 2756 mailboxes and over 15,896 postal compartments. The territory is divided into 107 542 address segments that allow unambiguously defined coordinates of the recipient and organization of transport logistics for nearly 1,000 trucks. Figure 2 shows the main transport flows in the Republic of Serbia.

Another important segment of logistics in the functioning of PostIB is security of payment. Namely, electronic commerce is still not able to solve the problem of security of electronic payments. The Post of Serbia, as the holder of the national payment system, has mechanisms to make the payments for goods and services more secure for the buyer and seller. A long tradition of paying on delivery ensures that the seller will definitely get his money. Business cooperation with commercial banks allows transfer of funds to be made on-line from one account to another, to form a referral payment at home or some other possibility. All other payments, such as PDV, taxes, customs fees, etc. can also be done at the Post Office. This means that in most cases the seller or the buyer can rely on Post of Serbia and PostIB when they need to solve their business problems.

In addition to offering storage capacities, commissioning, packing and addressing along with distribution of goods the offer becomes more attractive and interesting.

Fig. 2. Network for transport postal items

V. CONCLUSION

The idea of PostIB represents essential new approach to supply chain and provides many benefits for end participants in the chain. On the other hand the concept of PostIB creates opportunities to initiate appearance of postal items containing goods and thus achieve remarkable commercial results. The special importance of this concept is reflected in the opportunity to develop rural areas, increasing their competitiveness in the market, stopping the negative demographic processes through higher employment and creating a new economic perspective. The Post as a state institution has an obligation to maximally contribute to that aim and PostIB is one way to make it happen.

REFERENCES

[1] P. Watson, K. C. Gupta, “EM-ANN Models for Microstrip Vias and Interconnects”, IEEE Trans., Microwave Theory Tech.,vol. 44, no. 12, pp. 2395-2503, 1996.

[2] 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.

[3] http://business.benchmark.rs

32-bit Development Platform for Graphical Interfaces

Boyko Kazakov[1], Tihomir Brusev[2], Boyanka Nikolova[3]

Abstract – **In this paper is examined the technology behind 32 bit embedded system with touch screen interface. A key focus of the study is on development of a hardware device based on a Cortex-M3 microprocessor. A printed circuit board (PCB) is designed with Cadence OrCAD and achieved results are examined.**

Keywords – **CORTEX-M3, 32-bit, LCD, Touch screen.**

I. Introduction

Today's embedded systems are built upon 8/16 bit microcontroller architecture. With the availability of new low-power 32-bit architectures, there is an opportunity to increase performance, improve accuracy, and achieve greater power efficiency in these applications. It's also important that, higher processing capability makes it possible to implement new and effective features, including advanced control algorithms and next-generation interfaces such as GUI-based displays, rapid signal processing and capacitive touch sensing.

With the expansion of the Smart phone and personal data assistants their significance is rapidly increasing into the modern-day life.

The wide spread of open source platforms don't have even slightly the capabilities of those devices. And most of the multimedia devices which exist are just too expensive for the average user. Therefore this low cost, graphical interface unit is introduced. Taking full advantage of the 32-bit microprocessors.

Allowing creation of simple graphical applications, and helping understand the principles of graphical interface development. Easy to use pre-developed graphical library are introduced.

The major goals of designing one such system are reviewed and basic conception is introduced in Section II. Key considerations in developing PCB for the hardware with Cadence OrCAD software are shown in Section III.

[1]Boyko Kazakov is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: boyko.kazakov@gmail.com.

[2]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.

[3]Boyanka Nikolova is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: bnikol@tu-sofia.bg.

II. Conceptions and Overview

The 32-bit ARM Cortex M3 processor offers superior code density to 8-bit and 16-bit architecture (Fig.1) [1]. This has significant advantages in terms of reduced memory requirements and maximizing the usage of precious on-chip flash memory. The ARM Cortex-M3 processors utilize the ARM Thumb-2 technology which provides excellent code density. Code size comparison can be seen on Fig. 2. The comparison is created using relative EEMBC Caremark test size [2].

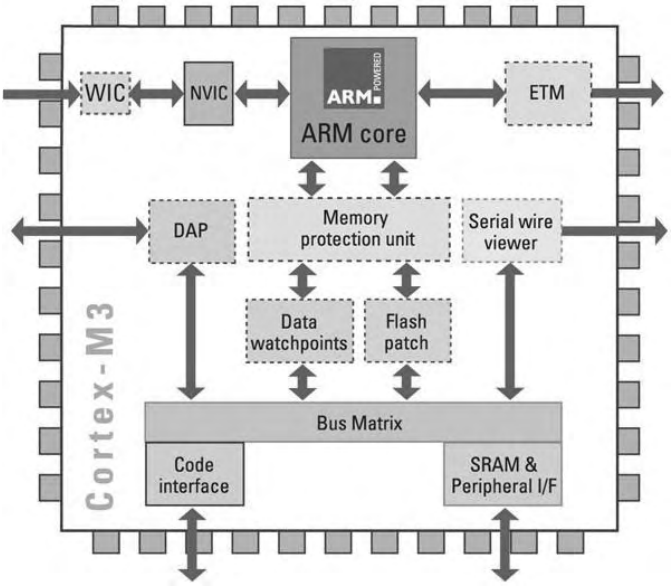


Fig. 1. Cortex-M3 architecture.

The whole picture is not complete without also considering that ARM Cortex-M3 processor instructions are more powerful. There are many circumstances where a single Thumb instruction equates to several 8/16-bit microcontroller instructions. This means that Cortex-M devices have smaller code and achieve the same task at lower bus speed. Comparison between different architecture is shown on Fig. 3. [3].

The demand for ever lower-cost products with increasing connectivity (e.g. USB, Bluetooth, IEEE 802.15) and sophisticated analog sensors (e.g. accelerometers, touch screens) has resulted in the need to more tightly integrate analog devices with digital functionality to pre-process and communicate data. Most 8-bit devices do not offer the performance to sustain these tasks without significant increases in MHz and therefore power and so embedded developers are required to look for alternative devices with more advanced processor technology.

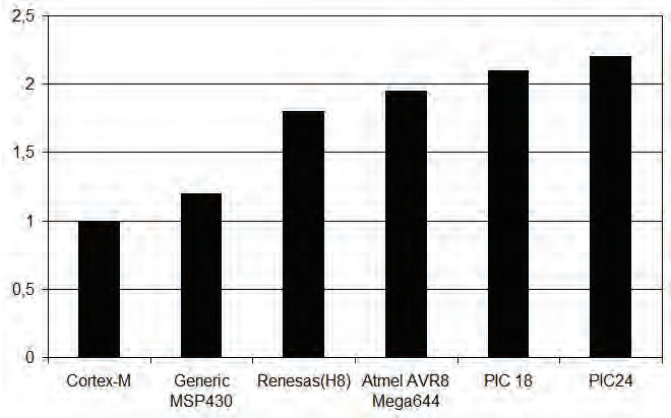


Fig. 2. Code size comparison.

The 16-bit devices have previously been used to address energy efficiency concerns in microcontroller applications. However, the relative performance inefficiencies of 16-bit devices mean they will generally require a longer active duty cycle or higher clock frequency to accomplish the same task as a 32-bit device [3]. The comparison for energy utilization can be seen Fig.4.

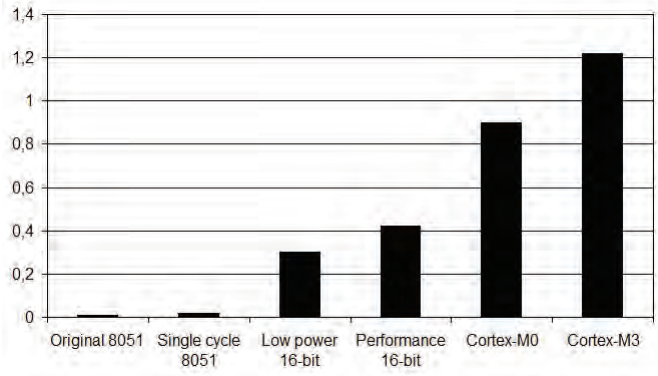


Fig. 3. Performance DMPIS/MHz.

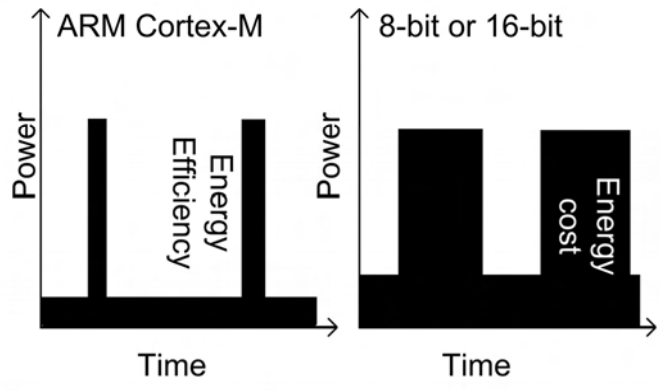


Fig.4 Energy utilization.

A. Device Concept

The designed system must have basic set of hardware needed for development. A simple Quarter Video Graphics Array (QVGA) with embedded LCD controller is chosen mainly on cost measure. The wide variety of low cost displays introduces the opportunity for one cheap and intended for open source communities' device. Resistive touch screens don't offer accuracy, and stylus must be used in order high accuracy to be achieved. Most of the popular developing platforms don't offer mobility. That's why the device must allow autonomous power supply. The device must feature connectivity, in order to be able to be manipulated with. The tendency in microelectronics is miniaturization. Therefore mini USB is suitable for one such system, because of its low volume and size. External memory must be accessible by the user, that's why micro SD connector is available.

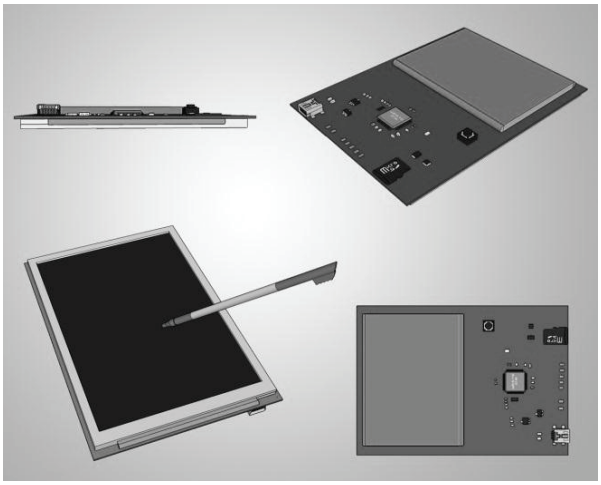


Fig. 5. Device Concept.

A basic conception of the device is designed before PCB creation and manufacturing process. The 3-dimensional view is shown on Fig. 4. The platform is handheld autonomous device that resembles modern cutting edge smart devices.

B. Software overview

In order to achieve high efficiency, image bit-blitting is used. Basicly, area called "window" is defined in the Graphic Display Data RAM (GDDRAM) and the incoming bit stream of filling data which is with predefined direction from the initialization of the display. This means that before graphic data sending, the microcontroller must calculate and store it into an intermediate buffer. Thus smooth frame rates are achieved. This way is much faster than sending first data for the designated pixel, and then its color value. An example of defined window is shown on Fig. 6.

The touch panel will be driven by a special touch controller. The incoming data read by the controller is prescaled and inserted into reasonable range of the display resolution. In order to be achieved accurate reading it is used a median count filter.

On Fig. 7 software conception is proposed. The Main function carries the initialization and functionality of the hardware and Shell functions manage the user applications, developed for graphical interfacing.

Initialization of the display consists in writing specific values to the LCD register. Manufacturers recommended are used.

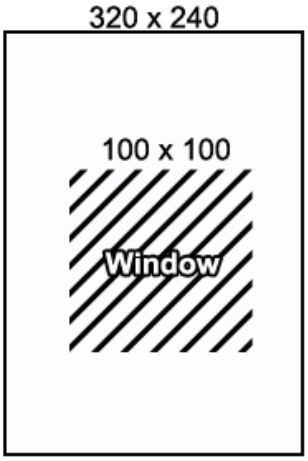


Fig. 6. Defined window area.

Users have a pre-developed graphics library on their disposal, which allows them displaying different shapes or images from the internal or external memory. The performance of the device allows manipulation with compressed or large files, thus introducing many possibilities to the application developers.

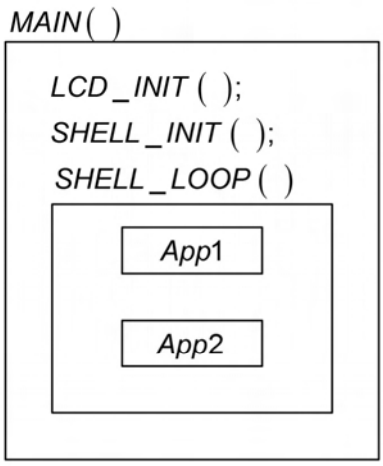


Fig. 7. Software structure.

III. Development Platform for Graphical Interfaces

All the parts chosen for the platform are mainly picked by intense research criteria or experimental

There is a LPC1343 microcontroller unit (MCU) provided on the board which belongs to the 32-bit CORTEX-M3 family from NXP [4]. The microcontroller alone interfaces everything on the board (Fig.8). Being effective in data processing, the LPC1343 is the right choice for development of devices with multimedia content. The MCU uses external 16 MHz quartz crystal oscillator, which is PPL-ed to 72 MHz clock speed. The LPC1343 MCU features 32KB flash memory and 8 KB RAM memory. It provides other integrated modules such as SSP controller with multi-protocol capabilities and fast-mode plus I2C bus interface communication, UART, up to 42 general purpose I/O pins, on chip PHY, 10-bit ADC.

The device features a 320x240 resolution TFT display covered with a resistive touch panel sensitive to touch. The TFT and touch panel together form a functional unit called a touch screen (Fig.8). The touch screen can be used to show images, videos and other graphic content, menu navigation etc. It makes it possible for the user to make interactive apps such as virtual keyboards and tablet inputs. Touch panel itself eliminates the need for additional buttons on the board.

Touch screen backlight is powered with switching transistor and can be adjusted via software with Pulse-Width Modulation (PWM). Thereby power consumption can be reduced when needed.

SSD1289 is an all in one TFT LCD Controller Driver that integrated the RAM, power circuits, gate driver and source driver into a single chip. It can drive up to 262k color amorphous TFT panel with resolution of 240 RGB x 320.

It also integrated the controller function and consists of 172,800 bytes (240 x 320 x 18) GDDRAM such that it interfaced with the MCU through 16-bit parallel interface and stored the data in the GDDRAM.

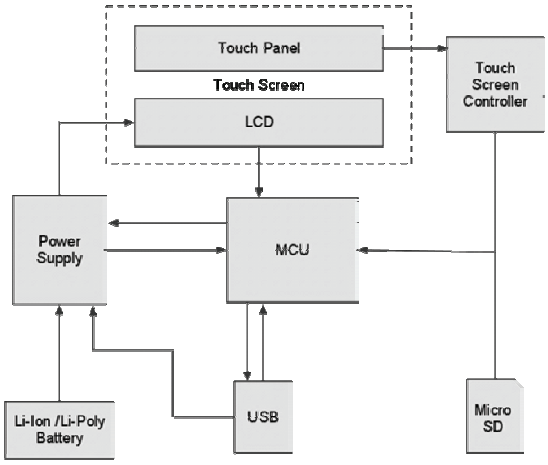


Fig. 8. Device Block Diagram.

SSD1289 embeds DC-DC Converter and Voltage generator to provide all necessary voltage required by the driver with minimum external components. A Common Voltage Generation Circuit is included to drive the TFT-display counter electrode.

An Integrated Gamma Control Circuit is also included that can be adjusted by software commands to provide maximum flexibility and optimal display quality.

SSD1289 can be operated down to 1.4V and provide different power save modes. It is suitable for any portable battery driven applications requiring long operation period and compact size.

An ADS7843 touch screen controller manages the touch panel and is interfaced also via SPI. This multi slave configuration reduces the pin count used on the MCU.

Since multimedia applications are getting increasingly demanding, it is necessary to provide additional memory. There is a build-in Micro SD connector for inserting micro SD card provided for the board. It enables the system to additionally expand memory space. SPI serial interface is used for communication between the MCU and the micro SD card.

There is a 5-pin LDO (Low-dropout regulator) which provides stable 3,3Volts for the device. The LDO with the battery charging unit and 2 Shottky diodes for the turn-on switch represent dynamic power supply, which makes possible, software to control power down of the device, monitoring the charging process of the battery via ADC, and taking full advantages of the power saving options of the LPC1343 MCU. The board may use one of three power supply sources:

- +5V from the USB-VBUS from the USB
- Li ion/li-poly battery connected to a designated connector provided on the board.
- +3.3V from the JTAG connector.

Battery charging unit has status pin. LED is connected to it and indicates when charging is complete.

The Shottky diodes are used for switch debouncing from the initial turn-on. Their absence may cause brown out voltage loops,

There is a mini USB connector provided on the board which is used for connecting to a PC. It allows stable 5 V source for charging. And also allows the device to enumerate as mass storage device (MSD) [5].

Debugging of the MCU can be performed via specially designated JTAG connector.

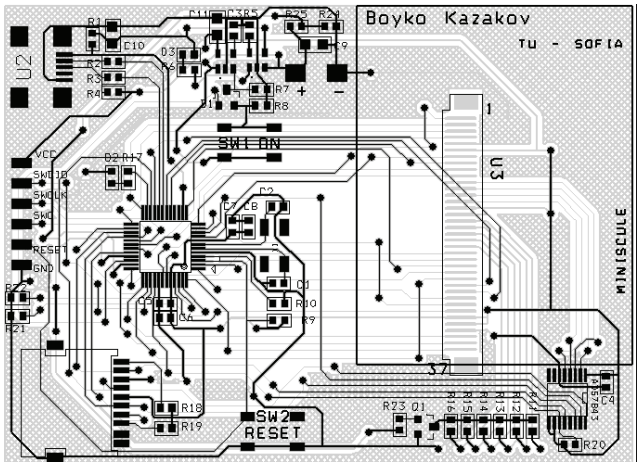


Fig. 9. PCB of the Device

The PCB is created for the device with Cadence Orcad [7] software and follows basic surface mounted technology (SMT) considerations. It features:

- Double layer PCB
- 10 mill tracks /12 mill for tracks involving the power supply, 50 mill via
- Main consideration for the board size and dimension is to fit on the back of the Touch screen. Thus 59x90 mm. is the optimal size for the most 3.2” modules there are.
- 6 pin 100mill SMD footprint is created for the JTAG debugger.

Footprints for the most of the parts are designed with the datasheet recommended dimensions. Special attention is given to routing of data buses, to reduce noise and cross talk influence. Also low via count is abide. A decoupling capacitors are placed also for reducing the noise. Connector for the touch screen is place with accuracy of a millimeter in order the display to be precisely centered on the bottom layer of the board. There is an area on the top layer, left unplaced with components for the battery installation.

IV. Conclusion

In this paper is examined the technology behind 32 bit embedded system with touch screen interface. A printed circuit board (PCB) is designed. The Cortex-M3 microprocessor is proven as suitable choice for such kind of a device. The developed platform can be used for creating of graphical user applications. With its simplicity the system is intended for the beginning of firmware developers.

Acknowledgement

The research described in this paper was carried out within the framework of Contract 102 НИ 199-3.

References

[1] www.codextr.com
[2] www.coremark.org
[3] “Moving from 8/16-bit to ARM Cortex-M” , www.arm.com;
[4] www.nxp.com;
[5] “Cortex-M3 Processor ” , www.arm.com;
[6] Donley D., “Migrating from 8-bit to 32-bit MCUs”, 2010;
[7] http://www.cadence.com.

Session PO7:

PO7 - ELECTONIC COMPONENTS, SYSTEMS AND TECHOLOGIES

Illuminance to Frequency Converter also Used for Conversion of the Ratio between Two Illuminances into a Number of Pulses

Tsanko Karadzhov[1], Ivelina Balabanova[2]

Abstract –**A circuit for conversion of illuminance to frequency and the ratio between two illuminances into a number of pulses has been developed. A simulation of the circuit performance has been carried out by means of PSPICE software and the transfer function of the converter taken by way experimentation.**

Keywords **– Illuminance, Frequency, Converter.**

I. INTRODUCTION

Optical to electric signal converters are integral parts of analogue or digital devices for measuring illuminance or light flux and they have a wide range of application in optoelectronics. This necessitates the design and development of such converters as well as the improvement of their parameters and functional capacity.

[1]Tsanko Karadzhov Technical University of Gabrovo, 4 H. Dimitar, 5300 Gabrovo, Bulgaria, E-mail: karadjov_st@abv.bg
[2]Ivelina Balabanova Technical University of Gabrovo, 4 H. Dimitar, 5300 Gabrovo, Bulgaria, E-mail: ivstoeva@yahoo.com

II. CONVERTER DESIGN

The 555 integrated timer is a specialized integrated circuit designed to generate square pulses of particular duration and frequency of repetition. One of the most frequently used operation modes with this timer is its involvement as astable multivibrator. If some of the components in the mastercircuit are replaced by optoelectronic components,then at the timer output it is possible to obtain pulses whose parameters are optically controlled. Figure 1 shows the design of illuminance to frequency converter also used for conversion of the ratio between two illuminances into a number of pulses.

Three timers type 555 have been used and photo diodes BPX 61 (product of SIEMENS) have been used as photo sensitive components. The photo diodes are included in the master circuit of the first timer – DD1. It controls the other two timers and determines the period throughout which the second and third timer will generate pulses with T=30.2μs and f = 33.11 kHz. The timer's period is approximately defined by means of the following expressions:

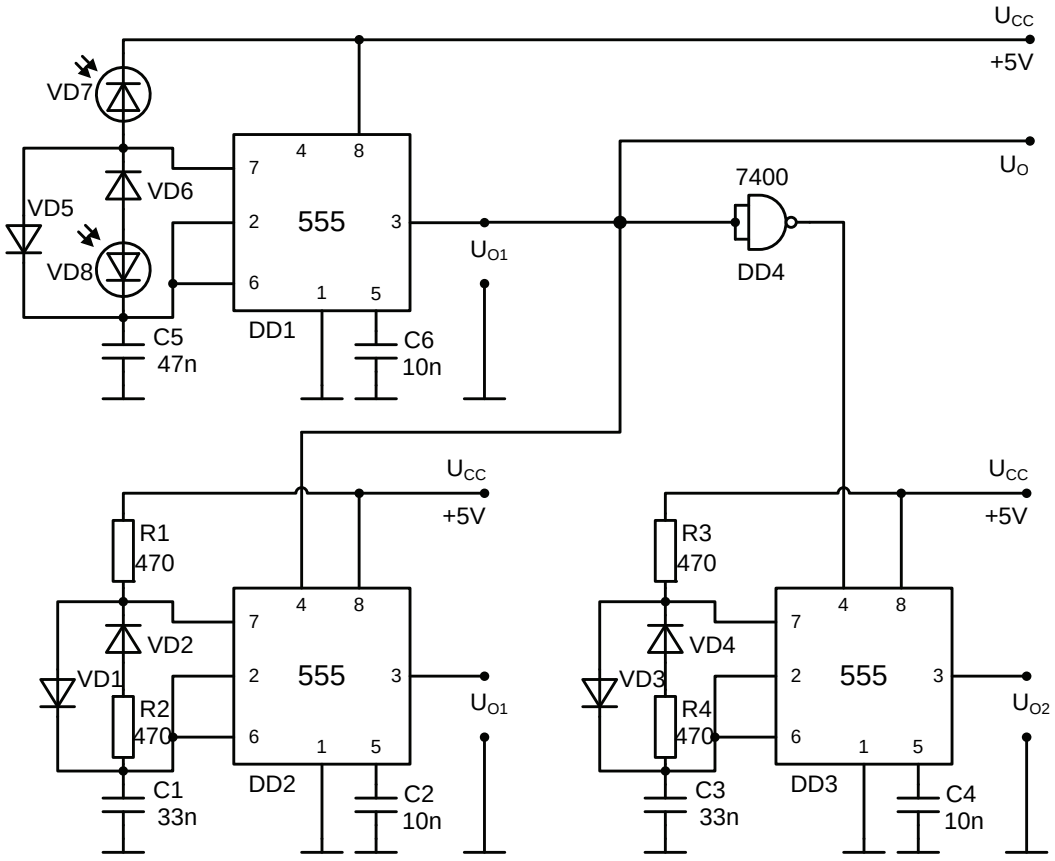


Fig.1 Converter design

Timer 2:

$$T \approx 0{,}7.C_1.(R_1 + R_2) \quad (1)$$

Timer 3:

$$T \approx 0{,}7.C_3.(R_3 + R_4) \quad (2)$$

Timer 1:

$$T \approx 0{,}7.C_5.(R_{VD7}(\Phi_1) + R_{VD8}(\Phi_2)) \quad (3)$$

Time diagrams of the circuit performance are shown on fig.2, fig.3 and fig.4 with different illuminance ratio of the two photo diodes.

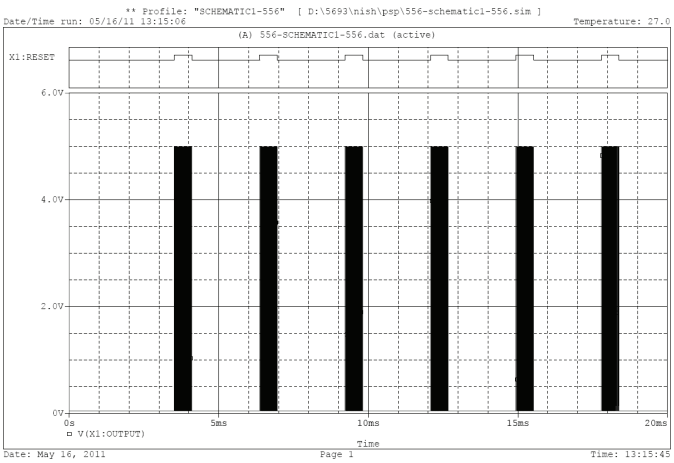


Fig. 2

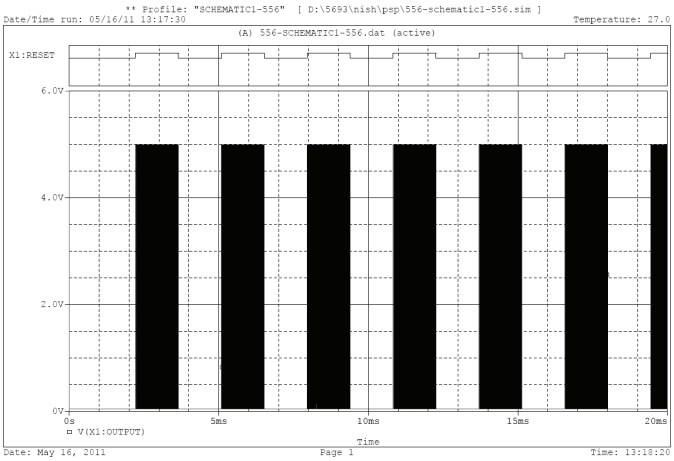


Fig. 3

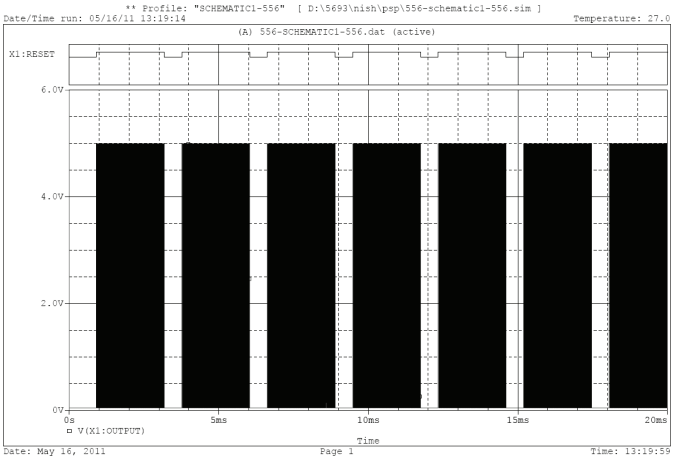


Fig. 4

Fig.5 shows the burst generated by timer 2.

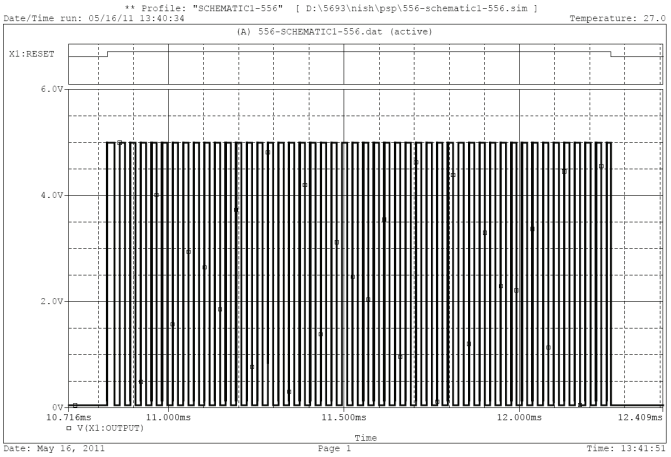


Fig. 5

Both charge and discharge of capacitor C5 occurs between voltage thresholds of the two comparators of the first timer ⅓ $U_{CC}$and ⅔U_{cc} by rule of the following laws.

$$U_{C5} = \frac{1}{3}U_{CC} + \frac{2}{3}U_{CC}\left(1 - e^{\frac{-t}{R_{VD7}(\Phi_1)C_5}}\right) \quad (4)$$

$$U_{C5} = \frac{2}{3}U_{CC}\left(1 - e^{\frac{-t}{R_{VD8}(\Phi_1)C_5}}\right) \quad (5)$$

Table 1 and fig. 6 contain the results from the measurement and the transfer function of the converter plus the dependence of he output frequency from the illuminance of f=f(E)

TABLE I

E, lx	T, ms	f, Hz
100	27.0	37.04
200	13.7	72.99
300	9.25	108.1
400	6.99	143.1
500	5.63	177.6
600	4.71	212.3
700	4.06	246.3
800	3.56	280.9
900	3.18	314.5
1000	2.87	348.4
1100	2.61	383.1
1200	2.40	416.7
1300	2.22	450.4
1400	2.07	483.1
1500	1.93	518.1
1600	1.82	549.5
1700	1.71	584.8
1800	1.62	617.3
1900	1.53	653.6
2000	1.46	684.9

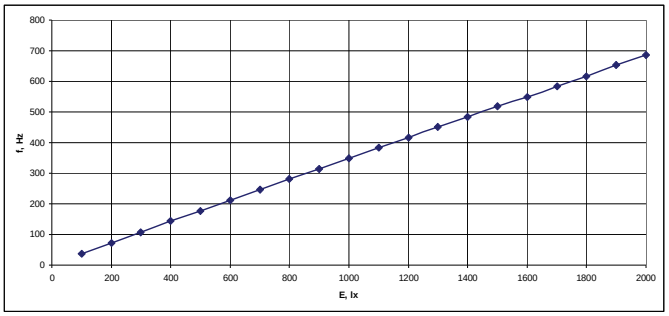


Fig. 6

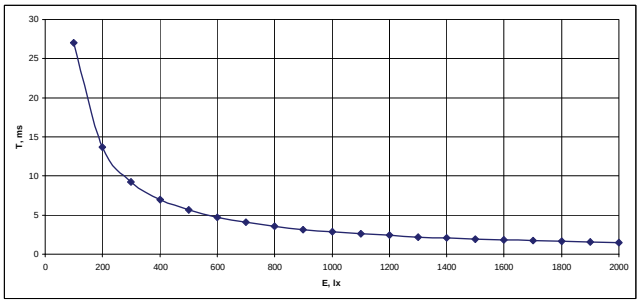


Fig. 7

Fig. 7 shows the dependence T=f(E).

Table II and fig.8 indicate the change in the pulse count(in percentage) within bursts at output U_{OI} depending on E1/E2

TABLE II

E1/E2, %	Брой импулси
10	10
20	20
30	29
40	38
50	48
60	57
70	66
80	75
90	84

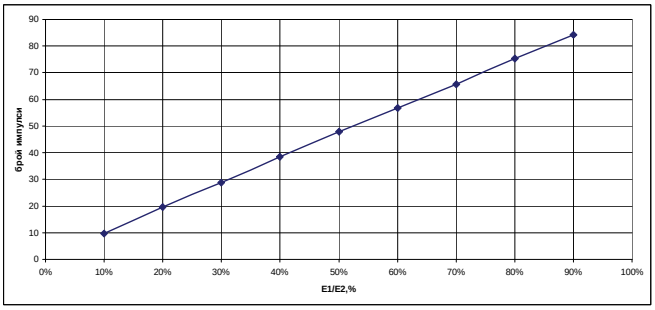


Fig. 8

The frequency of the pulses generated by timer 1 is much smaller than the pulse frequency of the other two timers. This determines the large number of pulses within a burst.

III. CONCLUSION

The newly developed multifunctional converter could be used in lux meters for measuring illuminance or the ratio between two illuminances. Output frequency and and pulse number could vary over a wide range. An advantage of the design is the linear transfer function which results from the linear dependence of the photo flux in the master circuit of timer 1 from illuminance.

REFERENCES

[1] Коев, Ц. К., Ръководство за лабораторни упражнения по Измервателни преобразуватели,ИК"Колонел", Габрово, 2002.

[2] Р. Трейстер, Радиолюбительские схемы на ИС типа 555, Мир, Москва, 1988

Improved Methodology for Design of Magnetic Components

Vencislav Valchev[1], Georgi Nikolov[2], Angel Marinov[3]

Abstract – **In this paper, an improvement to an existing methodology for design of magnetic components is presented. By incorporating the suggested improvements, it is possible to design magnetic components with nanocrystalline and ferrite soft magnetic materials. The component can use either natural or force convection for dissipating the generated by the losses heat.**

Keywords – **Magnetic components, Nanocrystalline, Ferrites.**

I. Introduction

There are lots of design methodologies that are used today ([1], [2]). However, most of them allow a design to be carried out for ferrite soft magnetic materials with natural cooling of the magnetic component. Using one of these methodologies - „Fast Design Approach" [3] as a basis,an improvement is proposed. The chosen methodology combines simple but accurate equations and easy to use graphics. It takes into account the effect of eddy current losses in magnetic component and in the windings. The approach categorizes the design cases into two major cases: saturated thermally limited design and non-saturated thermally limited design. The improvements that we proposed are concentrated in two of allfifteen steps.

II. Calculate the Heat Dissipation Capability P_H(Step 2)

In this step an estimation of the heat dissipation capability of the chosen core (in step 1) is made. The rule of the thumb used is:

$$P_h = k_A a b \quad (1)$$

where

k_A is a coefficient with typical value of 2500 W/m^2;
a and b are the two largest dimensions of the component in [m].

The value for k_A=2500 W/m^2 is selected for a general case of magnetic component design where ferrite materials is used and natural cooling is used. The maximum working temperature is about 85°C, and the maximum ambient temperature as about 60°C. Such temperatures are typical for magnetic components working in confined enclosure.

Our first improvement is more accurate calculation of this coefficient for wider range of temperatures and cooling options. The proposed steps are valid when the following conditions are met:

- The magnetic component is placed in such way, that the cooling air can easily move around it.
- There are no heat sinks mounted on the component.
- Only magnetic components with EE cores are used, with windings on the centre leg.

The following steps are carried out in order to determine the value of k_A.

- Calculate the surfaces involved in the heat transfer;
- Calculate theParameter L as the total distance of the boundary layer;
- Select the speed of the cooling air – v;
- Calculate the total emissivity of the surface of the component ε_T.
- Calculate the heat transfer
- Determine the value ofk_A.

Equivalent surfaces of an EE transformer - S_{conv} andS_{rad}

As there is no heat sink mounted on the component the heat transfer by conduction is neglected. The heat transfer by convection and radiation have different mechanisms, and as a result, two different surfaces are involved in the heat transfer - S_{conv} andS_{rad}. The surface, which is used in the heat transfer by radiation, is reduced compared to the convection surface, because the efficiency of the radiation is decreased in adjacent areas. The surface for convection takes into account all open areas, that can be cooled down by the incoming fluid. The formulas for calculating S_{conv} andS_{rad} are shown in Eq.2 and Eq.3 corresponding surfaces are shown on Fig.1, and the

$$S_{rad} = 2\left(4S_1 + 2S_2 + S_3 + 2S_4 + 2S_7 + 2S_8\right) \quad (2)$$

$$S_{conv} = 2\left(2S_5 + 2S_6 + S_3 + 2S_4 + 2S_7 + 2S_8\right) \quad (3)$$

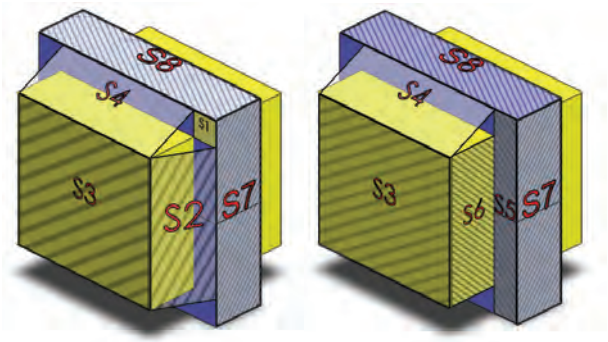


Fig.1Equivalent surfaces of an EE transformer, left for radiation; right for convection

The results for two transformers can be found in Table 1. The first chosen transformer is with core EE80 and ferrite soft magnetic material, while the second one is for nanocrystalline cut core - F3CC0010. As the F3CC0010 is UU type, 2 sets are used to make an EE core.

[1]Vencislav C. Valchev - TechnicalUniversity – Varna, Varna "Studentska" №1, Bulgaria, E-mail: vencivalchev@hotmail.com
[2]Georgi T. Nikolov - TechnicalUniversity – Varna, Varna "Studentska" №1, Bulgaria, E-mail: nikolov_george@hotmail.com
[3]Angel S. Marinov - TechnicalUniversity – Varna, Varna "Studentska" №1, Bulgaria, E-mail: igdrazil@abv.bg

TABLE.1

EQUIVALENT SURFACES OF AN EE TRANSFORMER

Parameter	Unit	E80/38/20	F3CC0010
S_{rad}	mm^2	23820	16080
S_{conv}	mm^2	25580	17196
L	mm	120	94

One can see that the equivalent surface used in the convective heat transfer is about 7% larger, than the surface used for radiation.

Total distance of the boundary layer - L

This parameter is equal to the distance that the cooling air makes around the component. It is usually half of the shortest air path (in the direction of air) - Fig.2. For the transformer shown on Fig.2c), the total distance of the boundary layer is:

$$L = c + 2b - 2(b - f) + 2\sqrt{(b - f)^2 + \left(\frac{e - d}{2}\right)^2} \qquad (4)$$

where:

a, b, c, d, e, f – are the dimensions of the core according to Fig.1

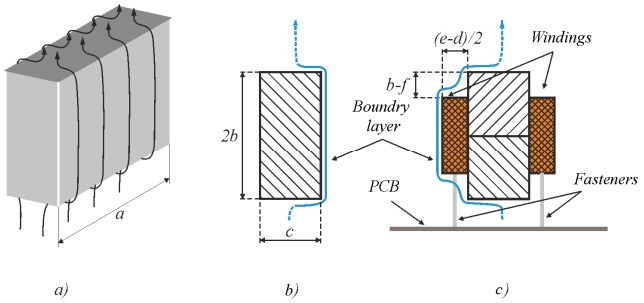


Fig.2 Total distance of the boundary layer

The results for the total distance of the boundary layerof the two cores, with the same orientation as shown on Fig.2c), are shown inTable 1.

Air velocity - v

To calculate the heat transfer, when force cooling is involved, the air velocity is needed. Usually the magnetic component is placed inside the enclosure of the equipment. With a particular fan that generates the airflow, one can calculate the air velocity, from the relationship between pressure rise and volume flow rate. However, it is difficult and time consuming to take into accounts all the variables that can influence the pressure rise. As a result, few experiments were conducted to investigate the air velocity in a real device. An inverter welding unit is chosen – S1700, manufactured by Struna Ltd. For comparison, an air duct with only one entrance and exit for the air is used. Simplified representations of both test setups are shown in Fig.3. The fan that is used to generate the airflow is PMD1212PMB1-A.

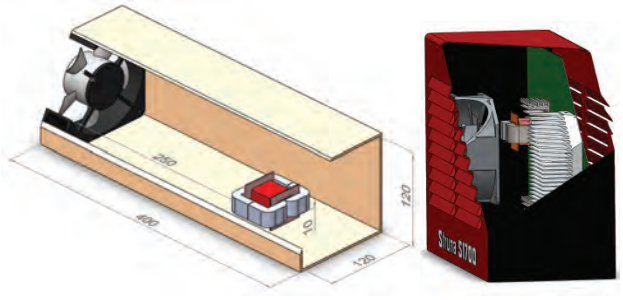


Fig.3Simplified representation of the two test setups

Two different transformers are used for the experiment. One of the transformer have a nanocrystalline core - F3CC0010, and the other a ferrite one - EE80. The winding fills about 60-70% of the window. Both transformers are placed in a way that allows the air to flow all around them. Several experiments are carried out with different rotation and position of the transformers, with or without additional components (PCBs, heat sinks, other large components) that simulates real device. The air velocity is measured at a distance of 1cm in front of the magnetic component with anemometer EA3000 with maximum error ±5%.

Table. 2 shows the results for the air velocity when the voltage of the fan is varied for two typical cases. The experiments are carried out relative humidity of air 65%, temperature 24°C, altitude 91mabove sea level.

TABLE. 2. AIR VELOCITY WHEN THE VOLTAGE OF THE FAN IS VARIED

Voltage of the fan		**V**	5	7	10	12	14	15
Air velocity	Air duct	m/s	1,9	2,7	4,1	4,7	5,4	5,6
	Welder	m/s	1,8	2,6	3,7	4,2	4,5	4,6

One can see that air velocity in the welding unit is about 10% lower than in the air duct. This important decrease in the velocity should be noted when the power dissipation of the magnetic components is calculated.

Heat transfer with forced cooling

When no cooling heat sink is mounded on the component, the conduction heat transfer can be neglected (only small percentage of the heat is transferred by the pins of coil former). The resulting heat transfer can be calculated by Eq.5.

$$q = q_{rad} + q_{conv} = \varepsilon_T \sigma S_{env} \left(T_w^4 - T_{amb}^4\right) + \alpha_c S_{conv} \left(T_w - T_{amb}\right) \qquad (5)$$

where:

q_{rad}, $q_{conv}$is the heat transfer rate for radiation and convection;

ε_T – total emissivity of the material;

σ – Stefan–Boltzmann constant – 5,6704.10^{-8} [W.m^{-2}.K^{-4}];

S_{rad} – the radiating area [m^2];

S_{conv} –area for convection heat transfer [m^2];

T_w–temperature of the surface [K];

T_{amb}–ambient temperature[K];

α_c –convection heat transfer coefficient of the material;

A simplified equation is used to find the convection heat transfer coefficient of the material α_c[4]:

$$\alpha_c = (3,33 + 4,8v^{0.8})L^{-0.288} \quad (6)$$

where:

*L*is the e total distance of the boundary layer;

v – velocity of the fluid.

When the total emissivity of the material is unknown, it can be determined by several different ways – with special paint, strips, or specialized tools. In the particular case, we use a specialized wireless non-contact infrared thermometer (PeakTech 5005USB), that can measure the temperature also with a thermocouple. The thermometer automatically calculates ε_T. The experiments shows ε_T = 0,82 for the tested nanocrystalline material and ε_T = 0,96 for the ferrites.

The expression (6)is consistent with the classical reference [5] up to v=12m/s as well with, "case 2" in [6]. The advantage of Eq. (6) is that it combines both natural ($v=0$) and forced convection ($v>0$) processes.

In the next calculations, it is assumed that the copper windings fill the window completely -the worst cooling case. The isolation between the windings is ignored in the calculations.

The areas S_2, S_3, S_4, $S_6$of the transformer shown in Fig.1 have total emissivity 0,80 (copper with enamel –[7]).

For all cut cores manufactured by Hitachi Metals (12 sizes) and all EE cores from Epcos (34 sizes), the corresponding heat dissipating capability is calculated. In the limited space of this paper, the results only for the cut cores of Hitachi Metals are shown in Table 3.

The following conclusions can be made.

- For natural cooling, the generated losses are dissipated effectively by radiation and convection. For transformers smaller than E32/16/9, more heat is dissipated by convection compared to radiation.
- For forced cooling, the losses dissipated by convection are several times more than those dissipated by radiation.
- When all conditions are same, the transformer with ferrite core can dissipate 5-10% more heat, because of the higher total emissivity.Nevertheless, the transformers with nanocrystalline soft magnetic materials have the advantage of higher working temperature.

TABLE 3. CUT CORES HEAT DISSIPATION CAPABILITY

T_w=120°C, ε_T=0,82; v=2,5m/s	**T_{amb}=30°C**					
	α_c	Q_{rad}	Q_{conv}	P_h	k_A	error
	W/ (m^2.K)	W	W	W	-	%
F3CC06.3	27,3	9,1	33,0	42,1	12809	-0,4%
F3CC0008	27,2	10,1	36,4	46,5	12765	0,0%
F3CC0010	26,3	11,5	40,7	52,3	12042	5,8%
F3CC016A	25,9	12,5	43,2	55,7	12831	-0,3%
F3CC016B	25,2	14,0	47,5	61,5	12209	4,7%
F3CC0020	24,9	15,0	50,1	65,1	12921	-0,8%
F3CC0025	24,5	17,7	58,5	76,2	11337	11,6%
F3CC0032	24,2	18,9	61,4	80,3	11942	7,0%
F3CC0040	23,9	20,1	64,2	84,3	12540	2,4%
F3CC0050	23,2	26,5	83,2	109,7	10342	19,6%
F3CC0063	23,0	28,0	86,6	114,6	10806	16,0%
F3CC0125	21,8	40,4	118,6	158,9	10426	19,2%

To find the value of the coefficient $k_A$the following equation is used:

$$k_A = \frac{P_h}{2ab} \quad (7)$$

where:

$P_h$is the calculated dissipating capability of the magnetic component;

a, b– are the dimensions of the component according to Fig. 2.

The results presented in Table 3 are analyzed andF3CC0008 is chosen as a reference. In the calculations later, its value of $k_A$is used. Using this coefficient has the advantage of very simple calculatnion of the heat dissipation capability of the core. The error from using one coefficient for all cores is shown in the last column of Table 3. The same analysis is done for the EE cores and E20/14/5 is selected as a reference. On the next graph are presented the results for this error for both cut cores and EE cores.

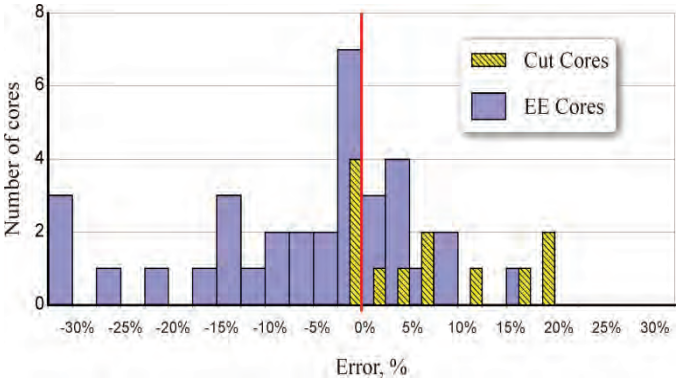


Fig.4 Distribution of the error from using one k_A: T_w=100°C (EE cores), T_w=120°C (cut cores), T_{amb}=30°C, v=0m/s, εT=0,96.

About 30% from the EE cores and approximately 40% from the cut cores have error less than ±2,5%. The largest error is observed when very little or very large cores are used, compared to the reference ones.

In all cases when high accuracy is required, it is best to calculate the results for the particular case. When using k_A some additional errors can influence the result. Sources of such errors are manufacturing tolerances, additional isolation, partial filling of the winding window and others.

Using the results for heat dissipating capability, two tables are built, that allows quick and easy determination of the coefficient k_A. The tables are for maximum ambient temperature of 30°C, but similar graphs can easily be drawn for any ambient temperature.

TABLE4. COEFFICIENT K_A ACCORDING TO THE WORKING TEMPERATUREAND AIR VELOCITY, FOR NANOCRYSTALLINE CUT CORES, TAMB= 30°C

Nanocrystalline cut cores		Air velocity,v[m/s]					
		0	1	2	3	4	5
Maximal working temperature,Tw, [°C]	70	2080	3680	4860	5930	6930	7880
	80	2660	4660	6140	7480	8720	9910
	90	3260	5670	7450	9050	10540	11970
	100	3900	6700	8780	10650	12390	14050
	110	4560	7770	10140	12280	14270	16170
	120	5270	8860	11530	13940	16180	18320
	130	5990	9990	12960	15630	18130	20500

TABLE5. COEFFICIENT K_A ACCORDING TO THE WORKING TEMPERATUREAND AIR VELOCITY, FOR FERRITE EE CORES, TAMB= 30°C

Ferrite EE cores		Air velocity,v[m/s]					
		0	1	2	3	4	5
Maximal working temperature,Tw, [°C]	70	2430	4390	5850	7160	8380	9550
	80	3110	5560	7380	9020	10550	12000
	90	3810	6760	8940	10900	12740	14480
	100	4550	7980	10530	12820	14960	17000
	110	5320	9240	12150	14770	17220	19550
	120	6120	10540	13810	16760	19510	22130
	130	6960	11870	15510	18780	21840	24750

III. FINDING THE PEAK INDUCTION $B_{P,DATA}$

In the original design methodology, the peak induction is taken from the datasheet of the material for specific temperature, frequency and core losses. However, this can lead to serious errors. The main reason for this is that the voltage waveform can influence the core losses. Usually the manufacturers' datasheets are for sinusoidal waveforms, while most of the power electronics nowadays work with square waveforms. Articles like [8], [9], [10]discuss this problem. Using the mathematical models for the core losses, proposed by the authors in [8], the peak induction $B_{p,data}$ can be calculated.

Using the core loss model for nanocrystalline and ferrite soft magnetic material is the second improvement to the "Fast design approach".

IV. CONCLUSION

In this paper, an improvement to the existing "Fast design approach" is proposed. Two of all fifteen steps are modified. As a result, the design of magnetic components with ferrite and nanocrystalline soft magnetic materials, with or without forced cooling is possible with the proposed improved methodology.

ACKNOWLEDGEMENT

The paper is developed in the frames of the project 'Development of specialized scientific infrastructure for investigation of wind and solar energy potential', № Д 002-48/10.12.2008, Ministry of Education, Youth and Science, Fund 'Scientific Research'.

REFERENCES

[1] E. Snelling, *Soft ferrites: properties and applications, Butterworth- Heinemann*, 2-nd edition, 1988.
[2] H. Whittington, B. Flynn, D. Macpherson, *Switched Mode Power Supplies: Design and Construction*, John Wiley & Sons, 1993.
[3] A. Van den Bossche, V. Valchev, *Inductors and Transformers for Power Electronics*, Boca Ration USA, CRC Press, ISBN: 157-444-679-7, 2005.
[4] V. Valchev, A.Van den Bossche, "Accurate natural convection modelling for magnetic components", Microelectronics Reliability, vol.43, №5, pp.795–802, 2003.
[5] P. Richter, "Elektrische Mashinen Allgemeine Berechnungselemente, Die Gleichstrommaschinen", Springler, Berlin, pp.318–321, 1924.
[6] A. Van den Bossche, V. Valchev, J. Melkebeek, "Improved thermal modeling of magnetic components for power electronics", // European Power Electronics Journal – EPE, vol.12, №2, pp.7–11, 2002.
[7] J.P Holman, , *Heat Transfer*, 8th ed., McCraw-Hill, 1997.
[8] G. Nikolov, V. Valchev, "Power loss model for nanocrystalline and ferrite soft magnetic materials with squarewave voltages", Annual Proceedings of the TechnicalUniversity in Varna – ISSN: 1311-896X, Varna pp.97-102, 2009.
[9] V. Valchev, G. Nikolov, "Comparison of nanocrystalline magnetic materials and ferrites used for power electronics transformers' – ICEST08, Nis, Serbia, 24-26 June, 2008, pp. 276-279.
[10] Jieli L., T. Abdallah, C. Sullivan, Improved calculation of core loss with nonsinusoidal waveforms, Industry Applications Conference, 36th IAS Annual Meeting, Conference Record of the 2001 IEEE vol.4, 30, pp.2203 – 2210, 2001.

Galvanomagnetic Device for Angular Displacement Measurement

Nikola D. Draganov[1], Totka A. Draganova[2]

Abstract – **Galvanomagnetic devices are found a very wide application for linear and angular measurement in spheres of instrumentation, automatics, motor industry, building machinery, geology, shipping. They are based on galvanomagnetic sensor for angular displacement and unit for information processing and visualizing. Block diagram and principled electrical scheme, operation algorithm and conversion characteristics of galvanomagnetic device for angular displacement measurement have been presented in this article.**

Keywords – **Galvanomagnetic sensors, Hall elements, Displacement measuring, Angular and Linear measuring.**

I. Introduction

Angular displacement measurement is pressing problem which is solved by creating of new more modern sensory mechanisms and modules for information collection and processing. Hall elements are most spreaded galvanomagnetic elements which enable on their basis to create sensors for angular displacement. They have high sensitivity, linear conversion characteristics and high operational reliability.

The problem connected with collection, treatment and indication of the information obtained by the sensors is very pressing for engineering practice. Using modern element base is possible to achieve good results in magnetic quantities measurement.

The purpose of the present elaboration is to create and investigate a galvanomagnetic device on a basis of experimental galvanomagnetic transducer for angular displacement with Hall elements and established analog-to-digital converter (ADC). The device is designed on a basis of a possible small number of electronic elements and enables easy measurement of angular displacement in one plane.

This elaboration can be a base of later investigations and implementations.

II. Presentation

Hall elements application for measurement of magnetic field, electrical and unelectrical quantities are wide discussed in [5].

[1]Nikola D. Draganov is with the Technical University of Gabrovo, Hadgi Dimitar, str.4, 5300 Gabrovo, Bulgaria,
E-mail: niko_draganov@mail.bg

[2]Totka A. Draganova is with the Technical University of Gabrovo, Hadgi Dimitar, str.4, 5300 Gabrovo, Bulgaria,
E-mail: totka_koeva@mail.bg

Block diagram of a synthesized device is depicted in Fig 1.

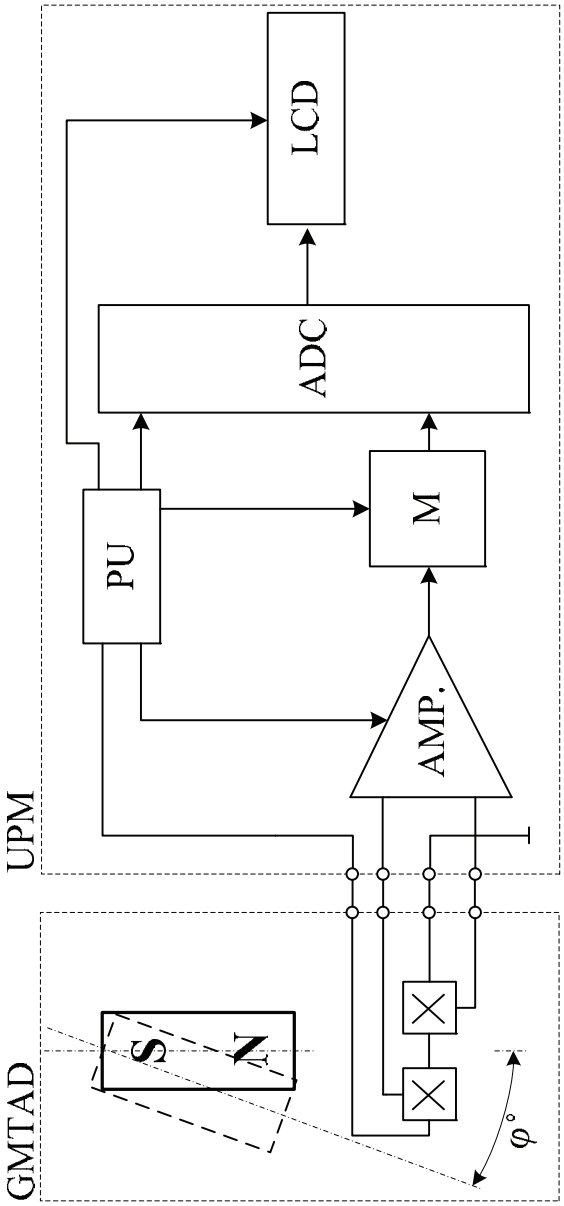


Fig.1. Block diagram of galvanomagnetic device for angular displacement.

It is composed of two modules – galvanomagnetic transducer for angular displacement (GMTAD) and unit for processing and measuring (UPM).

Module GMTAD consists of two uniform Hall elements, of the type VHE101, and a permanent magnet. Hall elements are immovably installed at a distance a=1mm in relation to a

permanent magnet which depending on purpose and operating way can turn if it is fixed in a center or can incline if it is fixed in upper edge. Detailed description of offered constructed variant of galvanomagnetic sensor for angular displacement is made in [1, 2].

GMTAD operation is based on double energy transformation. Angular displacement alternations is transformed into magnetic inductance alteration which by means of Hall elements is converted into electrical signal and magnetic field influence angle is changed [2]. Therefore the appeared angular displacement is detected by GMTAD and the obtained output signal is proportional to angular displacement.

Module UPM is designed from input differential amplifier (AMP), former (M), analog-to-digital converter (ADC), liquid crystal indicator (LSD) and power supply (PU). By means of this module is fulfilled sensory signal processing, transformation and measurement.

A schematic circuit diagram is depicted in Fig. 2.

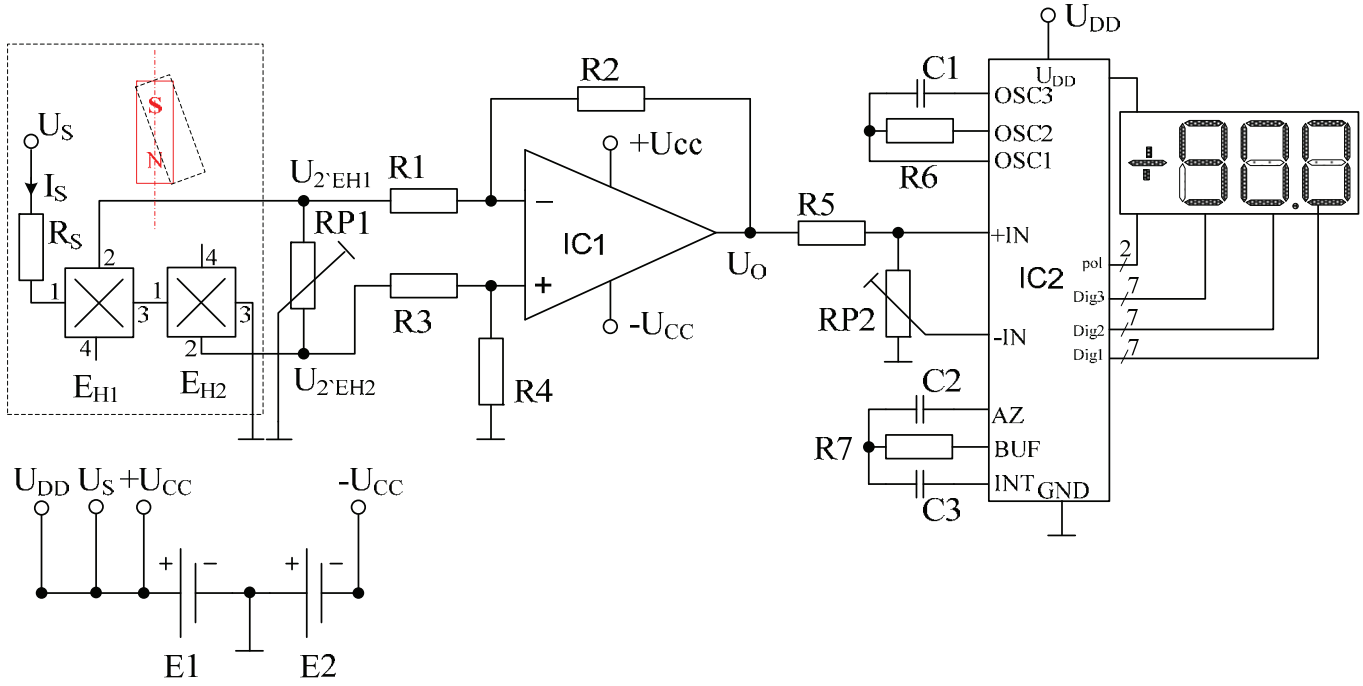


Fig. 2. A schematic circuit diagram.

Galvanomagnetic transducer for angular displacement is consisted of Hall elements E_{H1} and E_{H2}, of the type VHE101. They are supplied from accomplished by the high-resistance resistor R_S generator. The operating point of both connected in series galvanomagnetic elements depends on value and power of this resistor which is choice with parameters R_S =1kΩ and P_{RS}=0,5W. In this case Hall elements are supplied with rated control current I_S=5mA at which they have maximal magneto-sensitivity and more linear conversion characteristics.

A sensory signal (Hall voltage) received direct from both homonymous electrodes of Hall elements (pin 2 of $E_{H1}$and pin 2 of E_{H2}) is delivered to differential amplifier inputs. It is constructed on the basis of operational amplifier LM358. The inverting input of the operational amplifier (IC1) is enveloped in negative feed-back by resistor R1 and R2. Resistors R3 and R4 deliver symmetry on no inverting input.

At direct current I_S=5mA and without a slope (φ°=0) by means of potentiometer RP1 a voltage asymmetry between both homonymous electrodes of Hall elements (pin 2 of E_{H1} and pin 2 of E_{H2}) is eliminated (U_H=0). Amplified sensory signal is given to input of 10 Bit ADC realized by special integrated circuit ICL7106. It has high input resistance, good conversion characteristics, simple setting and possibility to indicate results on liquid crystal display without additional driver. By means of voltage divider consisting of resistors R5 and PR2 device setting and calibrating is provided. Clock frequency of ADC is set up by group R6 and C1. Capacitors C2, C3 and resistor R7 set up in ADC processing results refreshing time. Two dry batteries E1 and E2 (9V) are used to supply a circuit.

When a supply power is connected without applied incline the permanent magnet magnetic field will act in the same degree on both magneto-sensitivity elements E_{H1} and E_{H2}. As a result a voltage difference applied to the differential amplifier inputs (U_{H22}) will be zero. An angular displacement causes an incline of GMTAD permanent magnet (Fig. 1) to one of both Hall elements in according to displacement direction. Magnetic field begins to act more on one Hall element and less on other (Fig. 2). A voltage difference appears between pin 2 of $E_{H1}$and pin 2 of E_{H2}– $U_{2`EH1}$<$U_{2`EH2}$. This difference is amplified by differential amplifier (IC1) and by means of resistor R5 and potentiometer RP2 a reading of GMTAD is set up in degrees. If GMTAD is tilted to another direction the circuit operation will be analogous but the display will show a sign of minus. When angular displacement is changed output voltage difference between

pin 2 of E_{H1} and pin 2 of E_{H2} is now with opposite polarity, i.e. $U_{2`EH1}>U_{2`EH2}$ (Fig. 2).

GMTAD is investigated and conversion characteristics are obtained at the angle incline $\varphi°$ = (-90°÷+90°). The conversion characteristics families $U_O=f(\varphi°)$ obtained by experiment are depicted in Fig. 3 for three control current values I_S=1; 5; 10mA and a constant temperature (T_O=const).

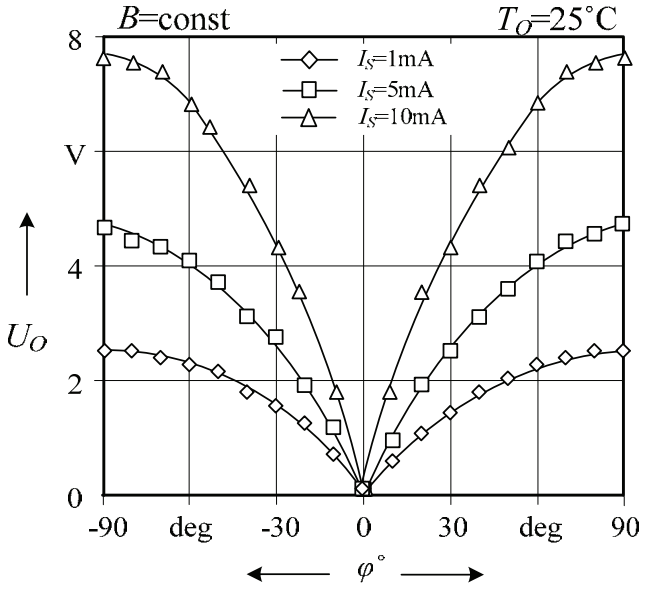


Fig. 3. Conversion characteristics $U_O=f(\varphi°)$.

The analysis shows that they are non-linear over the entire range of angle change and their symmetrical disposition in the first and second quadrants of coordinates. The biggest characteristic slope is obtained at initial angular displacement. For instance at $\varphi°$= (0 ÷ 60°) the slope is S_A= 0,03; 0,09; 0,065V/deg at I_S = 1; 5; 10mA respectively. At bigger angles displacement $\varphi°$= (60°÷ 90°) the slope is S_A= 0,01V/deg at I_S= 10mA, i.e. it is 6 time smaller than at small angular displacement.

III. Conclusion

A device for angular displacement measurement on the basis of galvanomagnetic transducer with Hall elements, differential amplifier and analog-to-digital converter has been synthesized and investigated.

Measuring block has been realized by means of conventional analog-to-digital converter (ICL7106) with accuracy (±5%) in view of there are not requirements to electronic elements limits and they do not influence on measurement accuracy.

The power supply can be realized external or by batteries which makes device portable.

Experimental conversion characteristics $U_O=f(\varphi°)$ of galvanomagnetic transducer for angular displacement are obtained at I_S=const and T_O=const. They show a wide change of measured parameter ($\varphi°$= -900 ÷+900).

The circuit design is characterized with accessible accomplishment, simple setting and calibration which opens up it to a wide group of specialists.Created galvanomagnetic device for angular displacement can find various application as electronic leveling appliance in construction, for incline measurement of moving means of transport, for determination of freight in a cargo ship and etc.

References

[1] Draganov, N. Experimental Galvanomagnetic Transducers of Linear and Angular Offset – Part II. Proceedings of papers – ICEST, Ohrid, Macedonia, 23-26 June, 2010.

[2] Draganov, N. Experimental Galvanomagnetic Transducers of Linear and Angular Offset – Part I. Proceedings of papers – ICEST, Ohrid, Macedonia, 23-26 June, 2010.

[3] Draganov, N., A.Aleksandrov. Galvanomagnetic Field Transducer with Analog Output. Proceedings of papers – ELECTONICS – 10,Sofia, Bulgaria, 28 May2010.

[4] Draganov, N. Research of Joint Operation of Hall Elements. Proceedings of papers – UNITEX-07, Gabrovo, Bulgaria, 23-24 November, 2007, Vol 1, p.210-214.

[5] Roumenin, C.S. Solid Etate Magnetic Sensors, Amsterdan.

Based of AMR Sensors Device for Multichannel Contactless Measurement of AC Current

Nikola D. Draganov[1], Totka A. Draganova[2]

Abstract-**Measurement and visualization of the AC current is fulfilled on the base of the AMR sensor and microcontroller PIC16F874A offered by Microchip®. Presented device has automatic mode of work. It can be connected to PC by means of the serial interface. Control algorithm enables collection, treatment, preservation and visualization of the information.**

Keywords – **AMR sensors, Contactless current measuring, microcontrollers, digital signal processing.**

I. INTRODUCTION

There are diverse electrical quantities. One of they is an electrical current. A lot of indirect and direct methods for its measurement are known. Very spread is indirect contact less method by a measurement of a magnetic field created by a flowing through a conductor electrical current. Connected in parallel bridge anisotropic magnetoresistors (AMR) are widely applied to contact less measurement of an alternating current in the modern installation. They have high sensitivity, wide frequency band, good linear characteristics and high reliability [6, 8].

The problem connected with collection, measurement, treatment and indication of the information obtained by the sensors is very pressing for engineering practice. Using modern element base is possible to achieve good results in an electrical current measurement.

The purpose of the present working out on a basis of AMR and microcontroller is to create a galvanomagnetic device which fulfills multichannel independent AC current measurement. It can be a base of later investigations and elaborations.

II. PRESENTATION

A device for a four channel independent AC current measurement is made. Its block schematic diagram is depicted in Fig.1. It is composed of measuring unit (MU), microcontroller (CPU), keyboard (KB), digital display (LCD), memory (EEPROM), serial interface (RS232) and power supply (PU). A measuring block is composed of sensor units (SU) which detect a created by a measured current I_{TEST} magnetic field and transform it in a electrical signal again. The SU number is defined by a microcontroller hardware. In a case of an utilized in this elaboration microcontroller PIC16C874A permissible SU number is five [5, 7].

Each sensor block is composed of magnetosensitive block (AMRS) and measuring converter (MC). AMRS is created on the basis of anisotropic magnetoresistance sensor of the type ZMY20M fixed immediately on network measuring conductor. By means of this sensor the created magnetic field is transformed to a voltage proportionally to an electrical current I_{TEST} value [3]. A schematic circuit diagram of galvanomagnetic device for four channel measurement of AC current is shown in Fig. 2.

A magnetoresistance sensor ZMY20M represents four magnetoresistors made on the base of a thin film permaloy connected in bridge circuit together in one case. The output

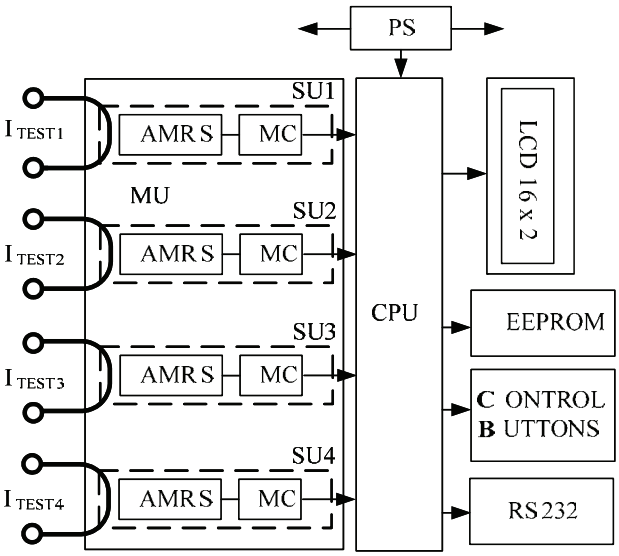


Fig.1. Block schematic diagram.

voltage of this kind sensor is proportionally to an applied external (working) magnetic field in direction B_Y. For the steady sensor operation it is necessary to apply perpendicular another magnetic field B_X created by a fixed in the case constant magnet. A conversion characteristic $U_O=f(B)$ at U_B=const is depicted in Fig. 4 [5, 7]. The sensor unit can transform AC current only in one range (I_{TEST}= 0÷10A). With a view to improve a conversion characteristic stability and toincrease a voltage level on the SU output a measuring converter (MC) is connected.

[1]NikolaD. Draganov is with the Faculty of "Electrical Engineering and Electronics" of Technical University of Gabrovo, Hadgi Dimitar, str.4, 5300Gabrovo, Bulgaria,
E-mail: niko_draganov@mail.bg

[2]Totka A. Draganova is with the Faculty of "Electrical Engineering and Electronics" of Technical University of Gabrovo, Hadgi Dimitar, str.4, 5300 Gabrovo, Bulgaria,
E-mail: totka_koeva@mail.bg

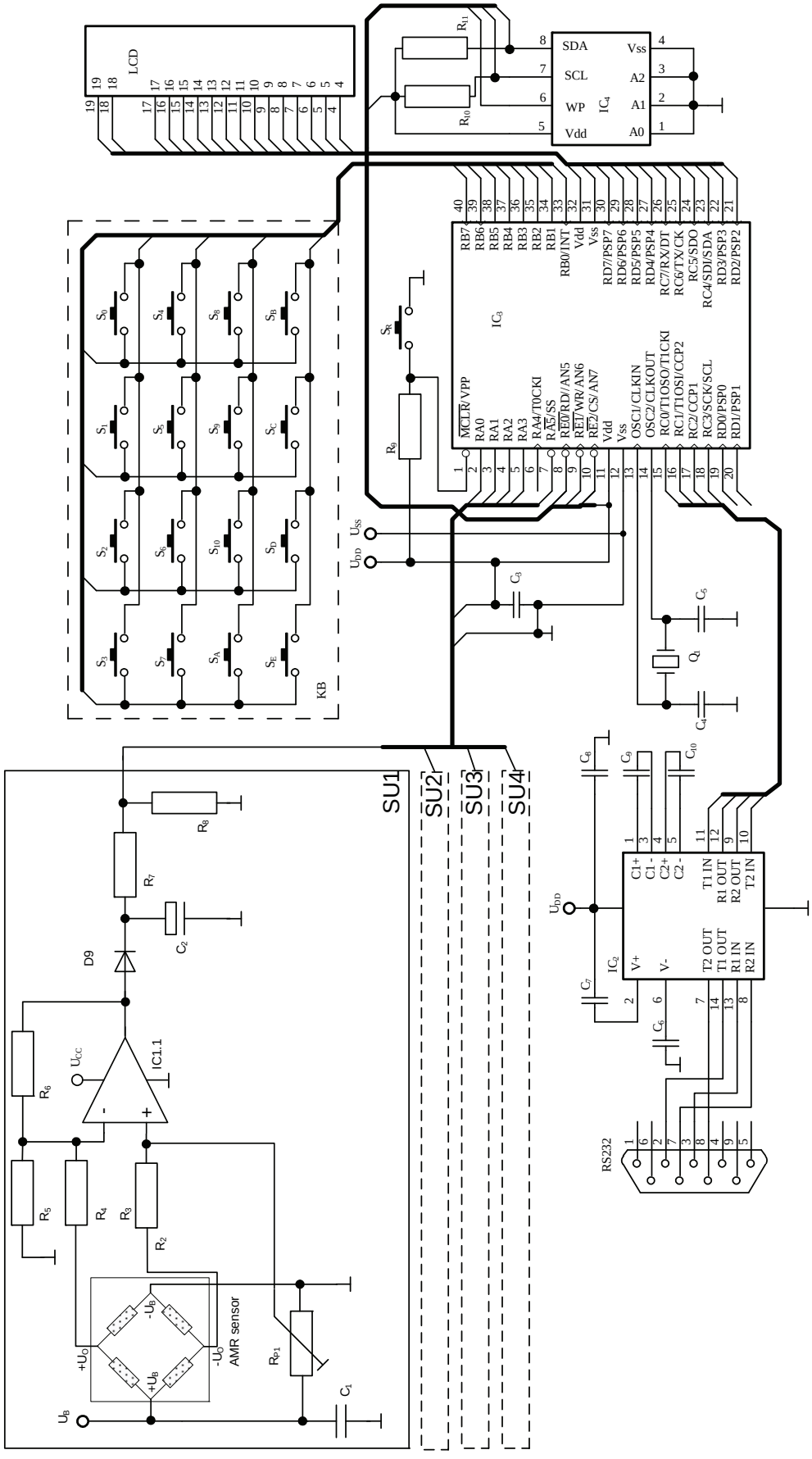


Fig. 2. A schematic circuit diagram

It is realized as classical DC amplifier with gain K=10. So the output information signal maximum value is U_{SUOMAX}= 2,51V at current I_{TEST}=10A (fig. 5).

By means of four independent SUs the galvanomagnetic device enables contact less measurement of four independent current signals. They can be receive as from one so from several signal sources simultaneously without to lose their accuracy by measurement. The received from SU signal is changed from 0 to 2,5V and therefore a reference voltage of the analog-to-digital converter (ADC) is chosen $U_{ref.}$= 5,1V. If a 8 bit discrete is used a minimal measured voltage will be U_{SUOMIN}=0,02V.

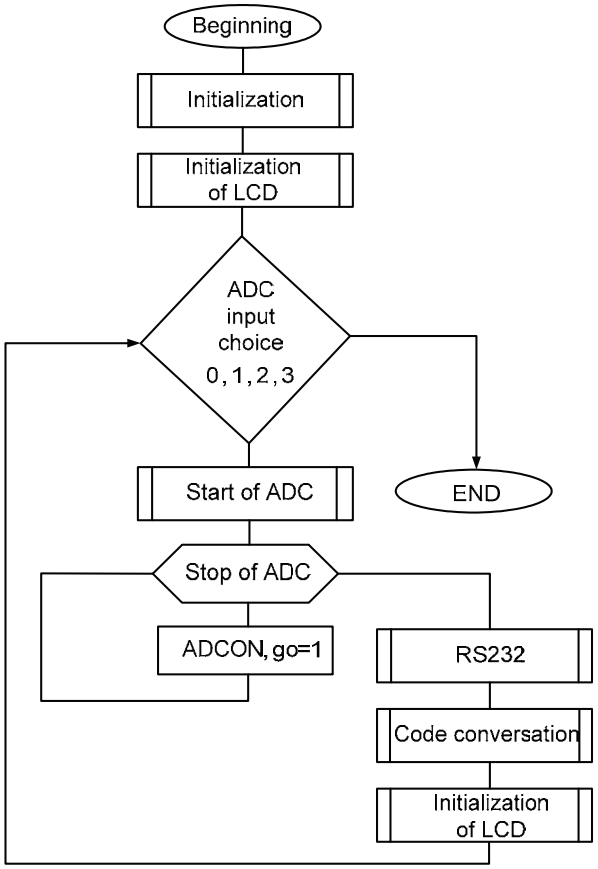


Fig. 3. Main program algorithm

Control organization, processing and monitoring are fulfilled by the main block realized by microcomputer PIC16F874A (Fig. 2).

Accuracy, reliability and measurement parameters are improved thanks to its functional possibilities at a decreased price and dimensions of the device.

The device operates at 20MHz frequency. A microcontroller internal memory (4kb) is enough for normal device processing. An additional 256kb memory of type 24LC256 is intended to buffer a large quantity of information obtained during a measurement of slow changed signals and later their values to send through serial interface RS232 to a personal computer (PC).

The device is equipped with 16 button keyboard which enables to extend its possibilities to assign an automatic or manual channel change and a time of results indication.

By means of intelligent LCD module the measured quantity values are indicated on 16 range display where for each channel the indications take turns in 3 seconds [2].

When power supply is turned on the necessary hardware resources are configured:

- Leads RA0÷RA3 and RA5 are configured as analog inputs through them the sensor unit signal are brought in;
- The keyboard control is fulfilled by port B as their leads (RB0÷ RB7) are configured like inputs;
- Through Port C leads RC0÷RC4 a galvanomagnetic device is connected to PC by RS232. The dispatched values are used for a graphical visualization in different time interval of a measured current change;
- To take out an information from a measurement on LCD display all leads (RD0÷RD7) of Port D are configured like outputs;
- Port E is used to connect a microcontroller to an external memory.

The measurement begins after an initialization fulfillment according to the depicted in Fig.3 algorithms. A measuring input is firstly chosen. Analog-to-digital converter (ADC) treats a input signal in time of 16 seconds. The obtained in binary code signal is sent to PC after that a measured quantity value is transformed in decimal code, respectively in ASCII code and is indicated on LCD screen during 3 seconds.

A measuring cycle of one channel depends on an utilized approach of a filtration. By means of software filter it is achieved an input analog average signal with a view to decrease accident fluctuations influence over a measured electrical current [2-5].

A second channel voltage measurement begins without delay after a finish of a foregoing one. So consecutively the four channels are read. The whole measuring cycle includes the time from a first channel switching on to the last one switching off together with the indication time of 14 seconds.

A measured quantity can be indicated on LCD module as electrical current or magnetic induction unities. By means of a microcontroller tabular transformation is formulated a linear dependence between a magnetic field and a obtained by a sensor output voltage.

III. CONCLUSION

A galvanomagnetic device forfour channel independent measurement and visualization of alternating electrical current on the base of magnetoresistance AMR sensor of type ZMY20M and microcontroller PIC16F874A produced byMicrochip® (Fig. 6) has been created.

A block schematic diagram and schematic circuit diagram have been elaborated. A main program algorithm has been submitted.

A measurement accuracy has been increased by means of software filter which eliminates accidental input quantity peak alterations [2].

The device enables to measure slow alterable in time electrical signals.

The external memory enables to keep large information.

It has been established measuring device errors because of:

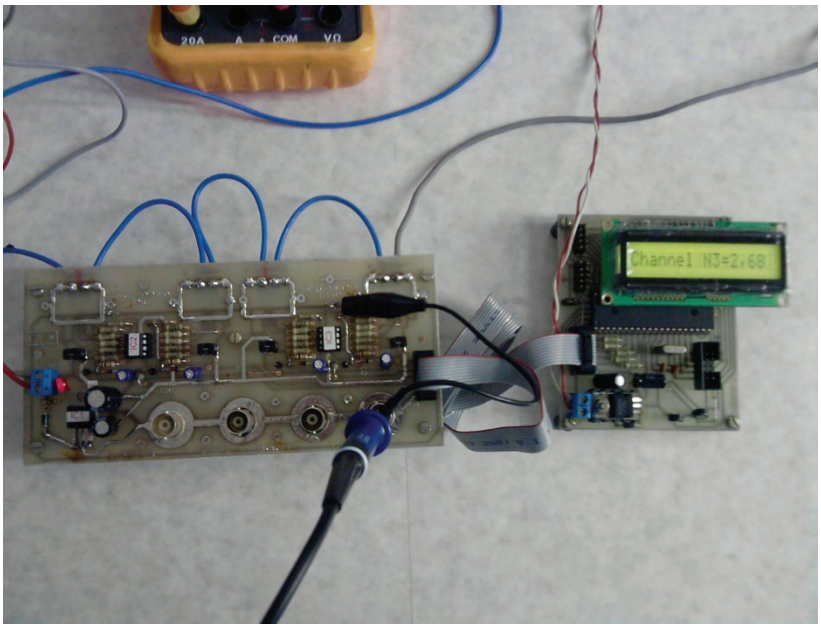


Fig.6. Model of device for four channel contact less AC current measurement.

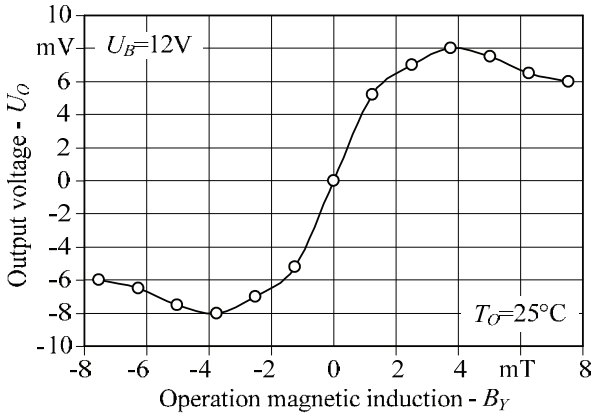


Fig.4. A conversion characteristic $U_O=f(B)$ at U_B=const of magnetoresistance sensor ZMY20M [8].

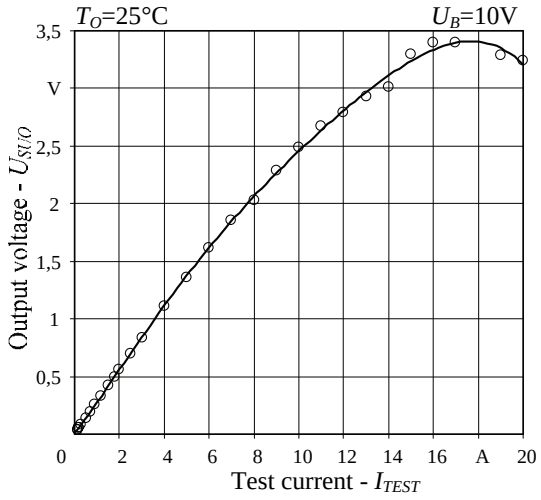


Fig.5.A conversion characteristic U_{SUO}=f(I_{TEST}),U_B=const and U_{CC}=const of sensors units.

- an analog-to-digital transformation which is 0,5% at maximum input ADC voltage U_{INMAX} = U_{SUO} = 2,5V and a discretization step 0,02V;
- analog block in result of temperature sensitivity of a magnetoresistance unit;

It has been made an error reduction by means of embedded in microcontroller program software filter and a sensor unit temperature compensation which decreases an error caused by surroundings temperature change.

Software provided operations have been controlled by 16 button keyboard. An interface has been realized by RS232 for a communication with PC.

The created galvanomagnetic device for contact less AC current measurement is widely applied as in research work so and in industrial installations for technological processes monitoring and control.

REFERENCES

[1] Aleksandrov, A., I. Simeonov. Galvanomagnetic device for measuring of electric current. Journal of the Technical University of Gabrovo, Vol. 15/2000, Gabrovo, 2000.

[2] Goranov, G., A. Iliev. Using of controller with common function in system for control of power. UNITECH`01 conference, Vol. 1, Gabrovo, 2001.

[3] Draganov, N., G. Goranov, A. Aleksandrov. Galvanomagnetic device for contact less measuring of DC current. UNITECH-08 conference, Vol. 1, Gabrovo, 21-22 Nov. 2008, p.142-146.

[4] Draganov, N., G. Goranov, A. Aleksandrov, P. Penchev. Galvanomagnetic device for measuring of magnetic field. Journal of the TechnicalUniversity of Gabrovo, Vol. 37/2009, p.70 - 73.

[5] Smrikarov, A., C. Vasilev. Microcontrollers. Ruse. Avangadr Print House, 2000.

[6] Roumenin, C. S. Solid state magnetic sensors, Amsterdam, Elsevier, 1994.

[7] Data sheet of PIC16F874A. www.microchip.com , April 2010.

[8] Data sheet of ZMY20M. www.zetex.com , April 2010.

TiO_2-Based Humidity Sensors with Different Dopants

Toshko Nenov

Abstract – **The paper presents results from the research of TiO_2-based humidity sensitive elements with different dopants, synthesized at various temperatures. The impact of dopants and sintering temperature on the parameters and characteristics of the sensing elements has been investigated. The results of the investigation of sensitivity, hysteresis of the characteristic and the reaction time are shown.**

Keywords – **Humidity sensors, ceramic, TiO_2, dopant.**

I. Introduction

Ambient air humidity or process gas is one of the important parameters determining the quality of outputs in many technological processes. Therefore, measuring gas humidity is very important and widely used in production and research. This requires development and testing of new and improvement of existing humidity sensors. One of the directions in the development of humidity sensors is the use of sensor elements based on ceramic materials. The use of ceramics for these purposes is determined by its following properties [1, 2]:

- easy adjustment of its microstructure by means of control of composition and conditions of synthesis;
- thermal stability and durability to environmental influences allowing for its use in high temperature processes;
- production using relatively simple operations;
- use of cheap materials in its production, which allows for obtaining relatively inexpensive sensors.

TiO_2 is of particular interest among the great variety of sensing element materials because of the possibility for easier control of the properties of sensing elements through dopants and the sintering temperature as well as its durability. TiO_2 easily forms titanates with dopants, which allows for modifying the parameters and characteristics of the sensing elements. It is relatively abundant and is used for other purposes in the electronics industry. It is also characterized with good durability.

In this respect, the present work describes the study of humidity sensing elements based on TiO_2 with different dopants, sintered at several temperatures.

II. Experiment

Table I contains the compositions of the studied experimental ceramic samples. In the study the dopant used are ZnO, SnO_2, Bi_2O_3, V_2O_5, PbO, CuO and $Na_2CO_3.10H_2O$. Apart from binary some ternarysamples have been investigated as well.

Toshko Nenov is with the Faculty of Electrical Engineering, Technical University of Gabrovo, 4, H. Dimitar Str., Gabrovo 5300, Bulgaria, E-mail: tnenov@tugab.bg

The experimental samples are obtained using ceramic technology [1]. Homogenization is carried out in a planetary mill Pulverizete-5 in corundum pot for 4 hours. Grinding of oxides is carried out simultaneously with the homogenization. Distilled water is used as a medium for mixing and grinding. Granulation gives plasticity to the ceramics and provides better compression of the samples. Polyvinyl alcohol is used as a binder.

The samples are shaped as discs with a diameter of 20mm and thickness of 3mm. The sintering time is 2 hours but the sintering temperature varies for the different samples.

Mechanical processing of the ceramic samples is performed before applying the conductive electrodes after which the samples are cleaned in an ultrasonic bath. Electrodes are obtained after sintering the silver-palladium paste applied on the sample at a temperature of 850°C for 30min. Finally pins are soldered to the resulting electrodes. Figure 1 shows the appearance of investigated experimental samples.

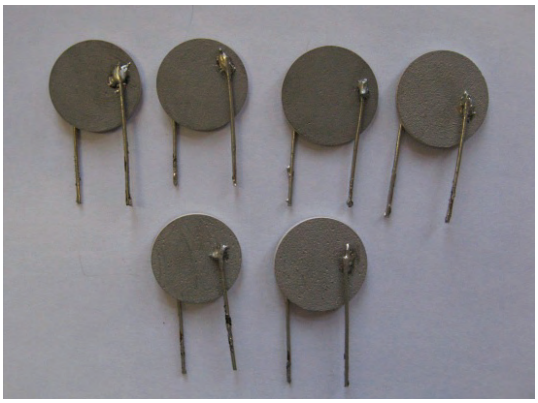


Fig.1. Experimental samples of humidity sensing elements

The humidity sensing characteristics of the samples were determined in conditions of controlled humidity through a relative humidity calibrator VAPORTRON H-100BL, product of BUCK RESEARCH INSTRUMENTS LLC [3]. The device supplies values of relative humidity in the range from 10 to 95%, with accuracyof 1.5%.

The active resistance and the impedance of the sensors, in the conditions of the humidity chamber were measured by an impedance analyzer Precision Impedance Analyzers 6505P product of Wayne Kerr Electronics Ltd. at a frequency of 1kHz and amplitude of 500 mV of the testing signal. The impedance analyzer permits evaluation of various parameters such as, phase angle, capacitance, resistance, inductance, and quality factor with a basic accuracy of 0,05% [4].

TABLE I
COMPOSITIONS OF THE STUDIED EXPERIMENTAL CERAMIC SAMPLES

Sample	Composition	Sintering temperature, °C
T1	95mol%TiO_2+5mol%Bi_2O_3	850, 950, 1050
T2	90mol%TiO_2+5mol %ZnO+5mol %Bi_2O_3	850, 900, 950, 1050
T3	90mol%TiO_2+5mol%Bi_2O_3 +5mol%SnO_2	850, 950, 1050
T4	50mol%TiO_2+49mol%SnO_2+1mol %V_2O_5	850, 950, 1050
T5	50mol%TiO_2+50mol%ZnO	850, 950, 1050
T6	95mol%TiO_2+5mol%$Na_2CO_3.10H_2O$	850, 900, 950, 1050
T7	97,5mol%TiO_2+2,5mol%$Na_2CO_3.10H_2O$	850, 950, 1050
T8	33,3mol%TiO_2+33,3mol%ZnO+33,3mol%SnO_2	850, 950, 1050
T9	95mol%TiO_2+5mol%PbO	850, 950, 1050
T10	95mol%TiO_2+5mol%CuO	850, 900, 950, 1050

III. EXPERIMENTAL RESULTS

Based on the investigations made it is determined that dopants influence the ceramics sintering process. A higher concentration of $Na_2CO_3.10H_2O$ in the T6 sample improves sintering compared to T7. There is better sintering in the samples containing Bi_2O_3, PbO and CuO.

The results from the study of the dependence of resistivity of samples T1, T7, T9 and T10 on the change of relative humidity are shown in Figure2, Figure 3, Figure 4 and Figure 5.

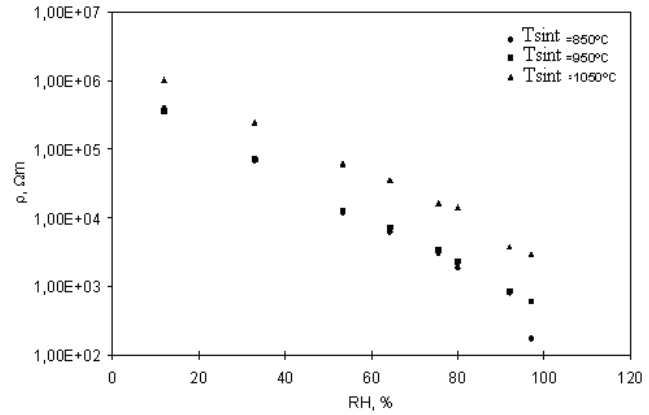


Fig.2. Resistivity dependence on humidity for sample T1

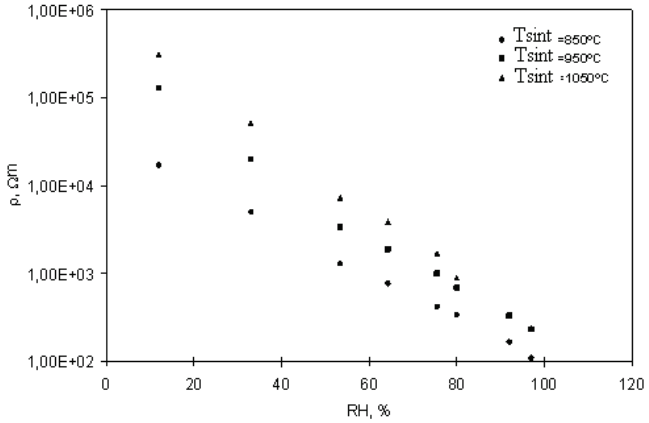


Fig.3. Resistivity dependence on humidity for sample T7

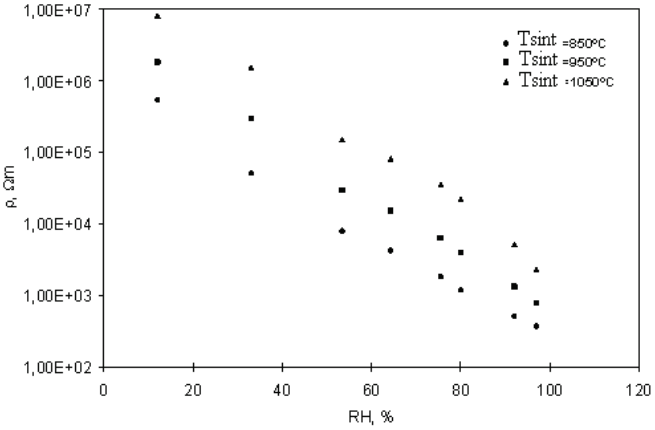


Fig.4. Resistivity dependence on humidity for sample T9

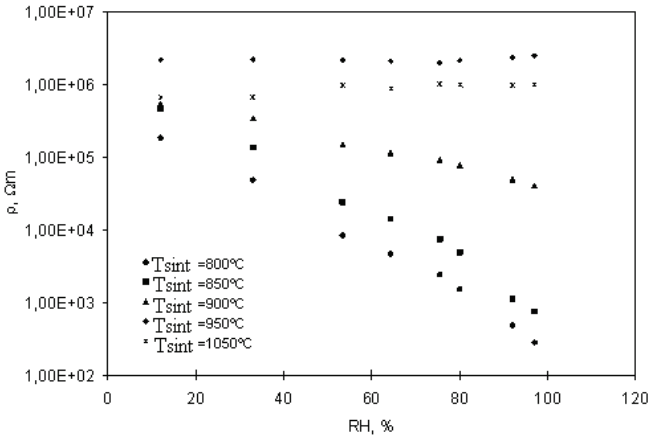


Fig.5. Resistivity dependence on humidity for sample T10

The characteristic function of ceramic humidity sensors with ionic conductivity can be approximated by the following dependence for a wide range of relative humidity:

$$\rho = \rho_0 \exp(-k_H RH) \quad , \qquad (1)$$

where ρ_0 is the sensor resistivity at 0% relative humidity, *RH* is relative humidity in % and k_H is a sensitivity coefficient.

TABLE II
SENSITIVITY COEFFICIENTS FOR THE SAMPLES

Sample	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
k_{Hi}										
k_{H1}, x10^{-3}										
RH_1 = 10%	76	57	72	83	92	61	85	78	92	66
RH_2 = 65%										
k_{H2}, x10^{-3}										
RH_1 = 65%	76	87	82	83	92	101	85	99	92	90
RH_2 = 95%										

TABLE III
HYSTERESIS VALUES FOR THE STUDIED SAMPLES

Sample	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Hysteresis										
F, %	1.3	1.4	1.3	0.6	1.9	11.7	1.6	1.2	1.5	1.7

To quantify the humidity sensitivity of the investigated samples the sensitivity coefficient k_H is used, which is determined by the dependence [1]

$$k_H = \frac{1}{RH_2 - RH_1} \ln\left(\frac{\rho_1}{\rho_2}\right) \quad , \qquad (2)$$

where ρ_1 and ρ_2 are the resistivities of the sensor at relative humidity RH_1 and RH_2 respectively ($RH_2 > RH_1$).

This coefficient allows for a comparative analysis of the characteristics of the studied samples. Since not all of the characteristics of the tested models are linear in a semi-logarithmic scale, they are divided into two linear segments - from 10 to 65% and 65 to 95% RH respectively. The sensitivity coefficients for these samples in both segments of their characteristic are shown in Table II.

All samples shown in Table II, are characterized with high sensitivity to relative humidity. Of particular interest are samples with linear characteristic in semi-logarithmic scale (T1, T4, T5, T7, T9). They have a constant sensitivity coefficient in the whole range of relative humidity.

For almost all sample the sintering temperature affects their resistivity. Only for samples T1 and T6 the sintering temperature in the ranges 850°C to 950°C and 900°C to 950°C, respectively, the sintering temperatures do not affect resistivity. This feature allows for good reproducibility of the characteristics of sensor elements based on T1 and T6.

Increasing the sintering temperature increases the resistivity of the samples. For some of them the characteristic changes from nonlinear to linear or it loses sensitivity to relative humidity. For example, for sample T2 raising the sintering temperature to 1050°C leads to loss of sensitivity. This also applies for sample T10 at sintering temperature above 900°C. For sample T3 increasing the sintering temperature leads to a reduction in the nonlinearity of the characteristic. Increasing the sintering temperature above 1050°C will lead to linearization of the characteristic. On the other hand, lowering the sintering temperature for sample T10 below 800°C has the same effect. Hence, by changing the sintering temperature of the studied samples one can control the type of dependence of resistivity on relative humidity. Additionally, this can be achieved by changing the percentage of dopant within certain bounds.

Hysteresis is found in the characteristics of all studied samples.

Hysteresis of the characteristic $\rho = f(RH)$ is defined as

$$F = \frac{\Delta\rho_{\max}}{\rho_{FS}} 100\ \%, \qquad (3)$$

where $\Delta\rho_{\max}$ is the maximum difference in resistivity of the sensor in the increase and decrease of relative humidity, and ρ_{FS} the range of change of resistivity to changes in relative humidity in the measurement range.

Hysteresis is similarly determined for the characteristics $R = f(RH)$. To study the hysteresis of sensor characteristics the resistance of the samples is first measured when relative humidity increases from 10 to 95%, and after that when it goes in the opposite direction. The hysteresis of the characteristic is determined in accordance with (3).Table III shows hysteresis values for the studied samples.

Figure 6 shows the hysteresis of the characteristic for sample T1 sintered at 850°C and Figure 7 shows the hysteresis of the characteristic for sample T7, also sintered at 850°C.

The change in resistance of the sensors during adsorption and desorption has been investigated as well.

The change in resistance during adsorption and desorption is used to determine the appropriate reaction time. This parameter determines the performance of the sensor, i.e. the time after which the change of humidity can be reliably reported.

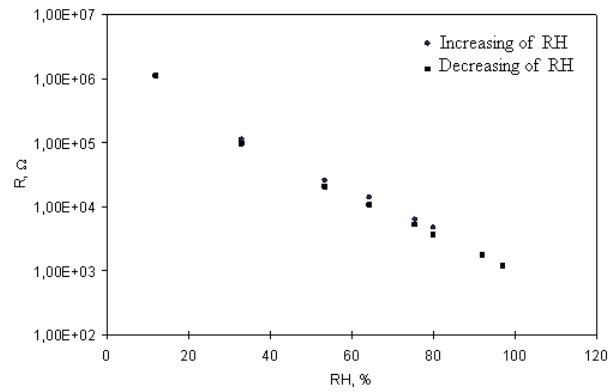


Fig.6. Hysteresis characteristic for sample T1 sintered at 850°C

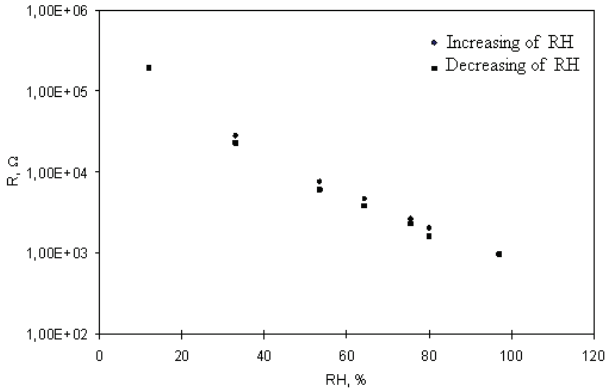


Fig.7. Hysteresis characteristic for sample T7 sintered at 850°C

Studying the change in sensor resistance during adsorption and desorption takes place at base relative humidity of 30% and 95%. Reaction time adsorption and desorption is defined as the time needed to reach $(R_{B1} + 10\% R_{FS})$ and $(R_{B2} - 10\% R_{FS})$, respectively, where R_{B1} and R_{B2} are the resistance values for the corresponding base humidities [5].

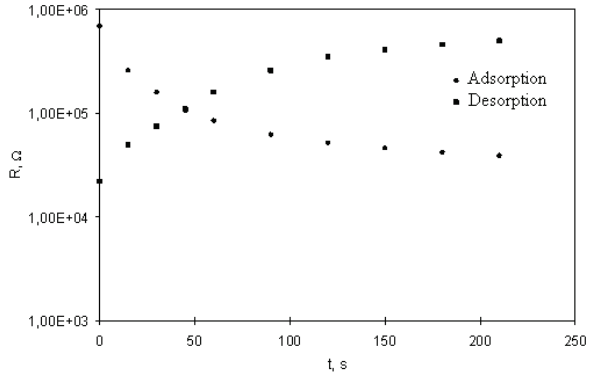


Fig.8. Resistance change for adsorption and desorption of sample T4, sintered at 1050°C

Figure 8 shows the resistance change for sample T4. The reaction time of the investigated sensors for adsorption is in the range 180…300s, and the reaction time for desorption is in the range 240…400s.

IV. Conclusion

Based on the investigations made on the influence of various dopants on the characteristics and parameters of the TiO_2-based ceramic humidity sensing elements the following conclusions can be drawn:

- Bi_2O_3, PbO, CuO and $Na_2CO_3.10H_2O$ have a significant impact on the sintering of the samples. One should expect similar influence of V_2O_5, but owing to its small concentration in the T4 sample, it is not clearly defined;

- The characteristics of the sensing elements with these dopants are linear in semi-logarithmic scale with the exception of samples with CuO;

- Samples with Bi_2O_3, PbO and $Na_2CO_3.10H_2O$ are characterized by high sensitivity for the range of change of relative humidity.

Acknowledgement

The results presented here are in fulfillment of contract № ДО 02-148/2008г. financed by the National Scientific Research Fund (Bulgaria).

References

[1] T. Nenov, S.Yordanov,*Ceramic Sensors: Technology and Applications*,Lancaster-Basel, TECHNOMIC Publ. Co. Inc., 1996.

[2] T. Seiyama, N.Yamazoe, H.Arai,"Ceramic humidity sensors". Sensors and actuators, vol.4, pp.85-96, 1983.

[3] "VAPORTRON H100BL/H100CL Series Precision Humidity Lab", Buck Research Instruments L.L.C., Boulder CO, 2009. (http://hygrometers.com/wp-content/uploads/VAPORTRON-users-manual-2009-12.pdf).

[4] "HF LCR Meter 6505. Technical data sheet", Wayne Kerr Electronics Ltd., 2008. (http://www.waynekerrtest.com/brochure/6500P%20Series.pdf).

[5] G. Asch. *Les capteurs en instrumentation industrielle, Paris*, BORDAS, 1991.

Application of Stress Redundancy and its Influence upon the Reliability of Electronic Elements and Systems

Toncho Papanchev[1], Anton Georgiev[2], Neli Georgieva[3] , Angel Marinov[4]

Abstract – **In this article is discussed the application of stress redundancy to achieve more reliable electronic systems. Here are researched some variations of the failure rate of several typical electronic elements in different stress conditions and ambient temperature. Diagrams of received dependencies are shown. In the paper is analyzed this form of redundancy that causes improving of reliability parameters.**

Keywords **– reliability, failure rate, stress redundancy**

I. INTRODUCTION

The modern way of life is unthinkable without the use of various in type and complexity electronic devices. Their quantity and variety is growing sharply, and their areas of application cover the whole spectrum of human activities. People in ever greater extent rely on electronic systems in ensuring the safety of complex and hazardous industrial processes.

An essential tool to enhance the reliability of electronic components and devices is introduction of so-called "redundancy". According to Gindev [1] redundancy is a measure of exceeding of a value or quantity. Hristov [2] defines redundancy as a generalised notion that covers capable to work devices, more than the necessary to fulfill the functional requirements according to the specifications of the system.

Gindev [1] formulates five types of redundancy: time redundancy, stress redundancy, structure redundancy, function redundancy and information redundancy. Structure and function redundancy is most commonly used as non-operating reserve, while the others are mostly used as an operating reserve.

The stress redundancy is applied mostly to the elements so they are not fully loaded when operate - by mechanical efforts, working power, flowing current, reverse voltage, etc. [2].

The presence of this type of redundancy carries out benefits in terms of reliability in two ways: first, slows up the development of aging processes and postpones entering in the period of parametric failures; and second, allows the elements to withstand higher transitory load, and thus reducing the probability of sudden failures during normal operation. The implementation of stress redundancy can be performed without complication of the scheme solutions or adding additional elements and blocks. Thus it becomes possible to increase reliability without increasing the price of the products.

Stress redundancy causes a better reliability of components and devices. Its application improves a probability of flawless work, a Mean Time To Failure (MTTF) and a failure rate. However, it is not appropriate to talk about fault tolerance – its improvement is achieved by applying other types of redundancy.

This article examines variations of reliability of electronic components and circuits when applying stress redundancy. These results are represented graphically.

[1]Toncho Papanchev is with the Faculty of Electronics, Technical University of Varna, Studentska 1, 9010 Varna, Bulgaria
E-mail: t.papanchev@tu-varna.bg.
[2]Anton Georgiev is with the Faculty of Electronics, Technical University of Varna, Studentska 1, 9010 Varna, Bulgaria
E-mail: georgiev_an@yahoo.com.
[3]Neli Georgieva is with the Faculty of Electronics, Technical University of Varna, Studentska 1, 9010 Varna, Bulgaria
E-mail: neligeorgieva@yahoo.com.
[4]Angel Marinov is with the Faculty of Electronics, Technical University of Varna, Studentska 1, 9010 Varna, Bulgaria
E-mail: igdrazil@abv.bg

II. THEORY

To examine the reliability of electronic devices first have to determine their structure in terms of reliability. There are described different types of structural schemes in the literature [1], [2], [3]. We will focus on the two main structures. If the failure of any component causes the failure of the entire system, it has a serial structure on reliability [3] – Fig. 1.

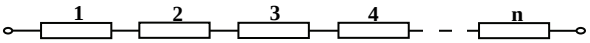


Fig. 1. System of *n* independent elements, connected in series

In contrast, the system which has parallel structure continues to function until fails and its last component [3], [4] – Fig. 2.

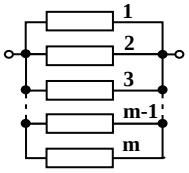


Fig. 2. System of m components, connected in parallel

Systems with serial structure in terms of reliability are the most widespread in practice [5]. To improve their reliability it is used stress redundancy. Therefore, an object of our further research and analysis is the stress redundancy in elements of electronic devices with serial structure.

The mathematical model that describes the reliability of the system with connected in series elements is based on Bernoulli's theorem [3]. According to this theorem, if failures are independent and the flow of failures is fixed in time, ordinary and without subsequent effects (this applies to the majority of electronic devices), the probability of flawless operation $P(t)$ of the system is evaluated by the product of probabilities of flawless operation $P_i(t)$ of its components [3]

$$P(t) = P_1(t) \cdot P_2(t) \cdot P_3(t) \cdot ... \cdot P_n(t) = \prod_{i=1}^{n} P_i(t) . \qquad (1)$$

The life of an electronic component can be divided into three periods during which the intensity of failures is changed in a different manner [1]. These are period of early failures, period of normal operation and aging period. During the first and third periods the failure rate alters significantly and acquires high values. During normal operation, the failure rate remains constant and acquires its smallest values determined only by the occurrence of sudden failures.

The probability of flawless work of electronic components in cases of sudden failures is described by an exponential law

$$P(t) = \exp\left(\int_{t=0}^{\infty} \lambda(t)dt \right), \quad (2)$$

where the parameter $\lambda(t)$ represents the failure rate of electronic components. The failure rate is the number of failures per hour.

When the failure rate is constant $\lambda(t)=\lambda$, the expression (2) acquires the type

$$P(t) = \exp(-\lambda \cdot t) \quad (3)$$

and the MTTF is

$$MTTF = \frac{1}{\lambda}. \quad (4)$$

The MTTF represents the average time of operation before failure.

Replacing (3) in (1) gives a new expression for the probability of flawless work of the system

$$P_{SYS}(t) = \exp(-(\lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_n) \cdot t) = \exp(-\lambda_{SYS} \cdot t) \quad (5)$$

and then the Mean Time Between Failures (MTBF) of the system is

$$MTBF = \frac{1}{\lambda_{SYS}}, \quad (6)$$

where λ_{SYS} is the sum of the failure rates of the system components

$$\lambda_{SYS} = \sum_{i=1}^{n} \lambda_i. \quad (7)$$

In other words, if we determine the failure rate of components in the system we can find the failure rate of the entire system. Then we can make an assessment of the other reliable indexes of the system - probability of flawless work and MTBF.

The purpose of reliability prediction is to be estimated reliability indexes of the elements and devices in order to identify existing weaknesses and to perform adequate actions to achieve the desired level of reliability.

Based on the exponential distribution law in "Military Handbook MIL-HDBK-217F" [6] has proposed the method "Part Stress Analysis" for predicting reliability of electronic components and systems. It takes into account the impact of various influences on the failure rate of electronic components to estimate the part failure rate

$$\lambda_p = \lambda_b \cdot \prod_{i=1}^{n} \pi_i, \quad (8)$$

where λ_p is the part failure rate; λ_b is base failure rate usually expressed by a model relating the influence of electrical and temperature stresses on the part; π_i is the factor of impact of the *i*-th influence.

They are taken into account all important influences of each element. For example, for capacitors they are operating temperature, the value of capacity, the voltage stress, its quality and environment, and for resistors they are working temperature, dissipation rate parameter, the power stress, its quality and environment.

We will consider an element with base failure rate λ_b and *m* numbers influencing factors. By X_i we denote the vector of values of the *i*-th influence, and with Π_i - vector of values of the factor of *i*-th influence. $\varphi_i(x_i)$ is the function describing the impact of the *i*-th influence. We can write

$$X_i = \{x_{i1}, x_{i1}, ..., x_{in(i)}\}, \quad i = \overline{1,m} \quad (9)$$

$$\Pi_i = \{\pi_{i1}, \pi_{i2}, ..., \pi_{in(i)}\}, \quad i = \overline{1,m} \quad (10)$$

$$\pi_{ij} = \varphi_i(x_{ij}), \quad i = \overline{1,m}, \quad j = \overline{1,n(i)} \quad (11)$$

where $n(i)$ is броя на стойностите на *i*-th influence; *m* is a number of influences.

It must be noted that the values x_{ij} of the influence X_i may be of numeric or linguistic type, while their factors of impact π_{ij} accept only numeric type.

Let the first two of influences are modified in their entire range, and the others accept their first values. Then we can write a matrix $\Lambda_1^{(1,2)}$ of the values of the failure rate with dimension $n^{(1)} \text{X } n^{(2)}$ as follows

$$\Lambda_1^{(12)} = \left\| (\lambda_b \cdot \pi_{31} \cdot \pi_{41} \cdot ... \cdot \pi_{m1}) \cdot \Pi_1^{'} \times \Pi_2 \right\|. \quad (12)$$

Performing consecutively multiplying all values of the coefficients of influence, we get *p* numbers two-dimensional matrices describing the values of the failure rate in *m*-dimensional area defined by the factors of influences, where

$$p = \prod_{i=3}^{m} n(i). \quad (13)$$

These results can be used to determine the limits of variation of influences for a given maximum failure rate

$$\lambda(x_{1j(1)}, x_{2j(2)}, x_{3j(3)}, ..., x_{mj(m)}) < \lambda_t, \quad (14)$$

where $\lambda(x_{1j(1)}, x_{2j(2)}, x_{3j(3)}, ..., x_{mj(m)})$ - failure rate in certain combinations of values of the influences X_1, X_2, X_3, …, X_m.

Examining the information received, we can choose the most optimal option for which we have an acceptable value of the failure rate and optimal parameters for practical implementation of the influential factors.

III. Experimental Results

Usually some of the influences are predetermined and can be adopted with specific values. Then the failure rate is considered when changing just one or two impacts.

We apply the proposed by MIL-HDBK-217F [6] methodology in the study of the impact of stress redundancy on the failure rate of several capacitors as shown in Table I.

The resulting failure rate is shown in failures per hour if not mentioned specially. The MTTF and MTBF are expressed in hours or years.

In practice there is previously known the quality of the elements, the environment in which they work, and the values of their capacity. This allows us to fix the value of the coefficients of these factors as listed in Table I.

TABLE I
EXAMINED ELEMENTS

	Type and capacity	λ_B, failures in million hours	π_Q	π_E	π_C	k, variant of Π_t	l, variant of Π_V
λ_{Al}	**Electrolytic Al_2O_3, 220 µF**	0.00012	3	10	3.46	2	1
λ_C	**Ceramic, 100 nF**	0.00099	3	10	0.81	2	3
λ_T	**Electrolytic Tantalum, 10 µF**	0.00040	3	10	1.7	1	4
λ_{CC}	**Ceramic multilayer, Chip, 100 nF**	0.0020	3	10	0.81	2	3
λ_{TC}	**Electrolytic Tantalum, Chip, 4.7 µF**	0.00005	3	10	1.43	1	4

The level of stress in capacitors is determined by the ratio between the maximum values of the applied operating voltage to the electrodes and the nominal voltage of capacitors. Range of variation of the stress is from 0.1 to 1.0 and the reporting values are set with the vector $X_V=\{x_{V1},x_{V2},\ldots,x_{Vi},\ldots x_{Vn}\}$ and

$$x_{Vi}=0.1+(i-1)\cdot\frac{1-0.1}{n-1},\ i=\overline{1,n} \tag{15}$$

where n - number of reported values of the stress.

The influence of the working temperature on the reliability parameters is investigated in the range -20°C ÷ 150°C divided into equal intervals. We set the values of the temperature with vector $X_t=\{x_{t1},x_{t2},\ldots,x_{tj}\ldots,x_{tn}\}$ and

$$x_{tj}=20+(j-1)\cdot\frac{150-20}{m-1},\ i=\overline{1,m} \tag{16}$$

where m - number of reported values of the temperature.

For further analysis we assume $n = 10$ and $m = 15$.

The factor of impact of temperature denotes with vector $\Pi_t^{(k)}=\{\pi_{t1}^{(k)},\pi_{t2}^{(k)},\ldots,\pi_{tj}^{(k)},\ldots,\pi_{tm}^{(k)}\}$, where k is the version of the factor depending on the capacitor's type [6]

$$\pi_{tj}^{(k)}=\exp\left[\frac{-E_a^{(k)}}{C}\cdot\left(\frac{1}{x_{tj}+273}-\frac{1}{298}\right)\right],\ j=\overline{1,m},\ k=1,2 \tag{17}$$

where: $E_a^{(k)}=\{0.15;0.35\}$ and $C=8.617\cdot10^{-5}$.

The factor of impact of voltage stress denotes with vector $\Pi_V^{(l)}=\{\pi_{V1}^{(l)},\pi_{V2}^{(l)},\ldots,\pi_{Vj}^{(l)},\ldots,\pi_{Vm}^{(l)}\}$, and l is the version of the factor depending on the capacitor's type [6]

$$\pi_{Vi}^{(l)}=\left\{\left(\frac{x_{Vi}}{0.6}\right)^{R_l}+1\right\},\ i=\overline{1,n},\ l=\overline{1,4},\ R_l=\{5;10;3;17\} \tag{18}$$

$$\pi_{Vi}^{(l)}=\left\{\left(\frac{x_{Vi}}{0.5}\right)^{3}+1\right\},\ j=\overline{1,m},\ l=5\ . \tag{19}$$

On Fig. 3 are shown the two variants of temperature factors. There is clearly seen that capacitors, for which is valid variant 2 are much more sensitive to increased operating temperature.

Fig. 4 presents the options of modifying the stress factor. There is a big difference between minimum and maximum values - from 0.1 to 6000. We apply logarithmic scale to get better informative graphic displaying. It clearly differentiates the limit value of the voltage stress by 0.6, to which the impact of stress on reliability is relatively low and approximately the same for all types of capacitors. Above 0.6 there are significant differences in values of the various stress factors. For example, under stress $x^{(V)} = 1$ we calculate values $\pi^{(V)3}=5.6$ for ceramic capacitors and $\pi^{(V)4}=5900$ for tantalum capacitors.

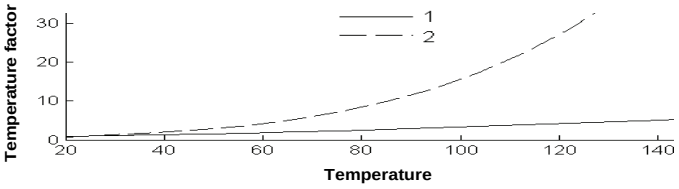


Fig. 3. Variants of temperature factor

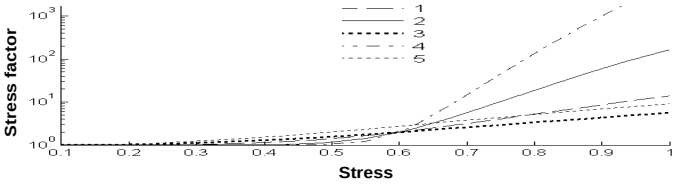


Fig. 4. View of the fifth variants of stress factor

Let us consider in detail the results for the electrolytic capacitor with aluminum oxide.

We calculate the matrix of the failure rate as a function of temperature and voltage stress factors

$$\Lambda_{Al}=\left\|(\lambda_b^{Al}\cdot\pi_Q^{Al}\cdot\pi_E^{Al}\cdot\pi_C^{Al})\cdot\left(\Pi_t^{(2)}\right)'\cdot\Pi_V^{(1)}\right\|. \tag{20}$$

If we use resulting data we can show the failure rate in various combinations of stress and operating temperature - Fig. 5.

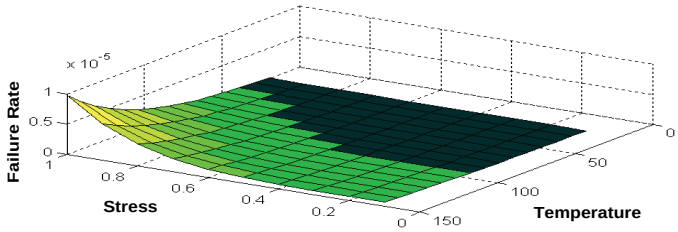


Fig. 5. Failure rate vs. stress and temperature

We can see the range of values of temperature and stress in which the failure rate changes in small range, followed by a significant and steep rise. Special attention should be paid at the level of stress when we expect high operating temperatures - over 60°C.

It is easy to identify eligible areas of change in operating temperature and stress on the placed requirements for the failure rate - Fig. 6.

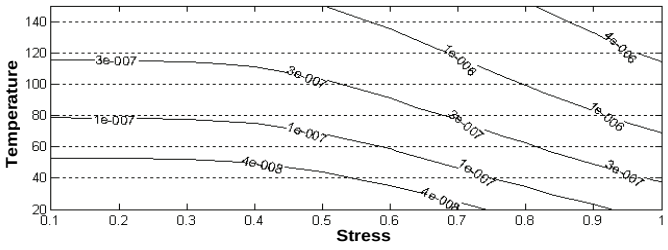


Fig. 6. Nomogram of failure rates

It is seen that when the working temperature is 60°C for example, the change of stress from 0.8 to 0.5 leads to three times decreasing of the failure rate.

By the same way we obtain data on other types of capacitors from Table I. On Fig. 7 there is shown the failure rate of examined elements as a function of voltage stress at operating temperature 50°C.

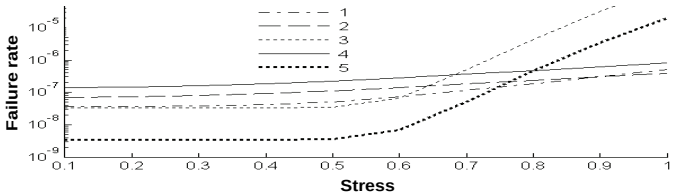


Fig. 7. An influence of stress levels on failure rate of capacitors
1 - Aluminum electrolytic capacitor; 2 - Ceramic multilayer capacitor; 3 - Tantalum electrolytic capacitor; 4 - Ceramic chip capacitor; 5 - Tantalum electrolytic chip capacitor

For example, let us have a system with 25 capacitors in it, equally distributed among the types counted in Table 2. Assumed operating temperature is 50°C. We calculate the failure rate of the system using Eq. (7)

$$\lambda_{SYS} = \lambda_R + 5\cdot\lambda_{Al} + 5\cdot\lambda_C + 5\cdot\lambda_T + 5\cdot\lambda_{CC} + 5\cdot\lambda_{TC}\,, \qquad (21)$$

where λ_R is the aggregated failure rate of the other elements of the system. Assuming $\lambda_R = 10.10^{-6}$ failures in hour. The results are shown in Fig. 8.

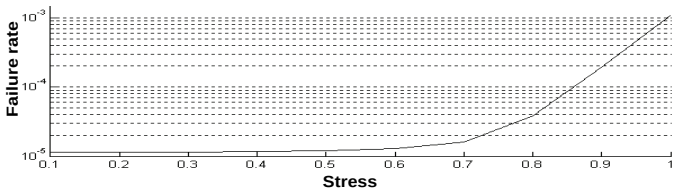


Fig. 8. An influence of different stress levels on the system failure rate

If we use Eq. 4 we could calculate MTBF depending on voltage stress. The results are shown in Fig. 9.

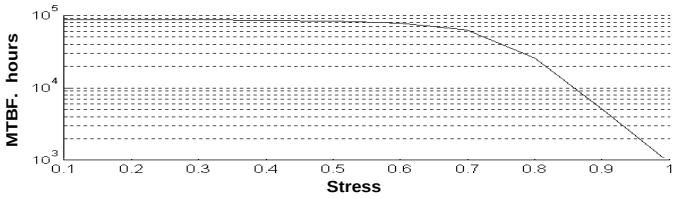


Fig. 9. An influence of stress levels on the system MTBF

Using Equation 5 we calculate the probability of flawless work of the device from the beginning of its operation up to a year - 8760 hours. The results are shown in Fig. 10.

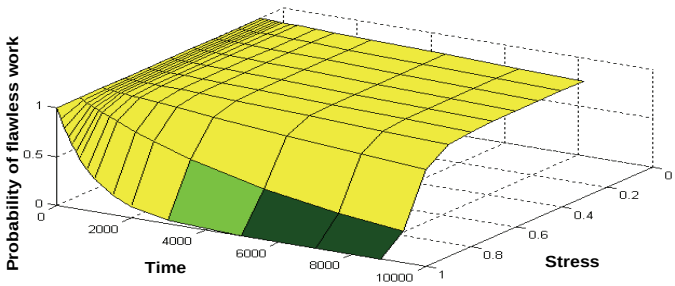


Fig. 10. Variation of system probability of flawless work

Table II shows several values of the probability of flawless work $P_{SYS}(t)$ at the end of the first year of operation under different stress conditions.

There is a relatively smooth reduction in the probability of flawless work in time when the stress is up to 0.7 and a sharp drop for higher values.

TABLE II
VALUES OF $P_{SYS}(t)$ AT THE END OF THE FIRST YEAR

STRESS	0.6	0.7	0.8	0.9
$P_{SYS}(t)$	0.89	0.87	0.71	0.18

In Table III we express the values of the MTBF according to several stress levels.

TABLE III
VALUES OF *MTBF* ACCORDING TO STRESS

STRESS	0.6	0.7	0.8	0.9
MTBF, years	8.89	7.15	2.95	0.59

It is clearly seen that when the voltage stress on the capacitors is low – 0.6 or less, we may expect the first failure of the device after nearly 9 years. Otherwise, if the voltage stress is 0.9 the expected time for the first failure is just 7 months.

IV. CONCLUSION

In this paper the application of stress redundancy to achieve more reliable electronic elements and systems is considered. Research is done and results are shown in improving reliability of electronic elements and systems by implementing a stress redundancy.

Stress redundancy leads to a better reliability of components and devices. Its implementation increases the probability of flawless work, the MTBF and reduces the failure rate. It is important to be applied in case where a relatively high working teperature is expected.

The proposed approach is applicable to different electronic components and devices and allows quick view of the possibility of improving the reliability by increasing the stress redundancy.

ACKNOWLEDGEMENT

The scientific researches, which results are presented in this publication, are made on the project “НП 3/2011” within the inherent by Technical University of Varna scientific-research activity, funded by the government budget.

REFERENCES

[1] Е. Гиндев, “ Увод в теорията и практиката на надеждността. Част 1. Основи на приложната надеждност”, Академично издателство “Проф. Марин Дринов”, София, България 2000.

[2] Х. Христов, В. Трифонов, “Надеждност и сигурност в комуникациите”, Нови знания, София, България 2005.

[3] А. Георгиев, Р. Пранчов, “Конструиране и технология на електронна апаратура”, ТУ–Варна, Варна, България 2003.

[4] D. P. Durkee, E. A. Pohl and E. F. Mykytka, “Input data characterization factors for complex systems affecting availability estimation accuracy”, Annual “Reliability and maintainability” 2002 Proceedings the International Symposium on Product Quality & Integrity, Seattle, Washington USA, pp.80-88.

[5] B. G. Najdenov, “Investigation into Non-Recursive Digital Filter in Correlated Jamming Environment”, Conference Proceedings, ICEST 2008, Nis, Serbia, June 25-27, 2008, pp. 518-521.

[6] “Military Handbook Reliability Prediction of Electronic Equipment MIL-HDBK-217F, Notice 2”, US Department of Defence, USA 1995.

FPAA Implementation of Phase-Independent Synchronous Detector for Spectrum Analyzer

Eltimir Stoimenov[1], Georgi Mihov[2], Ivailo Pandiev[3]

Abstract – **This paper presents a phase independent synchronous detector based on Field Programmable Analog Array (FPAA) devices. The proposed synchronous detector consists of two one-quadrant analog multipliers and two first-order low-pass filters (LPFs). The 'modus operandi' could be explained with the multiplication of the input signal with two additional reference sine and cosine signals with varying frequency. This way the resulting DC offset is proportional to the magnitude of the spectral component with frequency equals the frequency of the referent signal. The circuit is phase independent due to the output block for vector sum. The output voltage corresponds to the half amplitude of the specific spectral component. The functional elements of the structure are realized by employing configurable analog modules (CAMs) of the FPAA AN231E04 from Anadigm. The detector has relatively wide-band frequency response and can operate with 3,3V single supply voltage. The simulation results show good agreement with theoretical analysis.**

Keywords – **Analog signal processing, Spectrum analysis, Synchronous detector, Circuit for modulus calculation, FPAA.**

I. INTRODUCTION

The spectral analysis is quite a common task in nowadays electronics. Almost all devices doing spectral analyses rely on digital signal processing and more specifically - Fast Fourier Transform (FFT) [3]. Although digital signal processing possess many advantages like high noise immunity, high accuracy and great flexibility, depending from the application, this method could possess certain drawbacks. The most obvious of them is the need of sophisticated and relatively expensive microprocessor system. This could lead to increased overall power consumption and greater PCB requirements.

In the current paper we have recourse to a somewhat outdated method for spectrum analysis. It is based on purely analog circuits for signal processing and is already known in the literature [1],[2]. This method is implemented by a specific class of electronic circuits called "Synchronous detectors (or synchronous demodulators)". This kind of circuits is widely used in balanced modulation and demodulation, lock-in amplification, phase detection, and square wave multiplication [4]. The synchronous detectors are circuits which are highly selective and can obtain the amplitude of a spectral component with certain frequency of a noisy signal. The detector is controlled by an external signal with sinusoidal form. A detailed mathematical description of circuit operation is given in a section entitled "Principle of operation".

In addition, we have to note that the circuit described in this paper is planned to work in co-partnership with a specific digital control system introduced in [5]. In that way a Collaborative Analog and Digital Signal Processing (CADSP) system is realized.

II. PRINCIPLE OF OPERATION

The principle of operation of the phase-independent synchronous demodulation is based on multiplication of the input analog signal with sinusoidal and cosinusoidal signals with a certain frequency f_{CS}. The separation of the DC voltage proportional to the amplitude and the phase of the harmonic with frequency f_{CS} is performed by a low pass filter. The output voltage of the phase-independent synchronous detectors is obtained by vector sum circuit. As a result the output DC voltage is proportional only of the amplitude of the input signal with frequency f_{CS}.

The block diagram of the phase-independent synchronous detector implemented in our research is shown on Fig. 1. It's consisted by two one-quadrant analog multipliers, two LPFs and one output block for modulus calculation. The input x of the multipliers is connected to the input signal, while the y input is applied with sinusoidal (sin) and cosinusoidal (cosine) signals u_1 and u_2 from a quadrature generator. These signals are synchronizing and can be described by the fallowing equations:

$$u_1 = E\sin(\omega_{GEN}t) \quad u_2 = E\cos(\omega_{GEN}t) \tag{1}$$

Here E is the amplitude of the signals and ω_{GEN} is its frequency.

We will assume that the input signal u_i is a sine waveform with equation:

[1]Eltimir Stoimenov is with the Faculty of Electronics, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: e_stoimenov@tu-sofia.bg.
[2]Georgi Mihov is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: gsm@tu-sofia.bg
[3]Ivailo Pandiev is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ipandiev@tu-sofia.bg.

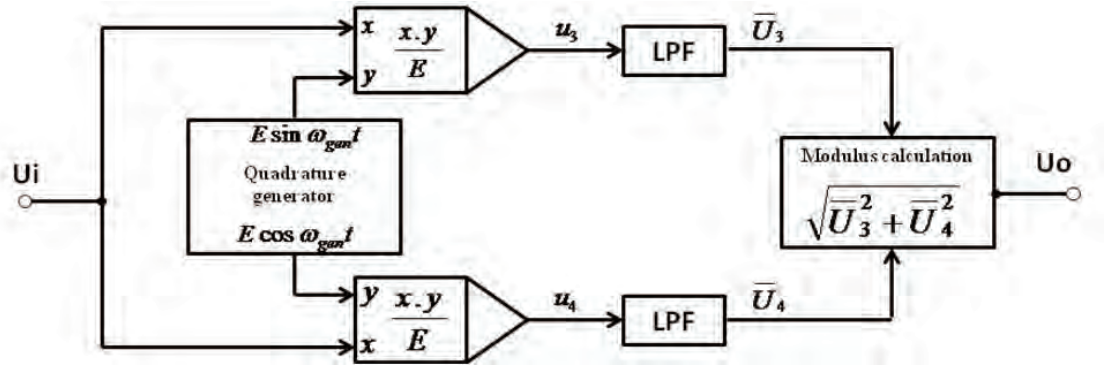


Fig. 1. Phase-independent synchronous detector block diagram.

$$u_i = U_{im}\sin(\omega_{in}t + \varphi) \tag{2}$$

where φ is the phase and u_{im} is the amplitude of the signal.

After multiplication the resulting signals from the analog multipliers can be described as:

$$u_3 = \frac{1}{K}[U_{im}\sin(\omega_{in}t + \varphi).E\sin(\omega_{GEN}t)] \tag{3}$$

$$u_4 = \frac{1}{K}[U_{im}\sin(\omega_{in}t + \varphi).E\cos(\omega_{GEN}t)] \tag{4}$$

where $\frac{1}{K}$ is a multiplication factor. Assuming that $K = E$ and, we can simplify the above equations to:

$$u_3 = U_{im}\sin(\omega_{in}t + \varphi).\sin(\omega_{GEN}t) \tag{5}$$

$$u_4 = U_{im}\sin(\omega_{in}t + \varphi).\cos(\omega_{GEN}t) \tag{6}$$

Using the trigonometric relations:

$$\sin\alpha.\cos\beta = \frac{1}{2}[\cos(\alpha+\beta) - \cos(\alpha-\beta)] \tag{7}$$

$$\sin\alpha.\sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)] \tag{8}$$

we can further evolve the above equations:

$$U_3 = U_{im}\sin(\omega_{in}t + \varphi).\sin(\omega_{GEN}t) =$$
$$= \frac{1}{2}U_{im}[\cos(\omega_{in}t + \varphi - \omega_{GEN}t) - \cos(\omega_{in}t + \varphi + \omega_{GEN}t)] \tag{9}$$

If we assume that the input frequency is equal to the frequency of signal from the DDS generator, or $\omega_{GEN} = \omega_{in}$ we have:

$$U_3 = \frac{1}{2}U_{im}\cos(\varphi) - \frac{1}{2}U_{im}\cos(2\omega_{GEN}t + \varphi) \tag{10}$$

Similarly for U_4 we get:

$$U_4 = \frac{1}{2}U_{im}\sin(\varphi) - \frac{1}{2}U_{im}\sin(2\omega_{GEN}t + \varphi) \tag{11}$$

If we analyze equation (9) we can infer that output signal from the multiplier is a sum of two sinusoids, one with low frequency and the other with high frequency. But when $\omega_{GEN} = \omega_{in}$ the first term of the signal is a DC and the second one is a sinusoidal component with doubled frequency. When we apply this signal to the input of LPFs with very low corner frequency only the DC component will pass. Consequently the output signals from the LPFs will look like equations (11) and (12).

$$\overline{U_3} = U_{im}\cos(\varphi) \tag{11}$$

$$\overline{U_4} = U_{im}\sin(\varphi) \tag{12}$$

We can generalize:

$$\overline{U_3} = \begin{cases} \frac{1}{2}U_{im}\cos(\varphi), \text{ when } \omega_{GEN} = \omega_{in} \\ 0 \qquad\qquad , \text{when } \omega_{GEN} \neq \omega_{in} \end{cases} \tag{13}$$

$$\overline{U_4} = \begin{cases} \frac{1}{2}U_{im}\sin(\varphi), \text{ when } \omega_{GEN} = \omega_{in} \\ 0 \qquad\qquad , \text{when } \omega_{GEN} \neq \omega_{in} \end{cases} \tag{14}$$

As it's obvious U_3 and U_4 are proportional to the amplitude U_{im} and the phase φ of the input signal.

The last block in the system performs modulus calculation of $\overline{U_3}$ and $\overline{U_4}$. Its output signal $\overline{U_O}$ is:

$$\overline{U_0} = \frac{1}{2}U_{im}\sqrt{\sin^2(\varphi) + \cos^2(\varphi)} = \frac{1}{2}U_{im} \tag{15}$$

As we can see, thanks to that block the output signal is only proportional to U_{im} and is phase independent to the input phase φ.

III. System Realization

The realization of the system described above could be quite challenging as it contains relatively complex and hard to attune analog blocks. Fortunately, recently the market introduced somewhat new class of devices called Field Programmable Analog Arrays (FPAA). These devices allow easy implementation of rather complex analog transfer functions. This could reduce the development time significantly and consequently, made FPAA our choice for realization.

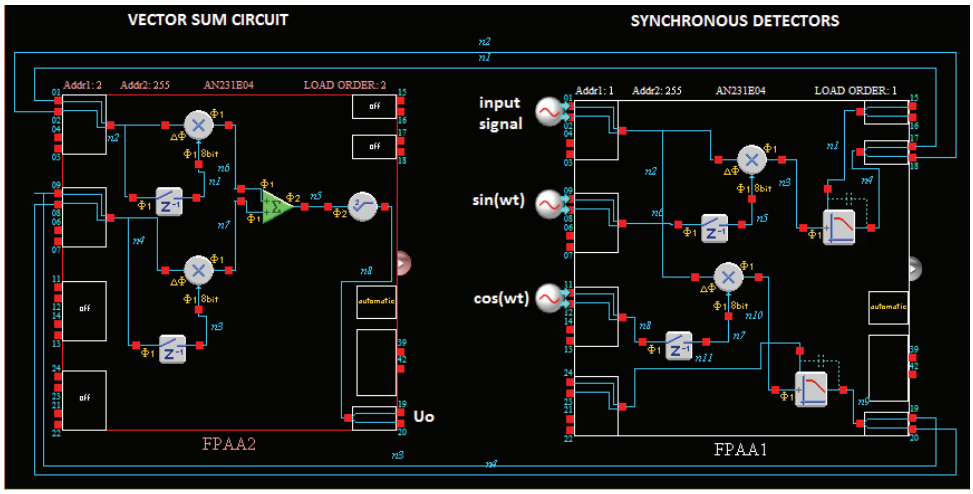


Fig. 2. FPAA configuration of synchronous detector from Fig. 1.

Any FPAA device is consisted by specific number of the so called Configurable Analog Blocks (CAB). These blocks implement Switching Capacitance (SC) technology and could be configure very flexibly. The CABs are surrounded by programmable interconnect resources for signal tracing from one CAB to other. Also FPAA incorporates analog input / output cells which can be set up in variety of operating modes. Using the software development environment the user can built the desired circuit with ease using numerous of predefined analog blocks. These blocks include amplifiers, sumators, integrators and many more.

The main drawbacks of FPAA technology are the reduced bandwidth and the increased power consumption compared to their application specific counterparts. Nevertheless we have to note that these disadvantages are not important to our system, as it haven't been planned to work with high frequency or autonomous supply. More detailed information about FPAA could be found in [6],[7].

As to the authors knowledge there is only one FPAA manufacturer on the market - Anadigm Inc. The company offers a wide range of silicon devices and software development tools. We have decided to use AN231K04-DVLP3 AnadigmApex Development Board as a hardware platform.

For realization of the system we've used Anadigm Designer 2 IDE. The screenshot shown on Fig. 2 presents our concept. The system is built with two FPAAs so two development boards should be daisy chained. Despite that intuitively the signal flow should be from the left Integrated Circuit (IC) to the right, the ICs are placed reversed (the input signals are applied to the right IC). The reason for this is technical and is due to the fact that the development boards could be daisy chained only from right to left (reverse to the signal flow). The right IC consist the two synchronous detectors and solves Eq. (2), (3) and (4). As the LPF should pass only the DC component, the corner frequency of the filter should be as low as possible. Therefore we used LPF with external capacitor which can achieve corner frequency as low as 10-2 Hz. The FPAA input analog signals are referenced to $VMR = +1{,}5V$ (Voltage Mid-Rail − VMR) and are limited to the range of 0 to $+3{,}3V$.

As already mentioned the multiplication factor K should be equal to the reciprocal of the amplitude E of the sine and cosine signals. So if we use signals with amplitude $E = 1{,}25V$ we can determine that $K = 0{,}8V^{-1}$.

The circuit implemented in the left FPAA performs the modulus calculation and solves Eq. (5). It is consisted by two multipliers operating as squarers, summator and square root extracting circuit.

The maximum operating frequency (corner frequency) of the detector is approximately 50 kHz.

IV. Simulation Results

The workability of the proposed synchronous detector shown on Fig. 2 is presented through simulation results. All the simulations were conducted using AnadigmDesigner2 IDE.

All the three simulations aim to show the ability of the system to obtain the spectrum of different kind of input signals. The simulation set up parameters are as follows: amplitude of the input signal (U_{im}) - 0,8 V; amplitude of the quadrature generator signals (E) - 1,25 V; multiplier multiplication factor (K) – 0,8 V^{-1}; all input signals DC offset - 1,5 V; Simulation time – 5ms.

We have to note that for simulation purposes the quadrature generator is replaced with two sinusoidal generators. The first of them forms the sine signal, while the other is set up with 90° phase shift so a cosinusoidal signal can be formed.

The simulation procedure is as follows: firstly an input signal with desired parameters is applied to the system. Then the frequency of the quadrature generator is incremented with some step and the respective output voltage is recorded. The iteration is repeated till the desired frequency band is achieved.

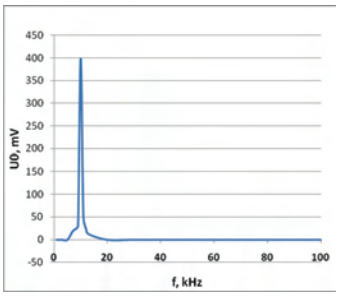
The simulation results are shown in Fig. 3. In the first case, shown on Fig. 3a, the input signal is a pure sinewave. Theoretically the spectrum of such a signal should look like a sharp spike. The maximum of this spike must be reached when x-axis become equivalent to the frequency of the input

signal. Also its amplitude should be $\frac{1}{2}U_{im}$ (400 mV in the current case).

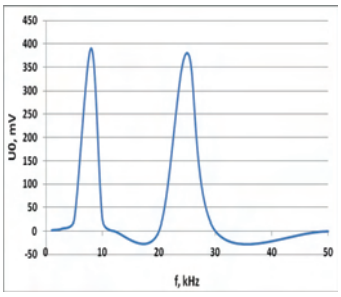
The second study is conducted with an input signal constituted by two sinewaves. The one sinewave is with frequency $\omega_{1in} = 8kHz$ and the other is $\omega_{2in} = 25kHz$. The amplitude should be still about 400 mV (Fig. 3b).

The third case, shown on Fig. 3c, examines squarewave input signal with frequency $\omega_{in} = 10kHz$. As it is well known from the theory the spectrum of such a signal is the *sinc* function with extremums multiple of the input frequency.

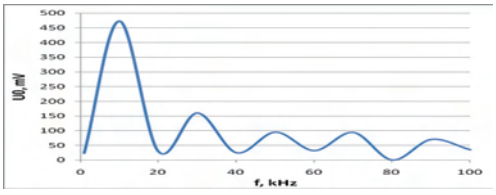
The last study is conducted with Electrocardiographic Signal (ECG) shown on fig. 3d. The analyzed signal is shown on fig. 3f. The spectrogram shown on fig. 3e is given with comparison purposes and it's obtained with specialized software "FFT Properties v5 trial" from Dew Research Inc[8]. Unfortunately there is poor agreement between the two spectrograms. Nevertheless we have to note that the spectrogram is quite complex and probably more simulation points will improve the situation (currently the spectrogram is built with 20 points). Also we have to mention that for simulation purposes the frequency of the ECG signal was increased 1000 times. The reason for this frequency increasing is that under simulation Anadigm Designer 2 doesn't allow the usage of LPF with external capacitors. As already mentioned, for simulation purposes we have been forced to use an ordinary LPF which have pretty higher corner frequency.



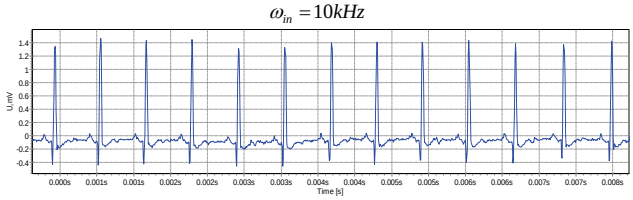
a) Pure sinewave with frequency $\omega_{in} = 10kHz$



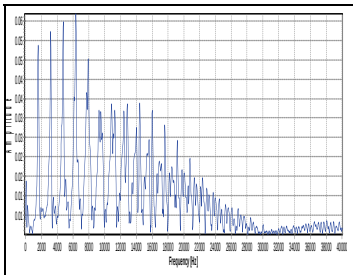
b) Sum of two sinewaves with frequencies $\omega_{1in} = 8kHz$ and $\omega_{2in} = 25kHz$



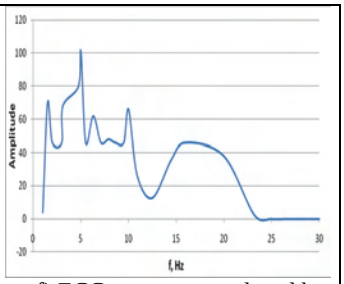
c) Square wave signal with frequency $\omega_{in} = 10kHz$



d) ECG Input Signal



e) ECG spectrum produced with FFT properties trial [8]



f) ECG spectrum produced by SD simulated with Anadigm Designer 2

Fig. 3. Simulation results with different input periodic signals.

V. Conclusion

In this paper a SD independent to the signal phase has been proposed. The SD circuit consists two one-quadrant analog multipliers, two first-order low-pass filters with external capacitors and output block for modulus calculation. The circuit is synchronized by two control signals with sine and cosine form. If the frequency of these signals varies linearly the SD could be used as a spectrum analyzer. A detailed mathematical description of the circuit operation was given. Also for confirmation of the theoretical predictions various simulations has been conducted.

Acknowledgement

This paper is a part of a project with contract № 102pd209-3/2010, which is sponsored by the research program of the Technical University of Sofia, Bulgaria.

References

[1] V. Tietze, Ch. Schenk, Halbleiter-Schaltungstechnik. 12. Aufla-ge. Berlin, Heidelberg, New York. Springer-Verlag, 2002.

[2] D. Sheingold, Справочник по нелинейным схемам Издателство „Мир", 1977.

[3] M. Hunter, A. Kourtellis, C. Ziomek and W. Mikhael, Fundamentals of Modern Spectral Analysis, IEEE, 2010.

[4] AD630: Balanced Modulator/Demodulator – datasheet. Analog Dev., Norwood, MA, 2004.

[5] E. Stoimenov, Design of Digital Control System of Spectrum Analyzer Built on MicroBlaze™ Processor, ICEST, 2011 (in press).

[6] http://en.wikipedia.org/wiki/Field-programmable_analog_array, last accessed April 18, 2011.

[7] http://anadigm.com/, last accessed April 18, 2011.

[8] http://www.dewresearch.com/fftp-main.html, last accessed April, 27, 2011

Design of Digital Control System of Spectrum Analyzer Built on MicroBlaze™ Processor

Eltimir Stoimenov[1]

Abstract –**The modern spectrum analyzers rely on purely digital signal processing methods such as Fast Fourier Transform (FFT). In some application multiresolution methods such as Wavelet are also applied. Regardless of the specific implementation the methods discussed above relies on digital system data processing. We are proposing a different approach for obtaining the spectrum of an input signal which relies on Collaborative Analog and Digital Signal Processing (CADSP). The current paper presents the design phase of a digital control system for a spectrum analyzer exploiting CADSP. The system is planned to be realized with Spartan3E Starter Kit equipped with onboard Field Programmable Gate Array (FPGA). It includes quadrature Direct Digital Synthesis (DDS) generating block, control unit, memory and some minor logical blocks ,all of them implemented in the FPGA. The control unit is based on Microblaze software processor with some additional periphery and it is in charge of managing all the system operations. These operations include - transferring the data from the DDS block to the onboard digital-to-analog converter (DAC) and servicing the user interface. The analog circuitry of the spectrum analyzer is realized with Field Programmable Analog Array (FPAA) and it is discussed in details in [1]. The proposed controlling system for spectrum analyzer is in a process of realization and experimental testing.**

Keywords – **Spectrum analysis, DDS function generators, Microblaze, CADSP, FPGA.**

I. Introduction

Spectrum analysis of an input signal is a common task in various fields of electronics. The modern approaches rely almost fully on digital signal processing and FFT in particular. This paper and [1] presents a different approach for spectrum analyzing based on mixed (analog and digital) signal processing. Using such an approach could poses some advantages like reduced power consumption, reduced size on chip and even higher computing performance, compared to its purely digital counterparts. In the heart of the spectrum analyzer mentioned above is the so called Synchronous Detector (SD). This purely analog circuit performs all the complex computing operations. Thanks to that circuit the harmonic composition of the input signal could be obtained. Detailed information about this kind of circuits could be found in[1], [2] and [3]. Nevertheless we should note that for proper operation of the analyzer the SD needs some additional circuitry. This includes circuitry for data collection and, more important, specific quadrature generator. The generator should be able to produce sine and cosine signals with varying frequency and. Also they should be with high signal to noise ratio (SNR) and swing not exceeding 3,3 V. The easiest way for generating such signals is the usage of DDS which can satisfy all these requrements. Detailed information about DDS can be found in [4].

This paper introduces a design of the system that will provide all the necessary functionalities needed by the SD for proper operation.

[1]Eltimir Stoimenov is with the Faculty of Electronics, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: e_stoimenov@tu-sofia.bg.

II. Block Diagram

The first stage in designing the system described above is the synthesis of the block diagram. The blocks needed for realization are: quadrature DDS generating block, controlling block (processor) and user interface.

A. Processor block

The processor block is responsible for the entire control tasks in the system. The processor should have sufficient memory and peripheral recourses. Its performance isn't critical as all the complex signal processing tasks are carried out in the analog part as described in section I.

B. Quadrature DDS generator block

The output signals from the DDS block are vital for proper operation of the SD. The required signal parameters are: simultaneous sine and cosine outputs, varying frequency up to several kHz, pick to pick swing limited to 0 - 3,3 V, signal to noise ratio (SNR) and frequency tuning resolution should be as high as possible.

C. User interface

The user interface should contain standard devices like LCD, push-buttons and RS232 interface. This block allows the user to communicate with the system and set all the parameters of interest.

D. Interface to the SD

As our system relies on the collaboration between analog and digital signal processing the problem for proper data exchange between the two parts arises. The natural solution of this problem is the usage of DAC and ADC. The DAC should convert the signals from the DDS block in appropriate analog form. On the other hand the ADC allows data receiving from the analog segment of the system.

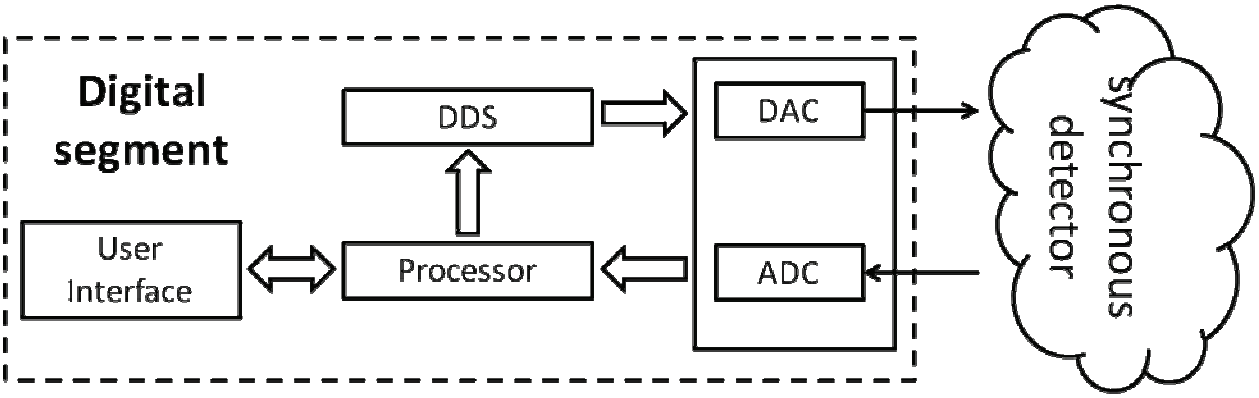


Fig. 1. Spectrum analyzer digital segment block diagram.

III. CONTROL SYSTEM IMPLEMENTATION

For implementation we've decided to use Spartan-3e Starter Kit from Xilinx Inc. This kit is very well known among the electronics engineering community and a plenty of information about it is available in internet. Moreover it possesses all the necessary blocks onboard. The kit is equipped with Spartan3E FPGA which will be extensively used in the system implementation. For more detailed information about the kit you can address to [7].

For software design environments Xilinx ISE and Xilinx EDK tools will be used.

If we refer to fig. 1 we can note that all the blocks, except ADC and DAC, could be implemented in the onboard FPGA. A block diagram with the components implemented in the FPGA is given on fig. 2. Also a detailed description will be given below.

Also now is the right moment to note that the starter kit has onboard devices like LCD, LEDs, pushbuttons and rotary encoder which are very convenient for user interface realization.

A. GPIO and Rotary switch encoder blocks

These blocks are part for the user interface which serves as a connection to the user. The GPIO block connects the system to the onboard pushbuttons, LCD and LED indicator. The Rotary switch encoder block filters the glitches from the externally connected rotary switch and that way unloads the processor from this task.

B. DDS Generating block

This block has a major role in the system and must provide the appropriate signals, needed by the SD circuit. The block is synthesized using Xilinx ISE prewritten IP core. The following key parameters were set:

- Type of output signals : Sine and Cosine
- Frequency tuning resolution : < 2Hz
- Output data width : 14 bits

C. Processor block

For system control our development exploits MicroBlaze software processor. The MicroBlaze core is a 32-bit RISC Harvard architecture soft processor with a rich instruction set optimized for embedded applications. The MicroBlaze processor, allows complete flexibility to select the combination of peripheral, memory and interface features that will give the exact system needed [5]. According our needs we've synthesized the processor putting up the following blocks and parameters:

- 32kB of Data and Instruction memory. Of course, Local Memory Buses and Memory Controllers are included.
- GPIOs – for connection with external or internal modules like LEDs, LCD and DDS block.
- Timers – for generating different time intervals
- Interrupt system - for servicing the interrupts from GPIOs and timers
- SPI for communication with DAC and ADC circuitry
- UART interface for PC connection
- Processor System Reset and Clock Generator.

The program algorithm of the processor is described in section IV.

D. ADC and DAC block

The Spartan-3E Starter Kit board includes ADC and DAC circuitry which are very convenient for connection with the analog circuitry. The DAC device is a Linear Technology LTC2624 quad DAC with 12-bit unsigned resolution. It is SPI-compatible, four-channel, serial Digital-to-Analog Converter. Two of the channels of the DAC are referenced to 3,3 V and the other twos to 2,5 V. We will use the 2,5 V referenced channels for sine and cosine production.

For analog signal input we are using the onboard analog capturing circuit. The circuit consists of a Linear Technology LTC6912-1 programmable preamplifier that scales the incoming analog signal. The output of the pre-amplifier connects to a Linear Technology LTC1407A-1 ADC. The reference voltage for the amplifier and the ADC is 1,65V. The

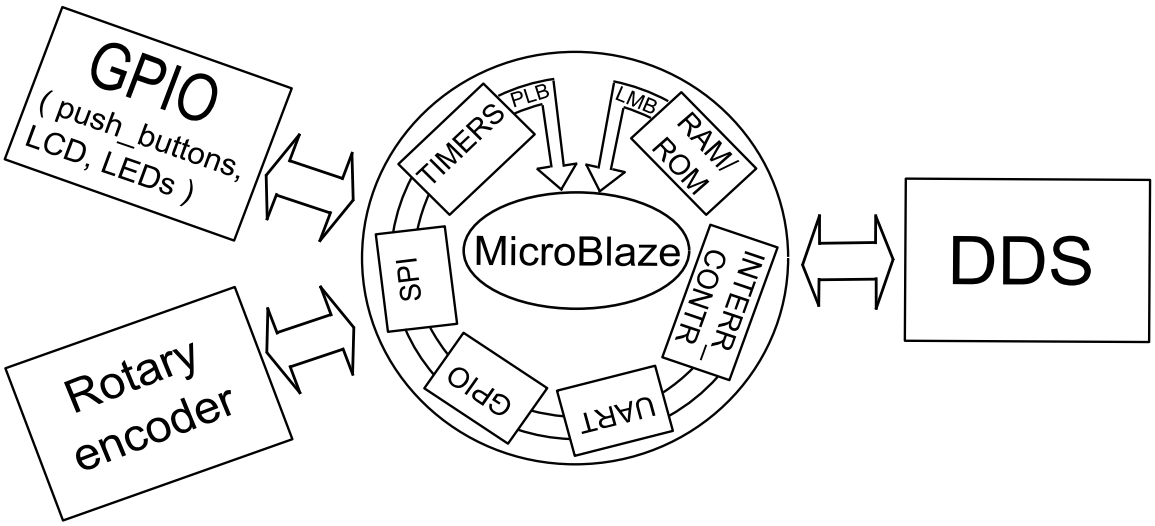


Fig. 2. FPGA – Block structure.

ADC presents the digital representation of the sampled analog values as a 14-bit, two's complement binary value [6], [7]. Both the pre-amplifier and the ADC are serially programmed or controlled by the FPGA [7].

IV. MICROBLAZE PROGRAM ALGORITHM

As already mentioned, the complex signal processing is performed by the analog part of the system so the processor program algorithm is quite simple. First of all the processor should control the output frequency of the DDS block in such a way that a linear sweep is formed. This is achieved in the main loop by incrementing the Frequency Tuning Word (FTW) on regular intervals Δt with some STEP value. When FTW reach the MAX value the process repeats. The user should take care for appropriate set of MAX, MIN, Δt and STEP constants through the user interface (push-buttons, rotary encoder and LCD).

Another important task of the processor is to transmit the output data from the DDS block to the DAC through the SPI interface. The program algorithm is shown on the Fig. 3.

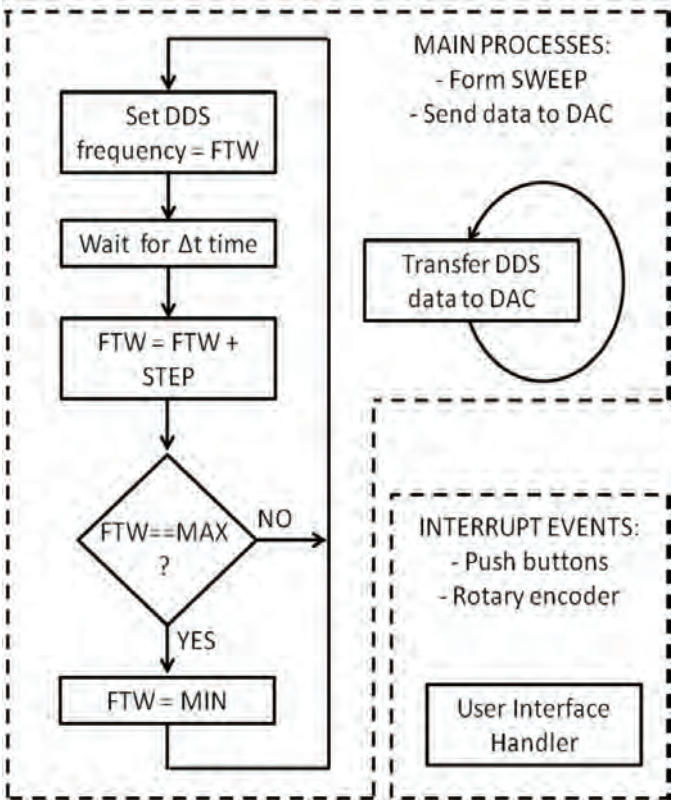


Fig. 3. Microblaze program algorithm.

V. Conclusion

CADSP is very interesting and relatively new engineering solution for signal processing. It exploits the computing power of digital as well as analog circuits which can greatly benefit the system parameters. The current paper presented the design of the digital control system which is a part of a Spectrum analyzer based on CADSP. The design is built around Microblaze soft processor which is charge for all the system control tasks. Moreover a DDS generator was implemented in the FPGA so a specific sine and cosine signals, necessary by the SD, could be produced. The author hopes that system will be realized soon and some experimental results could be reported.

Acknowledgement

This paper is a part of a project under contract № 102pd209-3/2011, which is sponsored by the research program of the Technical University of Sofia, Bulgaria. I'd like to thanks to Digilent Inc. for their Academic Program.

References

[1] FPAA-based Synchronous Detector Unsusceptible to the Signal Phase, ICEST, 2011

[2] V. Tietze, Ch. Schenk, Halbleiter-Schaltungstechnik. 12. Auflage. Berlin, Heidelberg, New York. Springer-Verlag, 2002.

[3] D. Sheingold, Nonlinera circuits handbook, Издателство „Мир", 1977

[4] W. Kester, *Analog-Digital Conversion*. Analog Devices, MA, 2004.

[5] http://www.xilinx.com/tools/microblaze.htm

[6] http://www.linear.com/

[7] Spartan-3E Starter Kit Board User Guide (v1.0), 2006.

Verification of Improved Methodology for Design of Magnetic Components

Georgi Nikolov[1], Vencislav Valchev[2], Emilian Bekov[3],Angel Marinov[4]

Abstract – **In this paper, a verification procedure for a proposed improvement to an existing methodology for design of magnetic components is presented. Two magnetic components are calculated and built. An oscilloscope and calorimeter loss measurement set-upsare used for the verification. The results show good matching between the improved methodology and the real measurements.**

Keywords – **Verification, Magnetic components, Nanocrystalline, Ferrites.**

I. INTRODUCTION

For the verification of the proposed design methodology in [1], two transformers are calculated and built. For the magnetic core of the first transformer, ferrite is used, while for the other one a nanocrystalline soft magnetic material is chosen. The transformers are for a typical power electronic device - an inverter welding unit. The input parameters required for the calculation are shown in TABLE 1. All parameters are the same, except the higher maximum working temperature for the nanocrystalline material.

TABLE 1. INPUT PARAMETERS FOR BOTH TRANSFORMERS

Input voltage	U_1	300V
No load output voltage	$U_{2,NL}$	60V
Full load output voltage	$U_{2,FL}$	26V
Maximum output current	$I_{2,FL}$	150A
Working frequency	f	80kHz
Air velocity	v	2,5m/s
Duty ratio of the voltage	D	10÷90%
Maximum ambient temperature	T_{amb}	30°C
Maximum working temperature for ferrite transformer	$T_{w,F}$	90°C
Maximum working temperature for nanocrystalline transformer	$T_{w,N}$	110°C

[1]Georgi T. Nikolov - TechnicalUniversity – Varna, Varna "Studentska" №1, Bulgaria, E-mail: nikolov_george@hotmail.com
[2]Vencislav C. Valchev - TechnicalUniversity – Varna, Varna "Studentska" №1, Bulgaria, E-mail: vencivalchev@hotmail.com
[3]Emilian Bekov - TechnicalUniversity – Varna, Varna "Studentska" №1, Bulgaria, E-mail: emo_bekov@hotmail.com
[4]Angel S. Marinov - TechnicalUniversity – Varna, Varna "Studentska" №1, Bulgaria, E-mail: igdrazil@abv.bg

II. DESIGN PROCEDURE FOR FERRITE TRANSFORMER WITH FORCED COOLING

The transformer is calculated using the "Fast design approach" [2] and the proposed improvements in [1].

First step is to determine the largest dimension of the component - a_{ch}. The transformer is with only one secondary winding. Usually the transformers that are used in power electronics have efficiency higher than 95%. This allows to simplify the calculation by assuming that the primary power is equal to the secondary one:

$$S_{tot} = \sum_{all\ windings} U_{rms}\, I_{rms} = 300.13 + 26.150 = 7800W \quad (1)$$

$$a_{ch} = \left(\frac{S_{tot}}{A}\right)^{1/\gamma} = \left(\frac{7800}{20.10^6}\right)^{1/3,16} = 0,0724m = 7,2cm \quad (2)$$

The parameters $\gamma$and A are calculated by the equations presented in [2]. Using these equations the value for $\gamma=3,16$ is calculated and $A=20\cdot10^6$ is chosen. As a result core EE80/38/20 is selected with largest dimension $a_{ch}=80mm$ and ferrite type 3F3 by Ferroxcube ([5]).

The next step in the design is to calculate the heat dissipating capability., For T_{amb}=30°C and $T_{w,F}$=90°C, by using the proposed improvements in [1] the value for $P_h=49,4W$ is calculated. Most often, the heat generated by the core and the windings is equally divided [3], [4]:

$$P_{h,cu} = P_{h,fe} = \frac{P_h}{2} = \frac{49,4}{2} = 24,7W \quad (3)$$

Using Eq. (3) the specific core losses can be calculated:

$$P_{fe,sp,v} = \frac{P_{fe}}{V_e} = \frac{24,7}{72300.10^{-9}} = 342kW/m^3 \quad (4)$$

The effective volume of the core $V_e=72300mm^3$ is taken from the datasheet.

By using the parameters from TABLE 1 and core loss model, described in [6], the corresponding peak induction $B_p=0,17T$ is calculated. This is lower than the saturation peak induction $B_{sat}=0,33T$ at 100°C.

To calculate the number of turns, first the peak-to-peak magnetic flux linkage must be found:

$$\psi_{pp} = V_{in}\frac{T}{2} = \frac{V_{in}}{2f} = \frac{300}{2.80.10^3} = 1,875.10^{-3}Wb \quad (5)$$

Then the number of turns for the primary and the secondary are:

$$N_1 = \frac{\psi_p}{\Phi_p} = \frac{\psi_p}{A_e.B_p} = \frac{1,875.10^{-3}}{392.10^{-6}.0,17} = 14,0\,\text{turns} \quad (6)$$

$$N_2 = N_1.\frac{V_2}{V_1} = 14.\frac{60}{300} = 2,8\,\text{turns} \quad (7)$$

The number of turns for the secondary are rounded up to 3. This leads to slight increase in the secondary voltage $U_{2,NL}=64,3V$.

To distribute the allowed total copper losses $P_{h,cu}$ among the windings, the coefficient α_i is used.

$$\alpha_1 = \frac{N_1.I_{rms,1}}{N_1.I_{rms,1}+N_1.I_{rms,1}};\quad \alpha_1 = 0,29;\quad \alpha_2 = 0,71; \tag{8}$$

$$P_{h,cu,1} = \alpha_1.P_{h,cu};\quad P_{h,cu,1} = 0,29.24,7 = 7,11W \tag{9}$$

$$P_{h,cu,2} = \alpha_2.P_{h,cu};\quad P_{h,cu,2} = 0,71.24,7 = 17,59W \tag{10}$$

The diameter of the copper wire used for the primary and the secondary windings is calculated by using Eq. (11) – Eq.(12).

$$d_1 \geq \frac{2}{\sqrt{\pi}}.I_1.\sqrt{\frac{\rho_c.l_{T1}.N_1}{P_{h,cu,1}}} = 1,17mm \tag{11}$$

$$d_2 \geq \frac{2}{\sqrt{\pi}}.I_2.\sqrt{\frac{\rho_c.l_{T2}.N_2}{P_{h,cu,1}}} = 3,97mm \tag{12}$$

Standard enamel copper wires type ПЕТ-2F with temperature grade+155°C are selected. The diameter for the primary windings is $d_{1,p}=1,30mm$ and for the secondary $d_{2,p}=5,00mm$.

With already selected wires, the copper ohmic and eddy current losses can be calculated:

$$P_{cu,ohm1} = \rho_c.l_{T1}.N_1\left(\frac{4}{\pi.d_{p,1}^2.p_1}\right)I_1^2 = 5,74W \tag{13}$$

$$P_{cu,ohm2} = \rho_c.l_{T1}.N_1\left(\frac{4}{\pi.d_{p,1}^2.p_1}\right)I_1^2 = 11,07W \tag{14}$$

One of the advantages of the "Fast design approach" is the calculating of the eddy current losses. The coefficientk_c is introduced. It shows how much the eddy current losses are larger than the ohmic losses:

$$P_{cu,eddy1} = k_{c,1}.P_{cu,ohm1};\ P_{cu,eddy2} = k_{c,2}.P_{cu,ohm2} \tag{15}$$

For calculating k_cEq. (16) is used:

$$k_c = m_e^2.k_{tf} \tag{16}$$

where

$m_e$is the equivalent number of layers;

k_{tf} – a coefficient used in the "Fast design approach".

The coefficient $k_{tf}$is determined by the graphs presented in [2]. To do this, first the equivalent frequency f_{eq} must be found.

$$f_{eq,1} = f.\left(\frac{d_{p1}}{0,5\,\text{mm}}\right)^2.\left(\frac{23.10^{-9}}{\rho_c}\right) = 540,8kHz \tag{17}$$

$$f_{eq,2} = f.\left(\frac{d_{p2}}{0,5\,\text{mm}}\right)^2.\left(\frac{23.10^{-9}}{\rho_c}\right) = 8,0MHz \tag{18}$$

The two coefficients can be found from the graphs. As a result the following values are obtained:

$$k_{c,1} = 0,98;\quad k_{c,2} = 5,51 \tag{19}$$

The results for the eddy current losses are:

$$P_{cu,eddy1} = k_{c,1}.P_{cu,ohm1} = 0,98.5,74 = 5,63W \tag{20}$$

$$P_{cu,eddy2} = k_{c,2}.P_{cu,ohm2} = 5,51.11,07 = 61,00W \tag{21}$$

Summing all results together give the total copper losses:

$$\begin{aligned} P_{tot} &= P_{cu,ohm1} + P_{cu,eddy1} + P_{cu,ohm2} + P_{cu,eddy2} = \\ &= 5,74 + 5,63 + 11,07 + 61,00 = 83,44W \end{aligned} \tag{22}$$

This is more than three times the maximum allowed copper losses 24,7W, calculated by Eq.(3).

Transformer with such losses will overheat and consequently be damaged, an optimization must be carried out in order to decrease the copper losses. Some of the possibilities are [7]:

- Increasing the wire diameter;
- Connecting several wires in parallel;
- Using Litz wire;
- Using copper foil or wires with rectangular cross sections.
- Use larger core

Increasing the diameter of the wire is used for partially filled layers. The diameter is increased until the layer is filled completely. For low number of turns (<5), this is not practical. In this case, it is best to use the second technique – wires in parallel. Usually 2÷4 wires are put in parallel. This allows the same cross section area to be obtained by using smaller wires, and hence smaller copper losses. The results from such optimizations are presented in Table2.

Case 0 is the default – with no optimizations. In case 1, the first two presented optimizations are used (increasing the wire diameter and connecting several wires in parallel). For Case 2, Litz wire is used. This results in lowering the copper losses up to the maximum value. However, because of the additional insulation, the cooling properties are worsen. Together with the fact that this design is "just on the limit", gives enough reasons to continue with the optimization.

Two sets of EE80/38/20 are used for cases 3 and 4. The MLT and the copper wire resistance are calculated more precisely, by taking into account difference in the MLT for the primary and the secondary windings and the resistance temperature dependence. This leads to lowering the losses bellow the maximum value safety margin of about 15% in case 4. With this step the optimization procedure for the ferrite transformer is over.

Similar design and optimization are carried out for the nanocrystalline core transformer. The results are presented in case 5. Two sets of the cut core F3CC0010 are used with nanocrystalline material FT-3M. The maximum working temperature for this material is 110°C. The temperature dependence of the wire resistance and different MLT are also taken into account.

Cases 4 and 5 use wires with square cross section. This results in easier winding and better utilizing of the window area. To use such wires, a conversion from rectangular to equivalent square wires is used. Different techniques exist for such conversions -[2] and [3].

The results from such optimizations are presented in Table2.

TABLE2. DESIGNED TRANSFORMERS

Parameter	Winding	Symbol	Unit	Ferrite core					Nanocrsyt alline
				No optimizations	Wires in parallel	Litz wire	2 x E80	2 x E80 and square wires	
				Case0	**Case1**	**Case2**	**Case3**	**Case4**	**Case5**
Number of turns	primary	N_{1p}	-	14			10		11
	secondary	N_{2p}	-	3			2		2
Magnetic induction		B	T	0,17			0,13		0,26
Diameter of the wire (with insulation)	primary	$d_{1,p}$	mm	1,30 (1,41)	3,55 (3,68)	0,20 (2,65)	1,00 (1,09)	2,0 (2,11)	1,50 (1,63)
	secondary	$d_{2,p}$	mm	5,00 (5,14)	4,00 (4,13)	0,20 (4,97)	4,5 (4,64)	3x7,5 (3,8x7,9)	0,7x32 (1,3x32,6)
Wires in parallel	primary	p_1	-	1	1	100	1	1	2
	secondary	p_2	-	1	4	23350	1	1	1
Equivalent number of layers	primary	m_{E1}	-	1	1	10	1	1	1
	secondary	m_{E2}	-	1	1	26,5	1	1	2
Equivalent number of turns in a layer	primary	n_{E1}	-	14	14	140	10	10	22
	secondary	n_{E2}	-	3	12	79,4	2	4,43	46
Copper fill factor in the direction of the layer	primary	η_1	-	0,34	0,93	0,53	0,19	0,38	0,89
	secondary	η_2	-	0,28	0,93	0,3	0,17	0,28	0,98
Copper fill factor in the direction perpendicular to the layer	primary	λ_1	-	0,07	0,18	0,1	0,05	0,10	0,12
	secondary	λ_2	-	0,26	0,21	0,27	0,23	0,17	0,12
Equivalent frequency	primary	$f_{eq,1}$	kHz	540	3920	12,8	320	1280	719
	secondary	$f_{eq,2}$	kHz	8000	5120	12,8	6480	3670	199
Eddy current coefficient	primary	k_{c1}	-	1,0	9,4	0,4	0,4	2,3	3,4
	secondary	k_{c2}	-	5,5	10,6	1,0	4,2	3,7	1,4
Ohmic resistance	primary	$R_{dc,1}$	mΩ	34,0	4,7	15,1	23,2	5,9	10,0
	secondary	$R_{dc,2}$	mΩ	0,5	0,2	0,5	0,3	0,2	0,3
Ohmic losses	primary	$P_{cu,ohm,1}$	W	5,7	0,8	2,6	3,9	1,0	1,5
	secondary	$P_{cu,ohm,2}$	W	11,1	4,3	10,7	5,8	2,3	6,5
Eddy current losses	primary	$P_{cu,eddy,1}$	W	5,6	7,4	1,1	1,4	4,7	5,2
	secondary	$P_{cu,eddy,2}$	W	61,0	45,9	10,5	24,3	17,4	9,0
Copper losses	both	P_{cu}	W	83,8	58,5	24,8	35,5	25,4	22,2
Maximum allowed copper losses	both	P_h	W	24,7			29,8		25,9

The last step in the design of the transformer is to check, is there enough space to place physically all the windings in the transformer window:

$$k_{cu} = \frac{\sum_{i=1}^{n} p_i . N_i . \frac{\pi . d_{i,p}^2}{4}}{W_a} \quad (23)$$

where:

*p*is number of wires in parallel (or the litz wire strands);

k_{cu} – filling factor;

W_a – window area.

For round wires $k_{cu}=0,5 \div 0,8$ and for litz wires it should be $k_{cu}=0,3 \div 0,5$.

For all of the designed transformers, this coefficient fits to the reference values.

For conducting real experiments, the transformers from case 4 and 5 are built and can be seen in Fig.1.

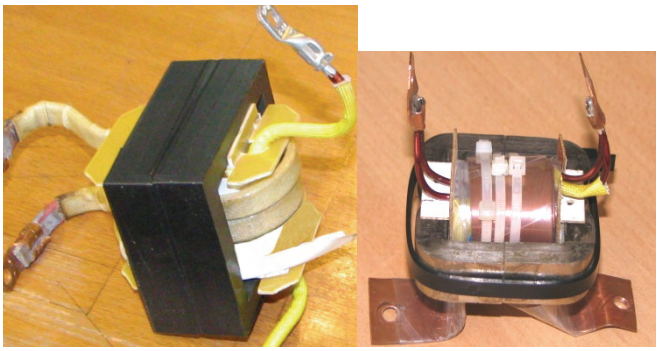


Fig.1Transformers from Case 4(left) and Case 5(right)

Short circuit and no load tests are carried out with the transformers. The system described in [8] is used for the loss measurements. The results are presented in TABLE3:

TABLE3. CALCULATED AND MEASURED LOSSES

	Winding losses				Core losses			
	calculated	measured	difference		calculated	measured	difference	
	W	W	W	%	W	W	W	%
Case 4	25,38	24,71	-0,67	-2,7	29,80	28,64	-1,16	-4,1
Case 5	22,19	21,04	-1,15	-5,5	25,91	25,8	-0,11	-0,4

One can see, that the measured values are about 3% less than calculated. This can compensate for some small manufacturing tolerances and rounding of numbers. This also means that with the proposed improvements it is possible to design magnetic componentsfor the typical conditions met in power electronics.

A comparison between the two transformers is done in TABLE4.

TABLE4. COMPARISON BETWEEN THE TWO TRANSFORMERS

		Case 4	Case 5	Difference	Improvement
Mass	**kg**	1,176	0,572	0,604	51,4%
Area	**mm2**	6080	4608	1472	24,2%
Volume	**mm3**	340480	211968	128512	37,7%
Price	**€**	2x(6,00÷7,00)	2x(7÷30)		0 ÷ -400%

It can be seen from the results in TABLE4, that using nanocrystalline core leads to significant improvement in the mass, area and the volume of the core. The higher price remains a disadvantage. However there are now some Chinese manufacturers (like Foshan Huaxin Microlite Metal Co Ltd 0) of nanocrystalline soft magnetic materials that offers several times cheaper price compared to the leading manufacturers (Hitachi metals - [10], Vacuumschmelze - [11]). This will only lead to decrease the prices of nanocrystalline cores in global aspect.

III. CONCLUSION

This paper gives a verification of the proposed "Improved methodology for design of magnetic components". Several transformers are calculated and two of them are actually built. Some experiments are conducted and the results show matching with the measured data better than 5%.

ACKNOWLEDGEMENT

The paper is developed in the frames of the project 'Development of specialized scientific infrastructure for investigation of wind and solar energy potential', № Д 002-48/10.12.2008, Ministry of Education, Youth and Science, Fund 'Scientific Research'.

REFERENCES

[1] V. Valchev, G. Nikolov, A. Marinov, Improved methodology for design of magnetic components, ICEST 2011, 29 june – 1 july, Niš Serbia.

[2] A. Van den Bossche, V. Valchev, *Inductors and Transformers for Power Electronics*, Boca Ration USA, CRC Press, ISBN: 157-444-679-7, 2005.

[3] E. Snelling, *Soft ferrites: properties and applications*, Butterworth- Heinemann, 2-nd edition, 1988.

[4] L. Jansson, Power-handling capability of ferrite transformers and chokes for switched-mode power supplies, Technical Note 31, Mullard Limited, 1975.

[5] Ferroxcube GmbH, <http://www.ferroxcube.com>.

[6] G. Nikolov, V. Valchev, "Power loss model for nanocrystalline and ferrite soft magnetic materials with squarewave voltages", Annual Proceedings of the TechnicalUniversity in Varna – ISSN: 1311-896X, Varna pp.97-102, 2009.

[7] V. Valchev, G. Nikolov, Comparison of nanocrystalline magnetic materials and ferrites used for power electronics transformers' – ICEST'08, Nis, Serbia, 24-26 June, 2008, pp.276-279.

[8] V. Valchev, G. Nikolov, "Advanced set-up for high frequency measurements of magnetic materials and components in electronics" – ICEST09, V. Tarnovo, Bulgaria, 2009

[9] Foshan huaxin Microlite Metals Co. Ltd, <http://www.fshuaxin.com>.

[10] Hitachi metals Ltd., <http://www.hitachi-metals.co.jp/e/>.

[11] Vacuumschmelze GmbH & Co. KG, <http://www.vacuumschmelze.de>

Development of Industrial Circuits with Semiconductor Diodes and Optoelectronic Elements

Elena Koleva[1], Ivan Kolev[2]

Abstract – **In this paper are considered practical optoelectronic circuits for industrial applications with improved parameters based on a combination of optoelectronic and electronic components.This ledto some benefits such as increasing the performance of the circuits protect input and output circuits for large-value reverse voltages and limit the pulse current through the emitters.**

Keywords – **Semiconductor Diodes, Light Emitting Diodes (LEDs), Optoelectronic Elements.**

I. INTRODUCTION

The main applications of semiconductor diodes in optoelectronic circuits are [3]:

- Protectionof LEDs and Laser diodes from reverse voltages;
- Inclusion of LED to alternating or two – pole voltage;
- Protection of input or output circuits of optoelectronic circuits by reverse or rise voltage, or fixing signals voltage levels (diode limiters);
- Limiters in current loop of the LED;
- Change of regimes work of transistor switches, acceleration circuits, differentiating circuits, non – saturation switches;
- Pick – up of current to DC/ AC circuits;
- Diode – resistor and diode – transistor logic circuits;
- Forming, threshold, comparing and relay circuits.

II. INDUSTRIAL CIRCUITS WITH SEMICONDUCTOR DIODES AND OPTOELECTRONIC ELEMENTS

A. Control of LEDs of non – saturated transistor switches – Fig. 1

The fast – action of the transistor switches is important for the control of light sources – LEDs and laser diodes. The non – saturated transistor switches has high performance by the saturated transistor switch in the circuit OE.

To not take VT_1 transistor in saturation mode and reduce of the fast – action of the switch, is used a fixed diode VD_1, which carried nonlinear optical negative feedback.

[1]Elena N. Koleva is Ph. D., Department ofElectronics, Technical University – Gabrovo, Street "Hadji Dimitar" No. 4, 5300 Gabrovo, Bulgaria, phone: +359 898 226 464,
e-mail: elena_ndpt@yahoo.com

[2]Ivan S. Kolev is Prof., Dr. Sci., Department of Electronics, Technical University – Gabrovo, Street "Hadji Dimitar" No. 4, 5300 Gabrovo, Bulgaria, phone: +359 898 634 633,
e-mail: ipk_kolev@yahoo.com

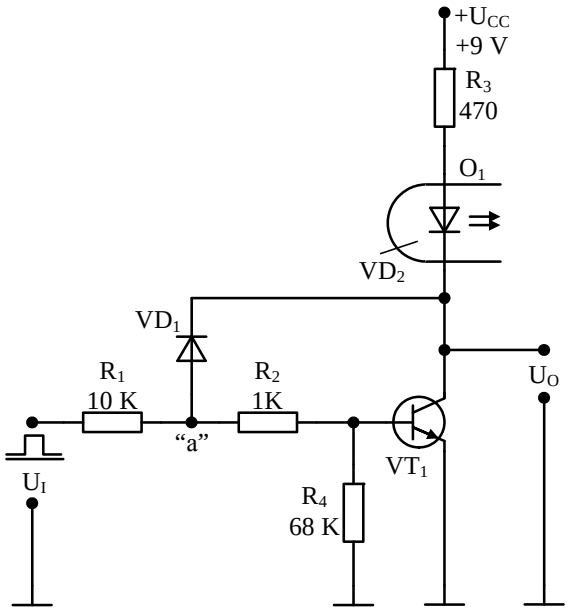


Fig. 1. Circuit for control of LEDs of non – saturated transistor switches

When the collector voltage of transistor VT_1 become – more than the value is Eq. 1:

$$U_{CE} > U_{F1} + U_a = U_{F1} + \frac{U_I - U_{BE1}}{R_1 + R_2}.R_2 + U_{BE1} =$$

$$= 0{,}7 + \frac{5 - 0{,}7}{10.10^3 + 1.10^3}.1.10^3 + 0{,}7 \approx 1{,}8\,V \qquad (1)$$

Diode VD_1 is the voltage and opens the collector of the transistor VT_1 is fixed at 1,8 V, but not as usual 0,1 ÷ 0,3 V. Times on and off the transistor VT_1 reduced to 20 %. The disadvantage of the circuit is – the small current through the LED of key non – saturated compared with saturated switch. Current through the LED ofthe non – saturated switch is Eq. 2:

$$I_F = \frac{U_{CC} - U_{F1} - U_{CE1}}{R_3} = \frac{9 - 1{,}2 - 1{,}8}{470} \approx 13\,mA \qquad (2)$$

For saturate switch is Eq. 3:

$$I_F = \frac{U_{CC} - U_{F1} - U_{CEsat1}}{R_3} = \frac{9 - 1{,}2 - 0{,}1}{470} \approx 16\,mA \qquad (3)$$

It is seen that the current through the LED at switch non – saturated reduced nearly 20 %.

B. Protection of input and output loops of optoelectronic circuits by over voltages

Protection of input loops from the negative (inverse) input voltages – Fig. 2 and Fig. 3.

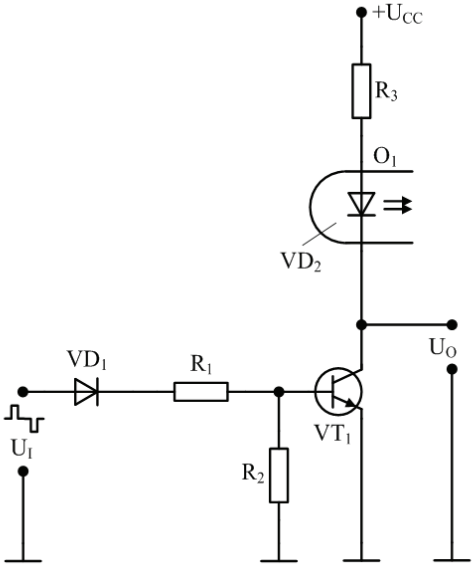


Fig. 2. Optoelectronic circuit for protection of input loop

In the circuit of Fig. 2 to optoelectronic switch is madeonly positive input voltages with an amplitude > UF_1 (0,7 V), where the circuit of Fig. 3 all positive input voltages, the negative input voltages are fixed at UF_1 = - 0,7 V.

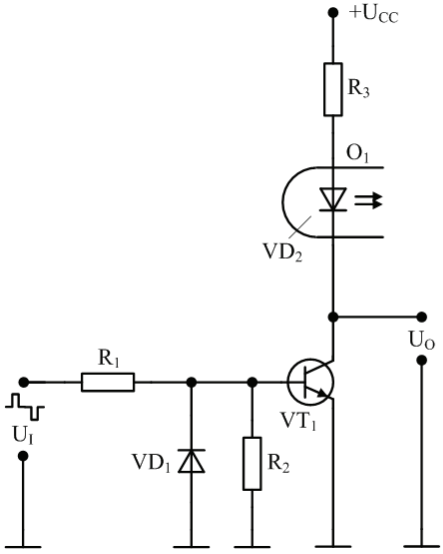


Fig. 3. Optoelectronic circuit for protection of input loop

Protection of the output loops of optoelectronic circuits from negative voltages – Fig. 4, Fig. 5 and Fig. 6.

In the circuit of Fig. 4 and Fig. 5 the diode VD_1 protects the transistor VT_1 and LED VD_2 by reverse voltage U_{CC}. The diode VD_1 can be LED, [4].

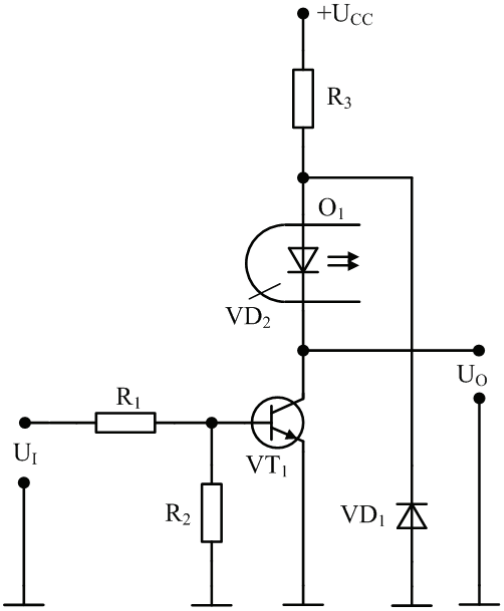


Fig. 4. Optoelectronic circuit for protection of output loop

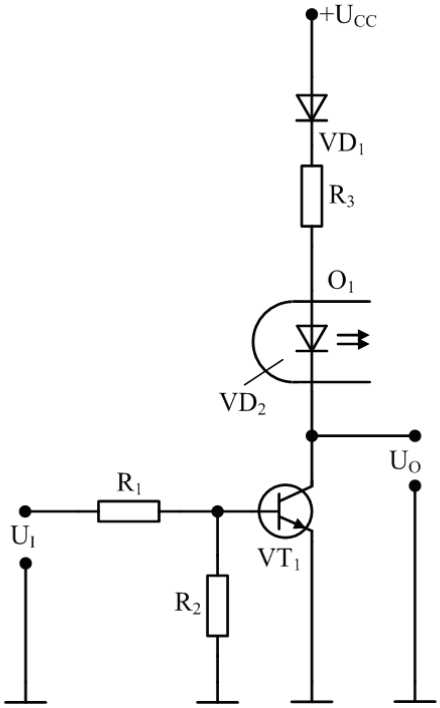


Fig. 5. Optoelectronic circuit for protection of output loop

In the circuit of Fig. 6 diodes VD_1 protect the circuit from reverse of the supply voltage. In the circuit of Fig. 6 the protection from reverse supply voltage U_{CC} is done either with diodes VD_1 or diodes VD_2.

The LED VD_3 radiated in reverse supply voltage U_{CC}.

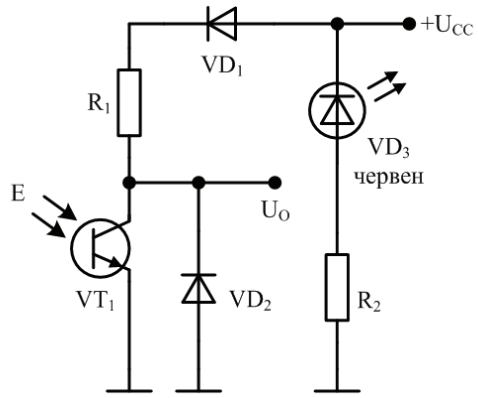


Fig. 6.Optoelectronic circuit for protection of output loop

Limitof the pulse current through the LEDs. In the work of the LED with short pulses (10 μs) and power current pulses (1 ÷ 2) A used a low – omnic resistor (10 Ω) or more LEDs several series connected diodes to limit current during LED – Fig. 7 and Fig. 8, [1].

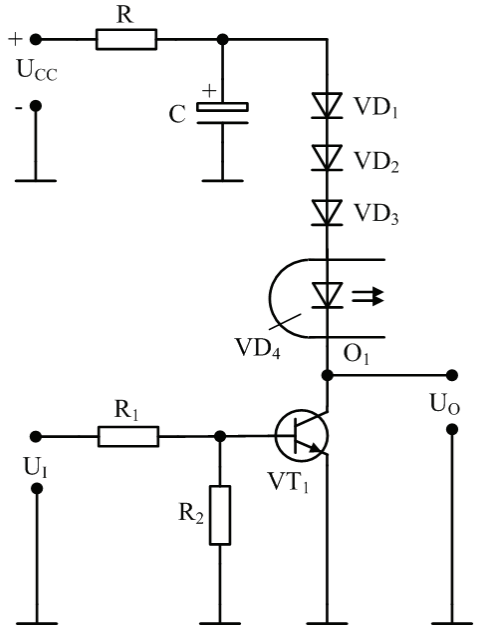


Fig. 7. Circuit of limit of the pulse current though the LED

Powerful electrical pulse is obtained either by discharging the capacitor C in LED – Fig. 7 or the inclusion of the LED for a short time to the supply voltage – Fig. 8.

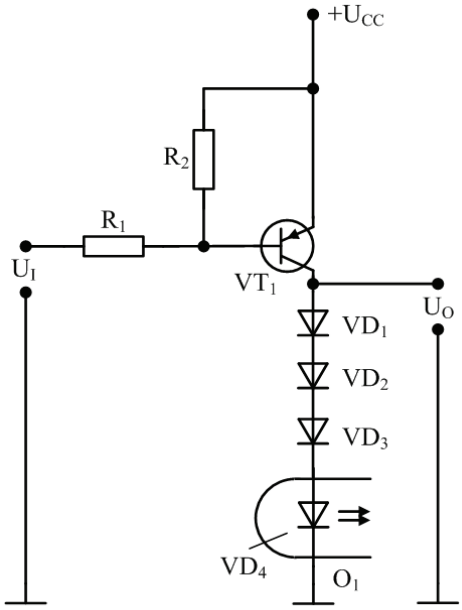


Fig. 8.Circuit of limit of the pulse current though the LED

In both cases the pulse current through the LED is Eq. 4:

$$I_{FP} = \frac{U_{CC} - U_{F4} - U_{CEsat1}}{R_{F1} + R_{F2} + R_{F3}} = \frac{9 - 1{,}2 - 1{,}3}{1+1+1} \approx 2\ A \quad (4)$$

Typically, transistors VT_1 are darlington U_{CEsat} and the voltage is high (more than 1 V).

C. Increasing of the fast – action of LEDs in saturated transistor switches– Fig. 9, Fig. 10

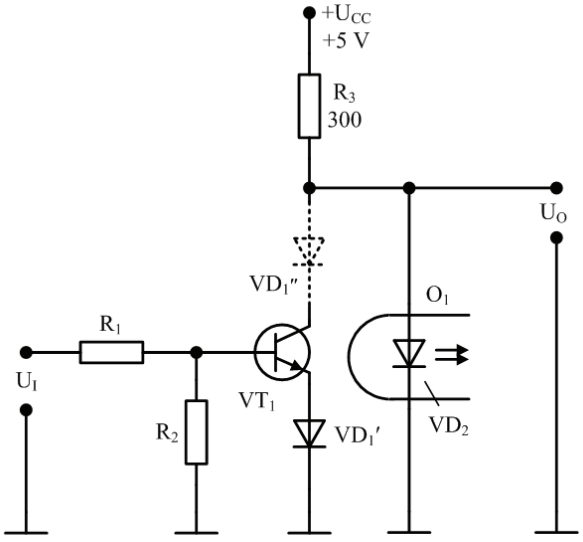


Fig. 9. Circuit for increasing of the fast – action of the LED in saturated transistor switch

When the transistors VT_1 are blocked (U_I = 0 V – Fig. 8 and U_I = 5 V – Fig. 9) during the current LED are Eq. 5 and Eq. 6:

$$I_F = \frac{U_{CC} - U_{F2}}{R_3} = \frac{5 - 1{,}2}{300} \approx 13\, mA \quad (5)$$

$$U_{F2} = U_{CEsat1} + U_{F1} = 0{,}1 + 0{,}7 = 0{,}8\, V \quad (6)$$

When the transistors VT_1 are in on – regime the voltage on the current IF_1 trough the LED is not zero, as in the switchesin Fig. 1 ÷ Fig. 7.

Low current flowsin the LED, several hundred μA, [2]. This inclusion leads to increased of the fast – action of the LED to 20 %. In the circuit of Fig. 8 when the output voltage U_O by saturated and blocked transistor VT_1 are Eq. 7 and Eq. 8:

$$U_{OH} = U_{F2} = 1{,}2\, V \quad (7)$$

$$U_{OL} = U_{CEsat1} + U_{F1} = 0{,}1 + 0{,}7 = 0{,}8\, V \quad (8)$$

The diode VD_1 can be incorporated into the collector or the emitter circuit. Unlike the circuit in Fig. 9 and Fig. 10, the diodeVD_1 is replaced by the LED and included two additional diodes VD_3 and VD_4.

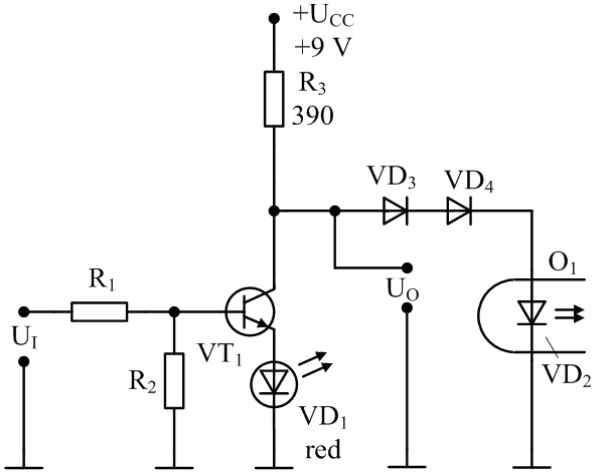


Fig. 10.Circuit for increasing of the fast – action of the LED in saturated transistor switch

When the transistor VT_1 is blocked (U_I= 0 V), current trough the LED is Eq. 9:

$$I_{F2} = \frac{U_{CC} - U_{F3} - U_{F4} - U_{F2}}{R_3} = \frac{9 - 0{,}7 - 0{,}7 - 1{,}2}{390} \approx 16{,}4\, mA \quad (9)$$

The output voltage U_O in this case is Eq. 10:

$$U_{OH} = U_{F3} + U_{F4} + U_{F2} = 0{,}7 + 0{,}7 + 1{,}2 = 2{,}6\, V \quad (10)$$

If U_I = 5 V, the transistor VT_1 is saturated and output voltage is Eq. 11:

$$U_{OL} = U_{CEsat1} + U_{F1} = 0{,}1 + 2 = 2{,}1\, V \quad (11)$$

Then voltage on the LED $VD_2$is Eq. 12:

$$U_{F2} = U_{OL} - U_{F3} - U_{F4} = 2{,}1 - 0{,}7 - 0{,}7 = 0{,}7\, V \quad (12)$$

and in LED – low current flows, hundreds of μA.

The current in the LEDVD_1 is Eq. 13:

$$I_{F1} = \frac{U_{CC} - U_{CEsat1} - U_{F1}}{R_3} = \frac{9 - 0{,}1 - 2}{390} \approx 18\, mA \quad (13)$$

Application of the developed circuits: protection of the LEDs and Laser diodes, input, output or supply chains of reverse voltage, current limiters, increase the fast – action of the transistor switches and the LEDs, current sensors, protection from phase – down failure.

III. CONCLUSION

Parts of the developed circuits increase the reliability of optoelectronic switches – Fig.2÷Fig.6, protecting them against improper inclusion and increased reverse voltages.

Increase of the fast – action of optoelectronic swiches – Fig. 1, Fig. 9, and Fig. 10.

REFERENCES

[1] Kolev, I. S. and E. N. Koleva. Optoelectronics. Devices. Elements. Aplications. Sofia, Technika publ., 2007.

[2] Kolev, I. S. and E. N. Koleva. Optoelectronic Sensors and Optoelectronic Security Systems. Gabrovo, Univ. publ. „Vasil Aprilov", 2009.

[3] Kolev, I. S. and E. N. Koleva. Industrial Optoelectronic Systems.Gabrovo, Expres publ., 2011.

[4] Koleva, E. N. Industrial Electronics. Gabrovo, Expres publ., 2010.

[5] Iliev T. and Plamen Danailov, "A method of recording and analizing the spectral structure of electrical signals obtained by optical sensors for vibration measurement of electrical machines", XV-th International Symposium on Electrical Apparatus and Technologies,Plovdiv, 31May – 1 June 2007.

Curve Fitting for Sensors' Analog Behavioural Modelling

Boyanka Nikolova[1], Milen Todorov[2], Tihomir Brusev[3]

Abstract – **This article summarizes the main methods for curve fitting of non-linear sensor's characteristics. The results are input data in SPICE model, built according to the basic measuring circuit within the sensor's product information. The curve fitting is achieved by MathWorks MATLAB®'s Curve Fitting Toolbox™ and National Instruments Multisim. Curve Fitting Toolbox™ supplements MATLAB features with data preprocessing capabilities, using parametric and nonparametric models from a library. On the other hand Multisim provides the necessary ABM (analog behavioral model) sources, which use mathematical and conditional expressions to set their output voltage or current. They may contain mathematical and conditional expressions that consist of circuit voltages, currents, time and other simulation parameters. ABM is an extremely powerful feature which provides an efficient way to macro model signal processes through non-linear mathematical and conditional expressions.**

Keywords – **Curve fitting, Simulation methodology, Intelligent sensors, MATLAB, Multisim.**

I. Introduction

At the core of any data acquisition system is interpretation of a voltage signal based on information about the analog sensor that makes the measurement intelligible. Typically, these are standard curves and equations specific to the type of transducer. Sensor calibration, however, takes this process one step farther by considering transducers on an individual basis. The sensor output voltage is mapped to a physical measurement based on metrics obtained from a specific sensor calibration. Although many sensors are linear over the limited range, these sensors exhibit a slight but progressively more nonlinear characteristic as the measurement range widens. Consequently, over an extended span, curve fitting is necessary if the system is to achieve a high level of precision.

With the facility of computation now available through digital computers and microprocessors, the problem of estimation of transducer's transfer characteristics is being increasingly tackled using software techniques. However, for inherent nonlinear sensors, a software solution depends upon the proper approach through mathematical modeling of the response curve [1, 2, 6].

The purpose of this paper is to assist engineers and scientists to implement the newly released Curve Fitting Toolbox in order to achieve more precise results. These results can be used as input data for the simulation models.

[1]Boyanka Nikolova is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: bnikol@tu-sofia.bg

[2]Milen Todorov is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: brusev@ecad.tu-sofia.bg

[3]TihomirBrusev is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: brusev@ecad.tu-sofia.bg

II. Parametric Models in MATLAB

TABLE I
CURVE FITTING TOOLBOX LIBRARY MODELS

Type of Fit	Description
Exponentials $y = ae^{bx}$; $y = ae^{bx} + ce^{dx}$	**a, b, c, d** – model's parameters.
Fourier Series $y = a_0 + \sum_{i=1}^{n} a_i \cos(nwx) + b_i \sin(nwx)$	$\mathbf{a_0}$ – models a intercept term in the data and is associated with the i = 0 cosine term; **w** – fundamental frequency of the signal; **n** – number of terms (harmonics) in the series and $n \in [1;8]$.
Gaussian $y = \sum_{i=1}^{n} a_i e^{\left[-\left(\frac{x-b_i}{c_i}\right)^2\right]}$	**a** – amplitude; **b** – centroid (location); **c** – related to the peak width **n** – number of peaks to fit, $n \in [1;8]$.
Polynomials $y = \sum_{i=1}^{n+1} p_i x^{n+1-i}$	**n+1** – order of the polynomial; **n** – degree of the polynomial, $n \in [1;9]$.
Power Series $y = ax^b$; $y = a + bx^c$	**a, b, c** – model's parameters.
Rationals $y = \frac{\sum_{i=1}^{n+1} p_i x^{n+1-i}}{x^m + \sum_{i=1}^{m} q_i x^{m-i}}$	**n** – degree of the numerator polynomial, $n \in [0;5]$; **m** – degree of the denominator polynomial, $m \in [1;5]$.
Sum of Sines $y = \sum_{i=1}^{n} a_i \sin(b_i x + c_i)$	**a** – amplitude; **b** – frequency; **c** – phase constant for each sine wave term; **n** – number of terms in the series.

Parametric fitting involves finding coefficients (parameters) for one or more models that fit to data [4, 8]. The data is assumed to be statistical in nature and is divided into two components: a deterministic component and a random component [3]. The deterministic component is given by a parametric model and the random component is often described as error associated with the data. The model is a function of the independent (predictor) variable and one or more coefficients. The error represents random variations in the data that follow a specific probability distribution. Systematic variations can also exist, but they will lead to a fitted model that does not represent the data well. To improve

the fit it may be necessary to increase the number of iterations. The fit is well behaved over the entire data range when the residuals are randomly scattered about zero.

After fitting data with one or more models is necessary to evaluate the goodness of fit. A visual examination of the fitted curve displayed in Curve Fitting Tool should be first step. Beyond that, the toolbox provides these methods to assess goodness of fit for both linear and nonlinear parametric fits [3, 4, 8, 9]:

- Residual analysis
- Goodness of fit statistics
- Confidence and prediction bounds.

These methods group into two types: graphical and numerical. Plotting residuals and prediction bounds are graphical methods that aid visual interpretation, while computing goodness of fit statistics and coefficient confidence bounds yield numerical measures that aid statistical reasoning.

Graphical measures are more beneficial than numerical measures because they allow to view the entire data set at once, and they can easily display a wide range of relationships between the model and the data. The numerical measures are more narrowly focused on a particular aspect of the data nad often try to compress that information into a single number. In practice, depending on data and analysis requirements it is necessary to use both types to determine the best fit.

When fitting data that contains random variations, there are two important assumptions that are usually made about the error:

- The error exists only in the response data, and not in the predictor data.
- The errors are random and follow a normal (Gaussian) distribution with zero mean and constant variance, σ^2.

The errors are assumed to be normally distributed because the normal distribution often provides an adequate approximation to the distribution of many measured quantities.

III. Curve Fitting and Simulation

On figure 1 is illustrated the proposed methodology for curve fitting and modeling methodology.

On step 1.1 the chosen characteristic must be represented as workspace vectors with the same length. To perform any curve fitting task, must be selected at least one vector of data:

- **X data** – selects the predictor data.
- **Y data** – selects the response data.

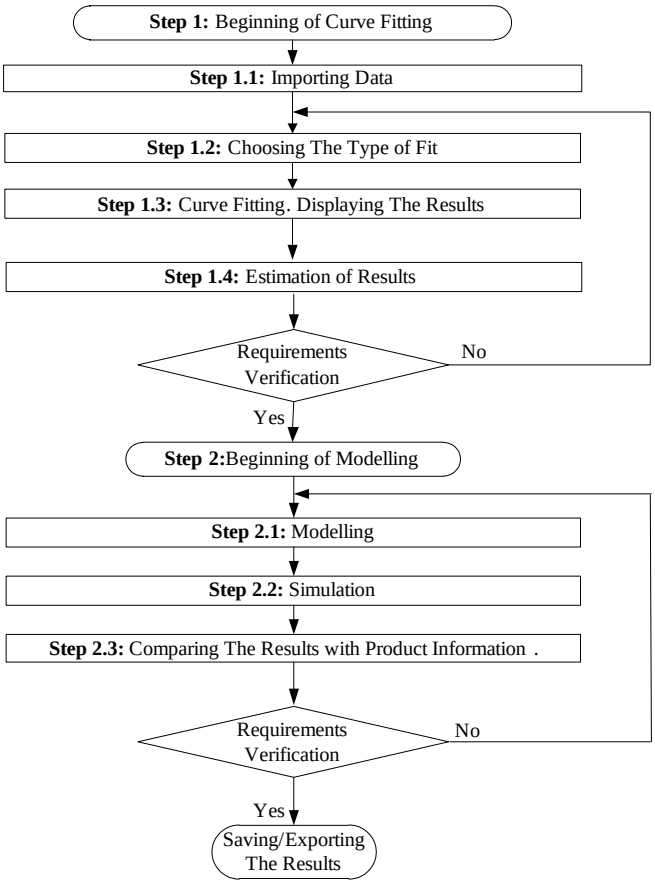


Fig. 1. Block diagram of curve fitting and modeling methodology.

Also it could be set the weights associated with the response data. If weights are not imported, they are assumed to be 1 for all data points.

On the next step 1.2 we choose one model from the library of parametric models. According to the shape and specificity of the sensor characteristic the appropriate selection must be done.

Step 1.3 displays the results of fitting.

The importance of step 1.4 consists in the analysis results. It must be examine the fitted curve, residuals, goodness of fit statistics, confidence bounds, and prediction bounds for the current fit. It is recommended to compare the current fit and data set to previous fits and data sets by examining the goodness of fit statistics. Figure 2 shows this in detail.

On step 2.1 the parts from sensor's basic measuring circuit should be substitute by ABM sources with. As input data for ABM sources is used the results of fitting. The idea of this is seen in detail on figure 3.

Once the model is ready we can run the simulation with given parameters – step 2.2.

The basic structure of the proposed model of gas sensor is shown on figure 4. The main purpose of model is to simulate the change of the output voltage as slope function of the input physical phenomenon. This is achieved with a block "Transfer Function". The environment is presented by block "Externalities". Changing the signal due to external factors (mainly temperature and relative humidity) are calculated in "Externalities" block and added to the sensor's signal by voltage summer.

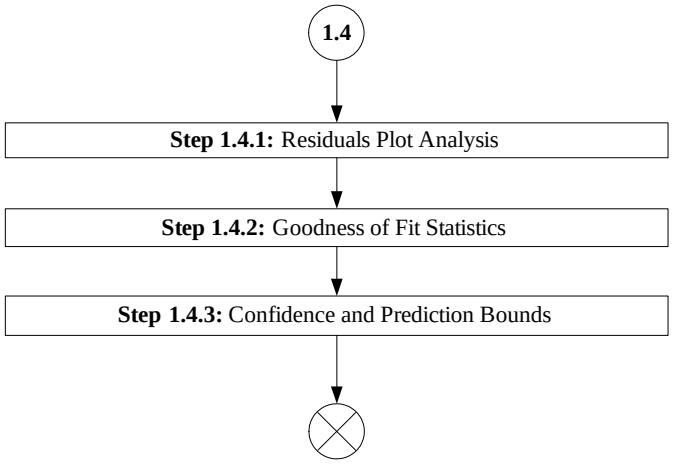


Fig. 2. Estimation of results.

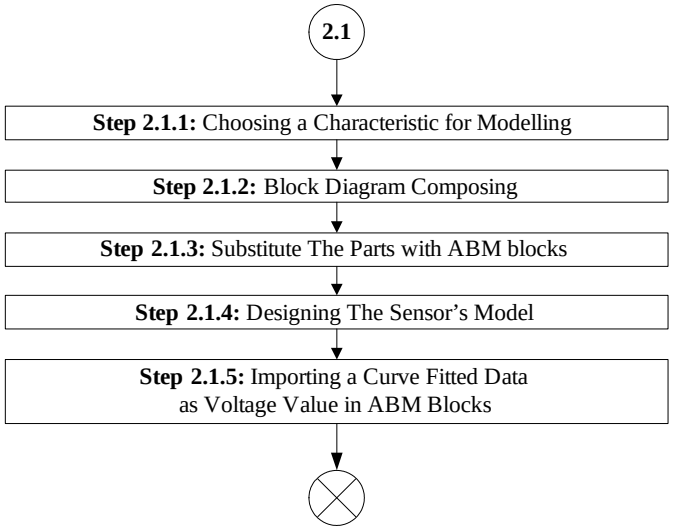


Fig. 3. Modelling.

TABLE II
CURVE FITTING TOOLBOX LIBRARY MODELS

Goodness of fit	Best-case value
Sum of Squares due to Error (SSE)	0
R-square (R^2)	1
Adjusted R-square (Adjusted R^2)	1
Root Mean Squared Error	0

The transfer function and other characteristics, which describe the effects of temperature and humidity, can easily be modeled by appropriate behavioral elements. The behavioral elements use mathematical and conditional expressions to set their output voltage or current. They contain mathematical expressions obtained by curve fitting. The main advantage of using a mathematical function is short simulation time and better convergence [7].

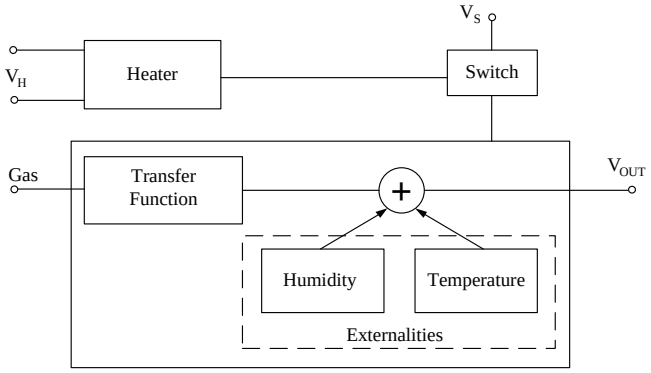


Fig. 4. Gas sensor's model.

Using the existing library of ABM blocks in National Instruments Multisim and figure 4 is developed behavioral model of TGS 2600 gas sensor – figure 5 [5, 7].

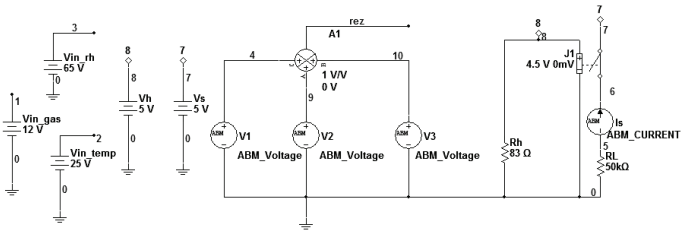


Fig. 5. Gas sensor's model in NI Multisim.

The data, fit, prediction bounds and residuals for fitting of full transfer characteristics are shown in figure 6. In table III are given the fitting numerical results for model's parameters and goodness of fit statistics. For a first fitting model 9th degree polynomial (Poly9) is chosen. For a second model is chosen power model with three parameters (Power2). As shown by results the residuals for Power Fit 2 are randomly scattered around zero and indicating that this model describes the data well. The same conclusion can be done comparing numerical values of goodness of fit statistics.

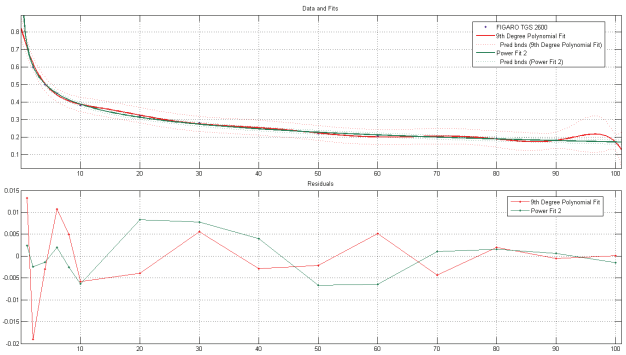


Fig. 6. The data, fit, prediction bounds and residuals for transfer characteristic of TGS 2600 sensor.

TABLE III
RESULTS FROM CURVE FITTING OF TGS 2600 TRANSFER CHARACTERISTIC

Type of fit		9th Degree Polynomial		Power Fit 2
Model		$f(x) = p1.x^9+p2.x^8+p3.x^7+p4.x^6+$ $+p5.x^5+p6.x^4+p7.x^3+p8.x^2+p9.x+p10$		$f(x) = a.x^b+c$
Coefficients	p1	-2.115e-015	a	0.9434
	p2	9.705e-013	b	-0.1878
	p3	-1.878e-010	c	-0.2257
	p4	1.997e-008		
	p5	-1.273e-006		
	p6	5e-005		
	p7	-0.001195		
	p8	0.01671		
	p9	-0.1302		
	p10	0.8215		
Goodness of fit	SSE	0.0008261		0.0003024
	R-square	0.9979		0.9992
	Adjusted R-square	0.9942		0.9991
	RMSE	0.01285		0.00502

Results of the simulation process in NI Multisim are shown in figures 7 and 8.

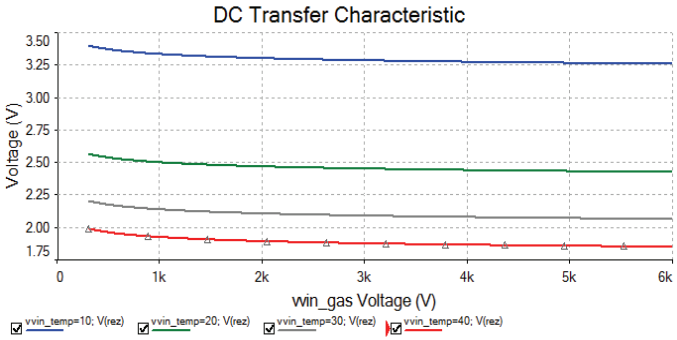


Fig. 7. Sensor's transfer function under the influence of temperature.

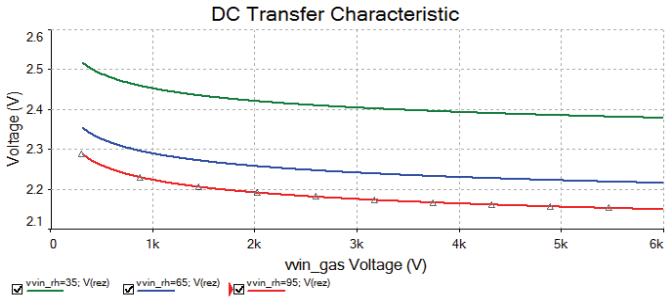


Fig. 8. Sensor's transfer function under the influence of relative humidity.

IV. CONCLUSION

This paper suggest an systematic approach for implementing curve fitting models and methods in order to achieve equation that precisely describe sensor's transfer characteristic and used as initial conditions for simulation process. As illustration experimental results from the whole process are presented.

ACKNOWLEDGEMENT

The research described in this paper was carried out within the framework of Contract 102 НИ 197-3.

REFERENCES

[1] B. Nikolova, G. Nikolov, M. Todorov. "Curve Fitting of Sensors' Characteristics", Annual Journal of Electronics, ISSN 1313-1842, Vol. 3, Number 1, pp.188-191, Sofia, 2009.

[2] B. Nikolova, G. Nikolov. "Design, Development and Calibration of Virtual System for Relative Humidity Measurement. Sensors & Transducers Journal, ISSN 1726-5479, Vol.93, Issue 6. Pp. 1-14, June 2008.

[3] *Curve Fitting Toolbox™ 3 User's Guide.* The MathWorks, Inc., 2001-2010.

[4] F. Riley, P. Hobson, J. Bence, *Mathematical Methods for Physics and Engineering.* Cambridge University Press, New York, U.S.A., 2006.

[5] FIGARO. TGS 2600 – for the detection of Air Contaminants. Product Information.

[6] J. Fraden, *Handbook of Modern Sensors: Physics, Designs, and Applications.* Springer, San Diego, California, U.S.A., 2004.

[7] Multisim. Component Reference Guide. National Instruments Corporation, January 2007.

[8] S. T. Karris, *Mathematics for Business, Science, and Technology with MATLAB® and Excel® Computations.* Orchard Publications, California, U.S.A., 2007.

[9] S. Weisberg, *Applied Linear Regression*, John Wiley & Sons, Inc., New Jersey, U.S.A., 2005.

Comparative Analysis of LCC Resonant DC-DC Converters

Nikolay D. Bankov[1], Georgi P. Terziyski[2] , Alexander S. Vuchev[3]

Abstracts: **Comparative analysis of LCC resonant DC-DC converters with inductive output filter has been carried out. The operation of the converters below and above their resonant frequency has been investigated. As a result from the analysis equations of the converters' output and regulating characteristics have been obtained.**

Keywords: **LCC resonant DC-DC converter**

I. INTRODUCTION

The resonant converters have found wide application in building powerful supply equipment. The converters with resonant tanks of third row [1] are most often preferred because they can work in the whole range of idle running voltage to short-circuit while maintaining the conditions for soft commutation of the steering keys. In [2] is discussed LCC resonant converter used for the construction of medical x-ray machines.

Harmonic analysis is often used for their theoretical investigation. In order to obtain results with acceptable accuracy during this process, the influence of only the first harmonics of the currents and voltages are taken into account [3], i. e. the "method of the first harmonic" is used.

The aim of the present work is to carry out an analysis of LCC resonant DC-DC converters with inductive output filters by the method of the first harmonic, to compare the resultant output and regulating characteristics of the LCC converters under consideration, thus defining their advantages and their drawbacks.

II. ANALYSIS OF THE CONVERTERS

The diagram of the converter under consideration is shown in fig. 1. It consists of an inverter (controllable switches $S_1 \div S_4$ with reverse diodes $D_1 \div D_4$), a resonant tank, a matching transformer (T_r), an inductive filter (L_F), and a load resistor (R_0).

Different configurations composed of LCC resonant tanks, each of which has been studied for the converter under consideration, are shown in fig. 2 [4].

[1]Nikolay D. Bankov is with the Faculty of Electrical Engineering and Electronic, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: nikolay_bankov@yahoo.com

[2]Georgi P. Terziyski is with the Faculty of Electrical Engineering and Electronic, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: georgi_terziyski@abv.bg

[3]Alexandar S. Vuchev is with the Faculty of Electrical Engineering and Electronic, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: avuchev@yahoo.com

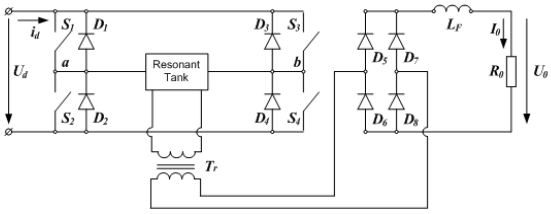


Fig.1. A resonant DC-DC converter with an inductive output filter

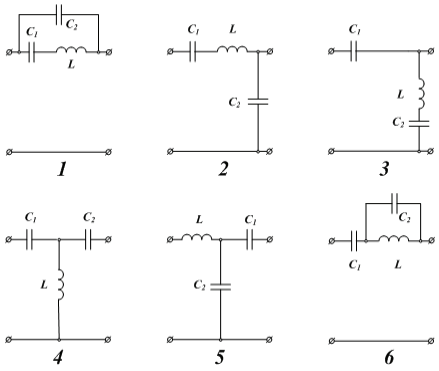


Fig.2. Configuration of LCC resonant tanks

For the purposes of the analysis it is assumed that all the elements in the diagram are ideal (no losses in them), the power devices switch from one state into another instantly, the matching transformer has a coefficient of transformation equal to unity, and the pulsations of the supplying U_d and the output voltages U_0 are negligibly small.

The following notations are accepted:

$U_0^{'} = U_0 / U_d$ - normalized output voltage;

$I_0^{'} = I_0 / (U_d / \rho_0)$ - normalized output current;

$R_0^{'} = R_0 / \rho_0 = U_0^{'} / I_0^{'}$ - normalized load parameter;

$\rho_0 = \sqrt{L / C_1}$ - wave resistance of the oscillating circuit;

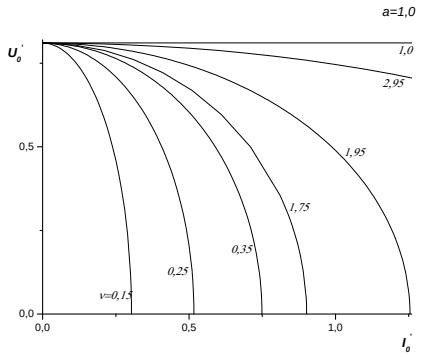
$\nu = \omega_s / \omega_0$ - frequency distortion of the oscillating circuit;

ω_s - operational frequency of the converter;

$\omega_0 = 1/\sqrt{LC_1}$ - resonant frequency of the oscillating circuit

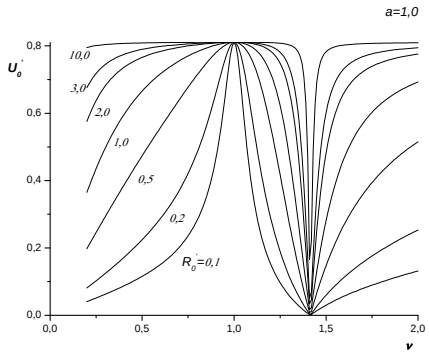
$a = C_2 / C_1$ - capacitance's ratio between the two capacitors in the oscillating circuit.

Families of similar output and regulating characteristics shown in fig.2 are obtained for the configurations of LCC resonant circuits at *a=1,0* and at different values of the frequency distortion ν and the normalized load parameter $R_0^{'}$. The latter are shown in fig. 3÷14. The equations of the output and regulating characteristics are given as well.



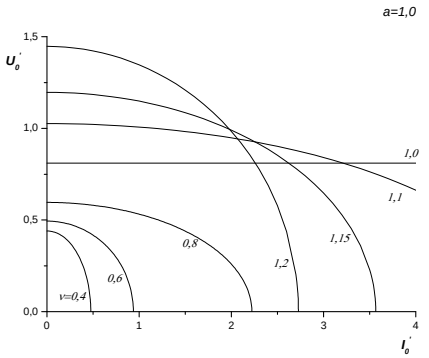
$$U_0^{'} = \frac{8}{\pi^2}.\sqrt{\frac{(\nu+\nu.a-\nu^3.a)^2-(\nu^2-1)^2.I_0^{'2}}{(\nu+\nu.a-\nu^3.a)^2}} \qquad (1)$$

Fig.3. Output characteristics for configuration №1 from Fig.2



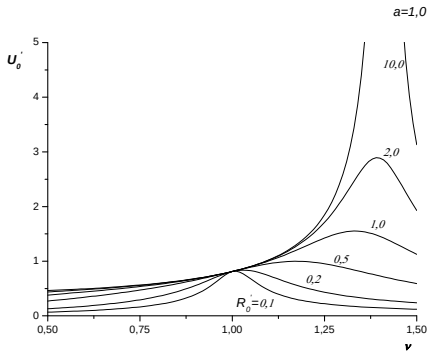
$$U_0^{'} = \sqrt{\frac{64.R_0^{'2}.(\nu+\nu.a-\nu^3.a)^2}{\pi^4.R_0^{'2}.(\nu+\nu.a-\nu^3.a)^2+64.(\nu^2-1)^2}} \qquad (2)$$

Fig.4. Regulating characteristics for configuration №1 from Fig.2



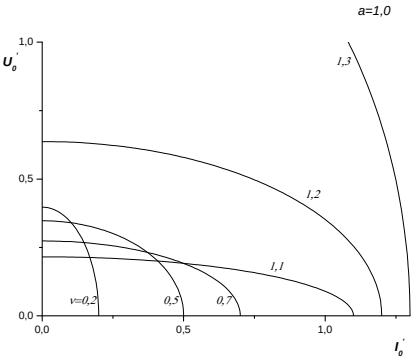
$$U_0^{'} = \frac{8}{\pi^2}.\sqrt{\frac{\nu^2-(\nu^2-1)^2.I_0^{'2}}{(-a.\nu^3+a.\nu+\nu)^2}} \qquad (3)$$

Fig.5. Output characteristics for configuration №2 from Fig.2



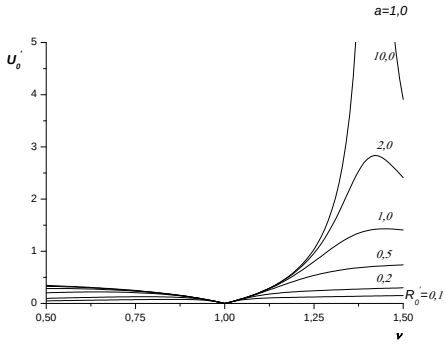
$$U_0^{'} = \frac{8.\nu.R_0^{'}}{\sqrt{64.(\nu^2-1)^2+\pi^4.(-a.\nu^3+a.\nu+\nu)^2.R_0^{'2}}} \qquad (4)$$

Fig.6. Regulating characteristics for configuration №2 from Fig.2



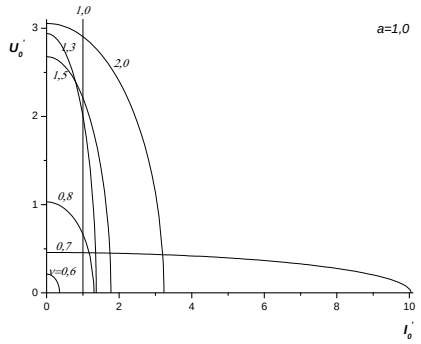
$$U_0^{'} = \frac{8}{\pi^2}.\sqrt{\frac{(\nu^2.a^2-I_0^{'2}).(\nu^2-1)^2}{(\nu^3.a-\nu.a-\nu)^2}} \qquad (5)$$

Fig.7. Output characteristics for configuration №3 from fig.2



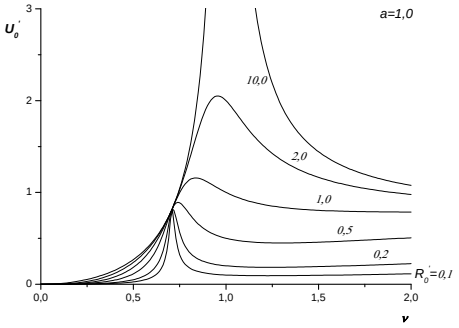
$$U_0^{'} = \frac{8.\nu.a.(\nu^2-1).R_0^{'}}{\sqrt{64.(\nu^2-1)^2+\pi^4.R_0^{'2}.(\nu^3.a-\nu.a-\nu)^2}} \qquad (6)$$

Fig.8. Regulating characteristics for configuration №3 from fig.2



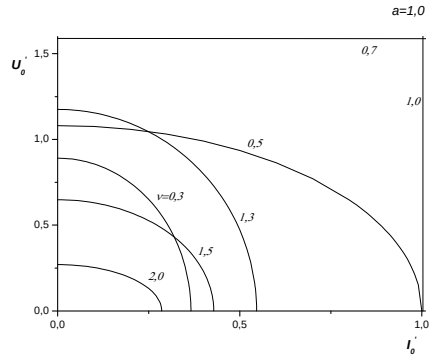
$$U_0^{'} = \sqrt{\frac{\nu^9.a^2 - (\nu^2.a + \nu^2 - 1)^2.I_0^{'2}}{(\pi^4/64).(\nu^3.a - \nu.a)^2}} \qquad (7)$$

Fig.9. Output characteristics for configuration №4 from fig.2



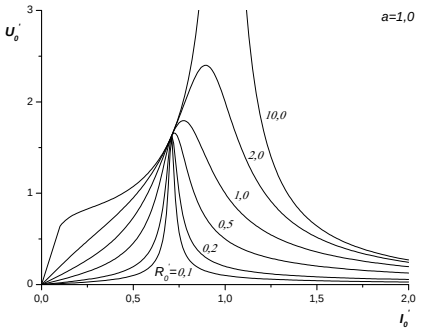
$$U_0^{'} = \frac{8.\nu^3.a.R_0^{'}}{\sqrt{64.(\nu^2.a + \nu^2 - 1)^2 + \pi^4.R_0^{'2}.(\nu^3.a - \nu.a)^2}} \qquad (8)$$

Fig.10. Regulating characteristics for configuration №4 from fig.2



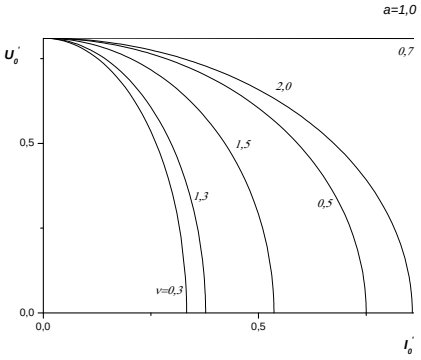
$$U_0^{'} = \sqrt{\frac{\nu^2 - (\nu^2.a + \nu^2 - 1)^2.I_0^{'2}}{(\pi^4/64).(\nu^3.a - \nu)^2}} \qquad (9)$$

Fig.11. Output characteristics for configuration №5 from fig.2



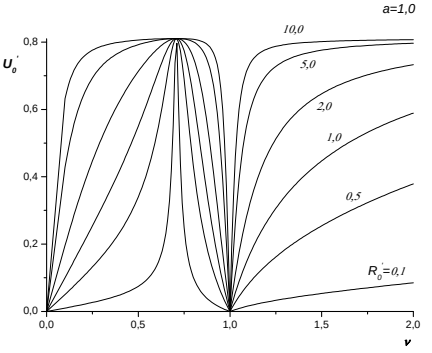
$$U_0^{'} = \frac{8.\nu.R_0^{'}}{\sqrt{\pi^4.R_0^{'2}.(\nu^3.a - \nu)^2 + 64.(\nu^2.a + \nu^2 - 1)^2}} \qquad (10)$$

Fig.12. Regulating characteristics for configuration №5 from fig.2



$$U_0^{'} = \frac{8}{\pi^2}.\sqrt{\frac{(\nu - \nu^3.a)^2 - (1 - \nu^2.a - \nu^2)^2.I_0^{'2}}{(\nu - \nu^3.a)^2}} \qquad (11)$$

Fig.13. Output characteristics for configuration №6 from fig.2



$$U_0^{'} = \sqrt{\frac{64.(\nu - \nu^3.a)^2.R_0^{'2}}{\pi^4.R_0^{'2}.(\nu - \nu^3.a)^2 + 64.(1 - \nu^2.a - \nu^2)^2}} \qquad (12)$$

Fig.14. Regulating characteristics for configuration №6 from fig.2

The output characteristics shown in fig. 3 are arranged concentrically and represent that, when the converter operates above and below its resonant frequency, with the increase of the operating frequency with $\nu = 0{,}15 \div 0{,}35$ and $\nu = 1{,}75 \div 2{,}95$, the short circuit current increases but the idle running voltage remains at a constant value. These characteristics are inherent to the voltage source, limited in current.

The output characteristics presented in fig. 5 show that, when the converter operates below its resonant frequency, with the increase of the operating frequency with $\nu = 0{,}4 \div 0{,}8$, the short circuit current and the idle running voltage increase. These are characteristics arranged concentrically and are innate to the current source of stable operation, even with short circuit.

When the converter operates above its resonant frequency, with the increase of the operating frequency to $\nu = 1{,}1 \div 1{,}2$, the short circuit current decreases but the idle running voltage increases. These are characteristics which intersect and are inherent to the voltage source, restricted in current. The output characteristics displayed in Fig.7 represent that, when the converter operates below its resonant frequency, with the increase of the operating frequency to $\nu = 0{,}2 \div 0{,}7$ the short circuit current increases and the idle running voltage decreases. These are characteristics which intersect and are inherent to the voltage source, limited in current.

When the converter operates above its resonant frequency, with the increase of the operating frequency to $\nu = 1{,}1 \div 1{,}3$, the short circuit current and the idle running voltage increase. These are characteristics that are concentric and are inherent to the current source of stable operation, even with short circuit.

The output characteristics given in fig. 9 manifest that, when the converter operates below its resonant frequency, with the increase of the operating frequency to $\nu = 0{,}6 \div 0{,}7$, the short circuit current and the idle running voltage increase, and at $\nu = 0{,}7 \div 0{,}8$ the short circuit current decreases, but the idle running voltage increases. These are characteristics that are concentric and are inherent to the current source of stable operation, even with short circuit.

When the converter operates above its resonant frequency, with the increase of the operating frequency to $\nu = 1{,}3 \div 1{,}5$, the short circuit current increases, while the idle running voltage diminishes, at $\nu = 1{,}5 \div 2{,}0$ the short circuit current and the idle running voltage increase. These are characteristics which are innate to the voltage source, limited in current. The output characteristics fixed in fig.11 and fig.13 are peculiar to the voltage source, restricted in current. It can be noted that all output characteristics are represented graphically by arcs from ellipses, while all regulating characteristics have a clearly defined maximum, whose value increases with increasing the value of the load resistor. The location of the maximum is displaced at the same time. It can be noted as well, that some of the obtained output and regulating characteristics are similar. For example, the ones, given in fig. 3÷4 are similar with the ones in fig. 13÷14. It can be seen from the graphs of the output characteristics (fig.3 and fig.13) that together with the increase in the operating frequency, the short circuit current also increases, while the idle running voltage stays the same. From the graphs of the regulating characteristics it is obvious (fig. 4 and fig. 14) that certain change in the output voltage is achieved by relatively small change in the operating frequency. Together with the other regulating characteristics (fig. 6, 8, 10 and 12) they show that with the increase of the value of $R_0^{'}$, the idle running voltage and the operating frequency also increase.

An area of operation can be noticed in some of the output characteristics, where the output voltage becomes higher than the supplying one, i.e., $U_0^{'} > 1$. It can be seen from figures 5,7,9, and 11.

III. Conclusion

Comparative analysis of LCC resonant DC-DC converters with inductive output filter has been carried out by the method of the first harmonic. Their operation below and above the operating frequency has been investigated.

As a result of the analysis are obtained output and regulating characteristics for the configurations of LCC resonant tanks. The output characteristics represent that on definite conditions, the converter can operate as a source of voltage limited in current or current source of stable operation, even with short circuit. The regulating characteristics manifest at what degree of increase in the value of the load resistor, there is also increase of the operating frequency of the discussed converter.

It has been established that the output voltage of the converter can have a higher value than the one of the supplying voltage.

The results from this investigation could be applied to designing LCC converters used as supplying devices of electrical arc welding aggregates, luminescent lamps, lasers etc.

IV. References

[1] Batarseh I., "Resonant Converter Topologies with Three and Four Energy Storage Elements", IEEE Transactions on Power Electronics, Vol. 9, No. 1, January 1994 pp. 64-73.

[2] Cavalcante, F. High Output Voltage Series-Parallel Resonant DC-DC Converter For Medical X-ray Imaging Applications, Dissertation, 2006.

[3] De Simone S., C. Adragna, C. Spini, G. Gatavari. Design-Oriented Steady State Analysis of LLC Resonant Converters Based on FHA. IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion, Speedam 2006 pp. S41-16 – S41-23.

[4] Yang, B. Topology Investigation for Front End DC/DC Power Conversion for Distributed Power System, Virginia Polytechnic Institute, Blacksburg, September 2003.

Overview of Automotive Network Protocols

Orlin Stoyanov[1], Georgi Krastev[2], Aleksandar Stoyanov[3]

Abstract – **This paper describes some of the different network protocols used primarily in the automotive industry but also throughout the aeronautical and various industrial markets. The first paragraph is dedicated to the Controller Area Network (CAN), the system which is often acting, as the "backbone" of a set of other networks, and intends to familiarize the reader briefly with the current state of the CAN protocol. The second paragraph begins with a detailed description of the LIN bus, generally described as a sub-type of CAN. At the end, several other protocols operating under X-by-Wire application fitted in motor vehicles are considered.**

Keywords – **Automotive control, High Speed Protocol, Vehicle network.**

I. INTRODUCTION

During the last few years there is an observed increase in the number of electronic systems in the automobiles. The large quantity of sensors built in the automobiles requires each new function to be controlled by ECUs. The field of multiplexed buses is constantly being enlarged in line with the car manufacturers which are making developments in the order to improve the safety, comfort and reliability. With the increasing number of electrical equipment in the automobile the need of efficient networking system is rising too. The use of such in-vehicle network technologies is expected to reach new frontiers in the near future.

Some of the important in-car technologies and protocols divided into two general classes, CSMA (Carrier Sense Multiple Access) and TDMA (Time division multiple access), are described as follows:

II. IN-VEHICLE NETWORK PROTOCOLS

Prior to begin brief introduction of this protocols, let's have a general overview to some of them. Fig. 1 illustrates the performance/cost ratio and the relative importance and uses of each protocol. It is not surprising that the faster the system, the higher the price to be paid. Of course, the relative costs shown on the horizontal axis are given for guidance only, with the usual reservations.

A modern vehicle includes a large number of intersection and interlacing layers of links for connecting multiple electronic control units, on-board systems and entertainment applications. To take more accurate idea of the situation, the diagram on the Fig. 2 shows architecture of a high class vehicle.

The CAN bus (Controller Area Networking) was defined in the late 1980 by Bosch, initially for use in automotive applications, but which is coming into use for linking distributed controllers and sensors in other fields. The official CAN specification has been released by ISO as 11898-1 (CAN data link layer protocol) [1].

CAN is a CSMA/CD protocol. It supports speeds of up to 1Mb/s so is an SAE class C protocol, suitable for real time control applications.

The CAN bus is a broadcast type of bus. This means that all nodes can receive the transmitted messages. They are not addressed to intended recipients, but message itself include the sender's identifier. This provides local filtering so that each node may react only on the messages with the correct identifier. Messages' identifiers give the priority of the message, so the priority of messages is decided at the design stage. There are two standards for CAN 2.0, called A and B. Part A describes the most common "standard" CAN frame. This frame supports 11 identifier bits. Part B is intended to describe the CAN frame in its "extended format". Because it was insufficient for some applications, the identifier value was changed from 11 to 29 identifier bits. The CAN 2.0A and 2.0B versions were designed to provide compatibility with any earlier version of the protocol. In full CAN, the CAN devices add filtration of the messages, and will only pass messages with specified identifiers on to its associated controller, so a controller is only interrupted by those messages the CAN terminal passes, that is those of interest to that controller [7,9].

The bus uses non-return to zero (NRZ) coding with bit-stuffing. The modules are wired to the bus: if just one node is driving the bus to a logical 0, then the whole bus is in that state regardless of the number of nodes transmitting a logical 1.

The CAN standard defines four different message types. The messages use a bit-wise arbitration to control access to the bus, and each message is tagged with a priority.

The scope of the CANbus protocol covers the physical and data link layers of the ISO/OSI model [9].

[1]Orlin Stoyanov,Faculty of Electrical and Electronic Engineering and Automation, Studentska St, 7017, Ruse, Bulgaria, E-mail: ostoyanov@uni-ruse.bg.

[2]Georgi Krastev,Faculty of Electrical and Electronic Engineering and Automation, Studentska St, 7017, Ruse, Bulgaria, E-mail: GKrastev@ecs.ru.acad.bg.

[3]Aleksandar Stoyanov, Faculty of Automotive and Transport Engineering, Studentska St, 7017, Ruse, Bulgaria,

E-mail:astoyanov@uni-ruse.bg

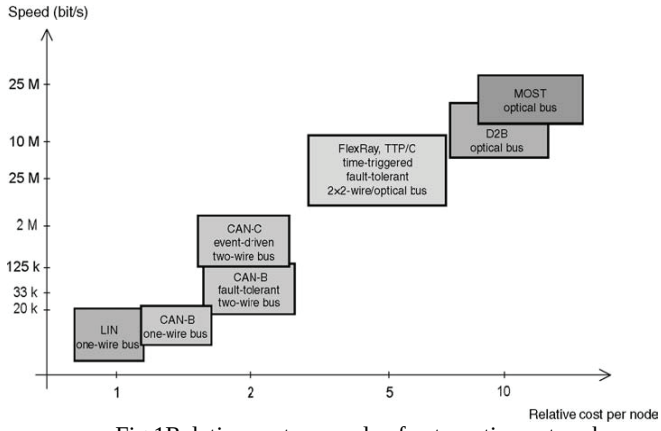


Fig.1Relative cost per node of automotive networks

LIN stands for Local Interconnect Network and is a low cost field bus. It is mainly intended to support the control of "mechatronic" elements found in distributed systems for motor vehicle applications, but of course it can be applied in many other fields. The LIN protocol concept is a multiplexed communication system whose level and associated performances fit below CAN's functionality. The primary and original purpose of LIN is therefore to provide a "sub-division" for CAN, with reduced functionality and lower costs.

The major developer of the LIN concept was the Motorola Company. A consortium was created, in March 2000, including the car manufacturers Audi, BMW, Daimler Chrysler, Volkswagen and Volvo Car Corporation, as well as Motorola Inc. and Volcano Communication Technologies AB.

LIN has single master/multiple slave architecture, therefore no need for arbitration. Speed is 20Kbit/s and it is specified to be most appropriate for SAE class A applications, the speed is actually at the lower end of class B. As it is time triggered, message latency is guaranteed. [7,8].

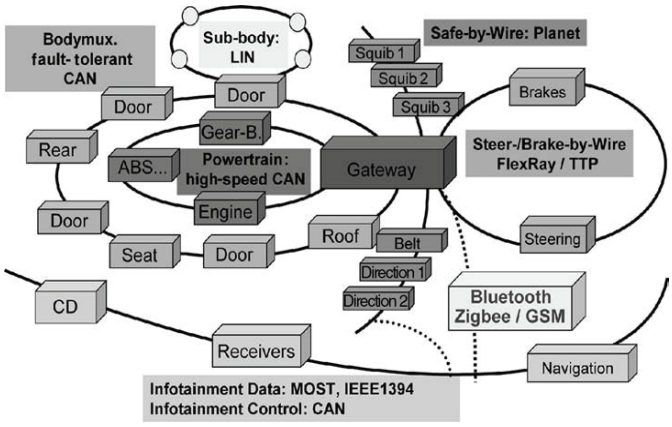


Fig.2 Conventional architecture of a high class vehicle

The increasing complexity of onboard systems arouses a demand for protocols providing "real-time" responses, deterministic solutions and high security. TTCAN (time-triggered CAN) was proposed by the CAN in Automation (CiA) group and the Bosch company at the beginning of 2002, which allowed CAN to be used for time triggered messages and increased determinism, reliability and synchronization over CAN.

TTCAN is a protocol layer at a higher level than CAN itself, without any modification of the data link layers (DLL) and physical layers (PL) of the latter. TTCAN is located primarily in the session layer of the OSI/ISO (International Standardization Organization) model. The aim of TTCAN is to keep the latency of each message to a specified value, independently of the load on the CAN bus itself. This protocol can be implemented on two levels:

- level 1 is limited to the transfer of cyclic messages;
- level 2 supports what is known as a 'global time' system.

TTCAN is based on a deterministic temporal exchange, based on a time window of a predetermined operating cycle, whose global operation can be represented in the matrix of rows and columns, which summarizes the general operating principle of this protocol.

All the messages travelling in the network between the CAN nodes are organized as elements of an X_Y matrix. This matrix time system consists of time windows organized in "basic cycles", identical time values (represented as the totality of each row of the matrix) and numerous time intervals (windows) during which transmissions are authorized (represented by the columns of the matrix). Thus, it defines the relationship between the time windows and the presence of messages in the network.

The TTCAN operating principle is based on the fact that one of the nodes of the network is responsible for organizing the time division and allocation involved. This is because, when the system starts up, one of the nodes allocates the reserved time phases to each of the others. The system thus becomes deterministic, as each node has the right to transmit at a precise moment known to it, for a closely specified period. Clearly, this does not constitute a real-time system at all, but if the complete cycle is executed quickly enough, there is a rapid return to the same node, and to all the participants this appears to be "real-time" network access [2,7].

The TTP/C system is a member of the family of time-triggered protocols. The "/C" indicates that it meets the criteria of class C of the SAE – Society of Automotive Engineers – for the real-time and fault-tolerant aspects of communication in the car industry. It was developed by Professor Hermann Kopetz of the Vienna University of Technology, Austria and subsequently adopted by the TTTech – Time-Triggered Technology Company. It was not originally intended for motor vehicle applications, but for industrial applications in general. TTPTM/C is designed on the principle of TDMA (time division multiple access) to the medium.

All activities are carried out at certain points in time, decided at system design time, rather than network activities being triggered by external events, as in a CSMA protocol such as CAN. As TTP is a TDMA protocol, latency is deterministic. There is a bus guardian that "guarantees" that no node can control communication media outside its transmission slot. This principle can resolve problems of interoperability between CPUs developed independently of each other. [4,7,10].

Volcano might be described as "TTP on CAN" and the Volcano web-site describes the protocol as CAN-based and

deterministic. The protocol is used by Volvo on the S80 and V70 cars, and is coming into use on Volvo buses. According to the Volcano Communications Concept, Volcano appears to be a technique in which the CAN network is integrated in such a way as to guarantee the latency of all the messages. It does this by specifying the latency and periodicity of messages at design time. This allows the maximum latencies to be calculated, so the system designer can specify the network set up in such a way as to modify these specifications to guarantee the specified parameters, by avoiding arbitration as far as possible. This seems to imply that the sending of network communication is time triggered rather than event triggered, so the description "TTP on CAN" seems a pretty good summing up.

This apparently means that network loadings can be considerably higher than using CAN conventionally, maybe 60% loading, whereas for latency of lower priority messages to be contained to reasonable limits, CAN loading may need to be around 10%.

FlexRay originated from the formation of a group of companies which conduct a profound technical analysis of the existing networks used in the automobile industry, namely CAN, TTCAN, TCN, TTP/C and Byteflight. The purpose of the analysis was to discover whether any one of them was capable of meeting all the technical and application requirements stated above. It was concluded that this study clearly showed that there is a lot more to be developed in this specific area, leading to the development of a new proposal, called 'FlexRay'. The findings concerning the existing solutions are summarized below:

FlexRay was designed to provide a communication system in which collisions for access to the medium are impossible; in other words, the nodes are not subject to arbitration on the transmission channel, and collisions should not occur in normal operation. However, collisions may arise during the starting phase of the protocol on the transmission medium. The physical layer does not provide any means of resolving these collisions, and therefore the application layer must take over to handle such problems.

In order to provide the system with the greatest flexibility of application, it is necessary to communicate at a exact known instant for a known length of maximum time (operate in "real time") with the assurance of being the only station present at that moment on the physical communication medium, making collisions not possible; to allow communication at a variable bit rate if required, and thus to require an unspecified communication time.

Communication in FlexRay takes place with the aid of recurring communication cycles. Each includes a 'static segment, a 'dynamic segment', an 'optional symbol window' and a phase in which the network is in idle mode, which is called the network idle time (NIT). This cycle is initialized by the network manager node.

III. Conclusion

The need of high technology automobile networks is critical considering the growing dependence of the automobiles on the smooth functioning of the electronics. The car manufacturers and suppliers are working together with the aim to standardize the in-vehicle network protocols. The growing demand for safety, comfort, reliability and entertainment requires incorporation of multiple protocols. This tendency is expected to continue with the manufacturers' implementation of more and more electrical devices in the automobile with the purpose of increasing the value of their products and meeting the requirements of the changing industry.

References

[1] CAN in Automation (CiA): CAN protocol specification http://www.can-cia.org/index.php?id=164

[2] CAN in Automation (CiA): Time-Triggered CAN http://www.can-cia.org/index.php?id=166

[3] FlexRay specification: http://www.flexray.com/

[4] Kopetz H., Real-Time Systems: Design Principles for Distributed Embedded Applications. Kluwer Academic Publishers, Boston, 1997

[5] Lawrenz, W. CAN System Engineering: From Theory to Practical Applications., 1997.

[6] Navet, N., F. Simonot-Lion. Automotive Embedded Systems Handbook. Taylor & Francis Group, LLC, 2009.

[7] Paret, Dominique. Multiplexed Networks for Embedded Systems. CAN, LIN, Flexray, Safe-by-Wire. John Wiley & Sons, Ltd, 2006.

[8] Rey, S. Introduction to LIN (Local Interconnect Network). Revision 1.0, 2003.

[9] The can protocol tourhttp://www.kvaser.com/can/protocol/index.htm

[10] TTTech website, http://www.tttech.com

A Matlab/Simulink Model of Piezoceramic Ring for Transducer Design

Igor Jovanović[1], Dragan Mančić[2], Vesna Paunović[3], Milan Radmanović[4], Zoran Petrušić[5]

Abstract – **In this paper a Matlab/Simulink model of thickness polarized piezoceramic rings, which is based on previously realized 3D matrix model of ring, is realized. Based on this Matlab/Simulink model it is able to compute all the relations between the input applied voltage and the output forces and velocities on every external surface. In order to compare the computed and experimental results, the input impedances for a piezoceramic rings with different dimensions are calculated and measured. Also, using the realized model the resonance and antiresonance frequencies for these rings are calculated and then compared with experimental results to validate the model. The proposed Matlab/Simulink model requires simpler implementation than mathematical model.**

Keywords – **Piezoceramic ring, Matlab/Simulink model, Resonance frequency characteristics, Input impedance.**

I. INTRODUCTION

Piezoelectric ceramics rings of different thicknesses and inner/outer diameters, especially lead zirconate titanate (PZT) rings, are widely used as active components in Langevine ultrasonicsandwich transducers for industrialapplications. In power applications of ultrasonic vibrations, piezoceramic rings are employed because of their good high electromechanical conversion efficiency.

Up to now various methods are proposed to model piezoceramic rings. Several one-dimensional (1D) models have been proposed to describe the principle modes of vibration of the piezoceramic rings in the thickness-extensional[1] and the radial modes [2].A three-dimensional (3D) approach is needed for proper modeling of piezoelectric transducer constructions with comparable lateral and thickness dimensions, such as e.g. sandwich transducers based on thick piezoelectric ceramic rings.Several 3D models also have been developed to analyze the piezoceramic rings[3,4]. Most of these3D models were based on simplified structures under simple loading (boundary) conditions.

In the present worka 3D Matlab/Simulink model to preview the behavior of a piezoceramic ring of any dimensions is proposed. The aim is to provide a simple and useful tool for the sandwich transducer design and optimization. The model is able to describe the composite vibrations both in the thickness and in the radial directions, and the coupling with the load and the backing.The piezoceramic ring is described, in the frequency domain, by previously realized approximated 3D piezoceramicring model [3]. Using the Matlab/Simulink model, the input impedance, as well as resonance and antiresonance frequencies are calculated and then compared to the experimental results.

Autors are with the University of Nis, Faculty of Electronic Engineering, A.Medvedeva 14, 18000 Nis, Serbia, E-mail: [1]igor_j@elfak.rs; [2]dragan.mancic@elfak.ni.ac.rs; [3]vesna.paunovic@elfak.ni.ac.rs; [4]milan.radmanovic@elfak.ni.ac.rs; [5]zoran.petrusic@gmail.com.

II. DESCRIPTION OF GOVERNING EQUATIONS

The typical piezoceramic ring geometry with outer radius *a*, inner radius *b*, thickness 2*L*, and with completely metallized ring-shaped surfaces is shown in Fig. 1a. Every ring surface is loaded by acoustic impedance Z_i, where v_i and F_i are velocities and forces on those contour surfaces P_i (i=1,2,3,4).

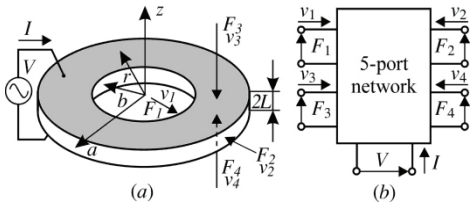


Fig. 1. Thickness-poled, electroded and loaded piezoceramic ring: (a)geometry and dimensions; (b) 5-port network representation

In paper [3],a 3D mathematical model is developed to describe the behavior of the thickness polarizedpiezoceramic ring.By means of this 3D model, the ring is modelled in the frequency domain as a five port system with one electrical and four mechanical ports, one for each external surface, by the following matrix form:

$$\begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ V \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{13} & z_{15} \\ z_{21} & z_{22} & z_{23} & z_{23} & z_{25} \\ z_{13} & z_{23} & z_{33} & z_{34} & z_{35} \\ z_{13} & z_{23} & z_{34} & z_{33} & z_{35} \\ z_{15} & z_{25} & z_{35} & z_{35} & z_{55} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ I \end{bmatrix} \quad (1)$$

Linear system of Eq. (1), which describes external behavior of the ring and relates the electrical (voltage *V* and current *I*) to the mechanical variables (forces F_i and velocities v_i) in frequency domain (Fig. 1b), is described in detail in paper [3]. The matrix elements are defined in Eq. (2), where $S=\pi(a^2-b^2)$ is ring area, $C_0=(\varepsilon_{33}^S S)/(2L)$ is the piezoceramic "clamped capacitance", c_{ij}^D are the elastic stiffness constants; ε_{33}^S is the clamped dielectric permittivity;$h_{ij}$are the piezoelectric tensor terms (i, j=1,2,3). $k_r=\omega/v_r$, $k_z=\omega/v_z$, $v_r=\sqrt{c_{11}^D/\rho}$ and $v_z=\sqrt{c_{33}^D/\rho}$ are the wave numbers and the phase velocities of the two uncoupled waves in the *r* and *z* directions respectively; ω is angular frequency; ρ is the piezoceramic density; J_1 and Y_1 are the first and the second kind of Bessel's functions of the first order.

$$z_{11}=\frac{-4\pi L}{j\omega}\left\{c_{12}^{D}-c_{11}^{D}\left[1-k_r b\left(A_1 J_0(k_r b)+B_1 Y_0(k_r b)\right)\right]\right\},$$

$$z_{22}=\frac{4\pi L}{j\omega}\left\{c_{12}^{D}-c_{11}^{D}\left[1+k_r a\left(A_2 J_0(k_r a)+B_2 Y_0(k_r a)\right)\right]\right\},$$

$$z_{12}=\frac{-4\pi k_r b L c_{11}^{D}}{j\omega}\left[A_2 J_0(k_r b)+B_2 Y_0(k_r b)\right],\quad z_{13}=\frac{2\pi b c_{13}^{D}}{j\omega},$$

$$z_{21}=\frac{-4\pi k_r a L c_{11}^{D}}{j\omega}\left[A_1 J_0(k_r a)+B_1 Y_0(k_r a)\right], \qquad (2)$$

$$z_{15}=\frac{4\pi b L h_{31}}{j\omega S},\ z_{23}=\frac{2\pi a c_{13}^{D}}{j\omega},\ z_{25}=\frac{4\pi a L h_{31}}{j\omega S},\ z_{35}=\frac{h_{33}}{j\omega},$$

$$z_{33}=\frac{c_{33}^{D}k_z S}{j\omega\,\mathbf{tan}(2k_z L)},\quad z_{34}=\frac{c_{33}^{D}k_z S}{j\omega\,\mathbf{sin}(2k_z L)},\quad z_{55}=\frac{1}{j\omega C_0}.$$

The constants in Eq. (2) are defined as:

$$A_1=\frac{Y_1(k_r a)}{j\omega\left[J_1(k_r b)Y_1(k_r a)-J_1(k_r a)Y_1(k_r b)\right]},$$

$$A_2=\frac{Y_1(k_r b)}{j\omega\left[J_1(k_r b)Y_1(k_r a)-J_1(k_r a)Y_1(k_r b)\right]},$$

$$B_1=\frac{J_1(k_r a)}{j\omega\left[J_1(k_r a)Y_1(k_r b)-J_1(k_r b)Y_1(k_r a)\right]}, \qquad (3)$$

$$B_2=\frac{J_1(k_r b)}{j\omega\left[J_1(k_r a)Y_1(k_r b)-J_1(k_r b)Y_1(k_r a)\right]}.$$

With this model, external behavior of the ring and all the transfer functions of the ring can be easily computed, taking into account the interaction with the surrounding mediaand the coupling between the thickness (T) and radial (R) modes.

III. MATLAB/SIMULINK MODEL

This section presents how the mathematical model of the piezoceramic ring, described in previous section, is implemented in Matlab/Simulink.

Matlab/Simulink is interactive software which has been used recently as design and development environment for model implementation in various areas of engineering and scientific applications[5].The graphical representation of models in Matlab/Simulink is based on block communication diagrams. Internally, the model is split into smaller separate functions blocks.

In order to demonstrate the advantages of this process, the Matlab/Simulink model of a piezoceramic ring is developed.At this point, it is possible to encapsulate the whole model in a Simulink blocks.A general schematic of this model is presented in Fig. 2.

Eqs. (1), (2) and (3) fully describe the model of the piezoceramic ring, which has been used for the simulation.

Realized model of a piezoceramic ring consist of two main blocks (Fig. 2). The first block gives all elements of the matrix in Eq. (1) as well as all the required coefficients a customized set of Eq. (4), which has been obtained through a series of simple mathematical operations. Calculating of these elements is based on entered characteristics and dimensions of used ceramic.

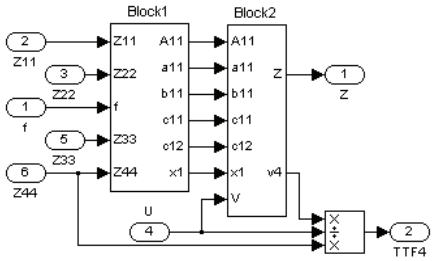


Fig. 2. Matlab/Simulink model of the piezoceramic ring

$$\begin{bmatrix}0\\0\\0\\0\\E_5\end{bmatrix}=\begin{bmatrix}A_{11}&A_{12}&A_{13}&A_{14}&A_{15}\\0&a_{11}&a_{12}&a_{13}&a_{14}\\0&0&b_{11}&b_{12}&b_{13}\\0&0&0&c_{11}&c_{12}\\0&0&0&c_{21}&c_{22}\end{bmatrix}\cdot\begin{bmatrix}v_1\\v_2\\v_3\\v_4\\I\end{bmatrix} \qquad (4)$$

where is $E_5 = -A_{11}\cdot a_{11}\cdot b_{11}\cdot V$ and:

$$\begin{aligned} a_{i,j} &= A_{1,j+1}\cdot A_{i+1,1} - A_{1,1}\cdot A_{i+1,j+1} \quad (i,j=1,2,3,4)\\ b_{i,j} &= a_{1,j+1}\cdot a_{i+1,1} - a_{1,1}\cdot a_{i+1,j+1} \quad (i,j=1,2,3)\\ c_{i,j} &= b_{1,j+1}\cdot b_{i+1,1} - b_{1,1}\cdot b_{i+1,j+1} \quad (i,j=1,2) \end{aligned} \qquad (5)$$

Since the calculation of coefficients in system of Eq. (4) requires model with a very complex structure, which includescomputing first and the second kind of Bessel's functions of the first order, the simplest way is to use embedded functions (*Embedded MATLAB Function block*). In this block,the functions that are not supported by Simulink, such as Bessel functions, is easy to declare with command *eml.extrinsic*.Also, since all the parameters to obtain these elements are previously defined, there is no risk of occurrence algebraic loops through this block.

The second part of the model use corresponding value obtained in the first block and solves the system of Eq. (4), and determines the input current I and input electrical impedance$Z=V/I$. The second block for model of a piezoceramic ring is shown in Fig. 3. The signal$x_1$is represented by the expression$x_1=c_{22}-c_{21}c_{12}/c_{11}$.

A brief overview of the internal structure of the others blocks is provided in Figs. 4 and 5. These figures contain the Simulink implementation of the mathematical models for determining of values for mechanical velocities on metalized ring-shaped surfaces (v_1 and v_2) and circular-curved surfaces (v_3 and v_4) of piezoceramic ring, respectively.

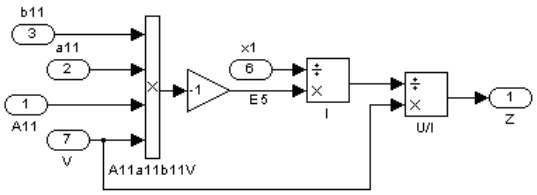


Fig. 3. Simulink subsystem fordetermining of input current$I$ by solving of equation system (4)

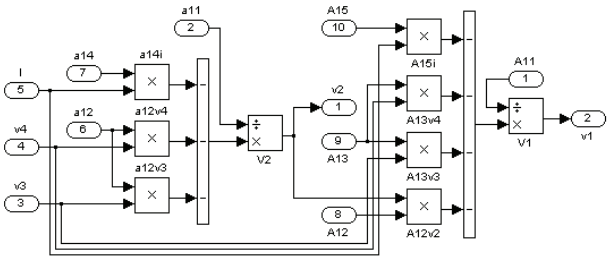


Fig. 4. Simulink subsystem for determining of values for mechanical velocities$v_1$andv_2, onmetalized ring-shapedsurfaces of PZT ring

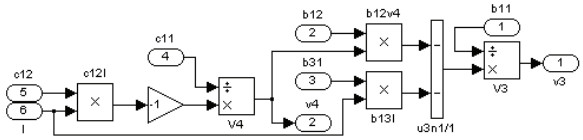


Fig. 5. Simulink subsystem for determining of values for mechanical velocities v_3 and v_4 oncircular-curved surfaces of PZT ring

Another powerful feature of the Simulink, called masking, is that it can simplify the use of the model by replacing many dialog boxes in a subsystem with a single dialog box. Instead of requiring the user of the model to open each block and enter piezoceramic parameter values, those parameter values can be entered on the mask dialog block and passed to the blocks in the masked subsystem. Fig. 6 illustrates how the mask dialog block for the piezoceramic ring looks like. The user has just to change the values of the parameters for different types of piezoceramic rings.

Loading the mechanical ports with the acoustical impedances of the surrounding media and by applying an *ac* voltage $V=V_0e^{j\omega t}$ to the electric port, it is possible to compute all the relations between the input applied voltage and the output forces and velocities on every external surface analytically, such as the electrical input impedance (V/I), the transmission (F_i/V) and the receiving (V/F_i) transfer functions.Our model can compute separate transfer functions for each external surface with arbitrary acoustic loads.

Thus development of the Matlab/Simulink model of the piezoceramic ring has been completed. The model provides numerous possibilities for the investigation ofpiezoceramic ringproperties. Verification of the created Matlab/Simulink model will be considered in the next section.

IV. Numerical and Experimental Results

The Matlab/Simulink model of the piezoceramic ring illustrated in Fig. 2, allows simulation of operation of the piezoceramic ring under different conditions. In order to obtain an experimental validation of the proposed model, the input electrical impedance of different piezoceramic rings are measured and compared with the computed results (if one assumes that the external medium is air).

Experimental input impedance curves of piezoceramic samples were measured with the frequency-sweeping apparatus (HP 4194A Network Impedance Analyzer). Seven samples of commercial PZT4 rings (Fig. 7) have been characterised. The PZT4 piezoceramic rings dimensions are given in Table I.

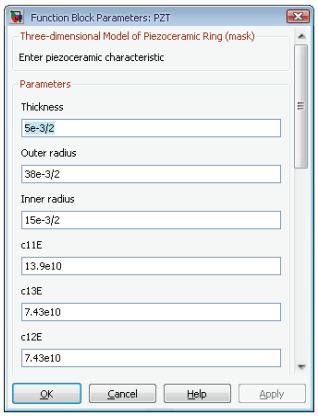


Fig. 6. Mask dialog block for the enter piezoceramic ring parameters

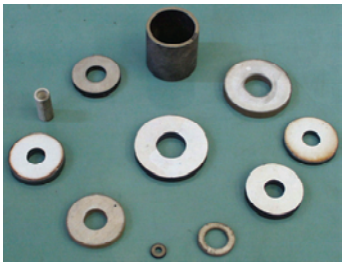


Fig. 7. Measured piezoceramic ring samples

TABLE I
PIEZOCERAMIC RINGS DIMENSIONS

Sample	1	2	3	4	5	6	7
2*a* (mm)	38	38	50	24	38	38	10
2*b* (mm)	15	13	20	15	13	13	4
2*L* (mm)	5	6.35	6.35	3	4	6	2

Fig. 8 shows a comparison between the measured and computed input electrical impedance for a 4th and 6th samplesof PZT4 rings. Computed results were carried out using PZT4 piezoceramic material constants [6].

As it can be seen, the form of the impedance curves is in accordance with experimental ones. According to piezoceramic rings dimensions, one of these vibrational modes is related to thickness mode (T) and the others are related to radial modes (R). Our model predicts with sufficient accuracy the first radial (R_1) and first thickness (T_1) modes, which are the mostly used in practical applications.

The resonance (f_p) and antiresonance (f_s) frequencies of piezoceramic rings are calculated with high accuracy. f_p is the frequency at which the electrical impedance of the ring reaches its minimum, and f_s is the frequency of minimum admittance. Using the proposed model, good agreement between simulated and experimental results for resonance and antiresonance frequencies is observed (Fig. 9). In this simulation up to four vibrational modes are presented for all piezoceramic rings.

As it is possible to see, the model is able to predict resonance and antiresonance frequencies with good accuracy both for the first radial mode, and for the first thickness mode. Because only these two modes are of relevance in the practical applications of piezoceramic rings as ultrasonic transducers, the model can be used as a simple and useful tool in transducer design and optimization. Difference between resonance and antiresonance frequencies obtained from simulation compared with experimental results in certain vibrational modes are occurrencesbecause this model includes only the thickness and radial resonant modes.

Results shown in Figs. 8 and 9 should not be used for very fine comparisons of measured and theoretical results, by using only typical values for the material constants [6]. The computed results can be improved by fitting the constants of the piezoceramic material. Further, some measured modes are

not predicted by the model, probably because they are shear modes. However, general trends can be observed.

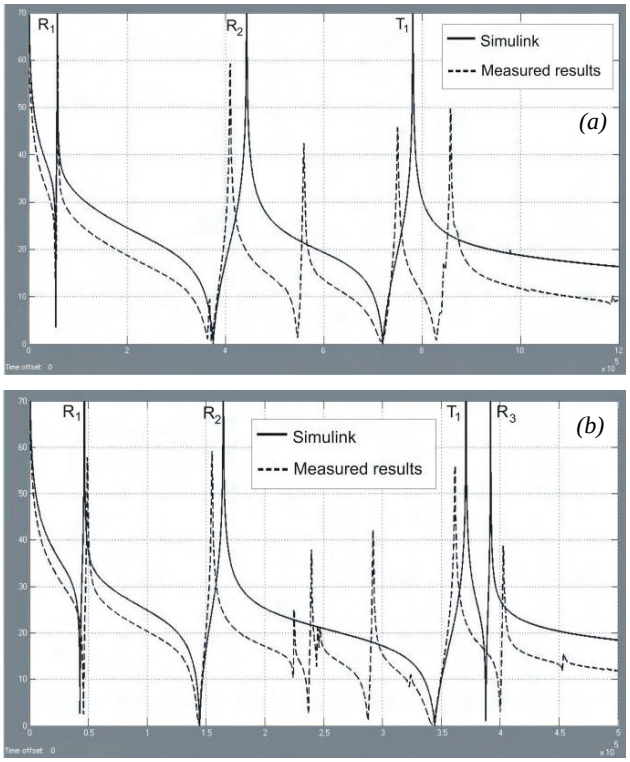


Fig. 8. Simulated and experimental input electrical impedance versus frequency for the4th sample (a)and 6th sample(b)

V. Conclusion

In this paper an accurate piezoceramic ring model valid to any diameter to thickness ratio is realized and demonstrated in Matlab/Simulink for a typical PZT rings. This model taking the interaction with the surrounding media into account is able to compute all the ring transfer functions, such as the input electrical impedance.External behavior of the piezoceramic ring is described in frequency domain by a system with four mechanical ports (one for each external surface) and one electrical port.

The piezoceramic rings with different dimensions are analyzed using the developed model. The comparison experimental and theoretical results are quite good and validate the new design approach. Firstly, the electrical input impedance of samples with different dimensions was computed by the model and compared with experimental results, obtaining a good agreement. After that, the resonance and antiresonance frequencies are calculated using the model and then compared to the experimental results. Such comparison also shows satisfactory agreement.

The Matlab/Simulink model gives a very good prediction of the piezoceramic ring behavior, although only the electrical impedance is studied when the ring is without outer load. It can be applied successfully to design and optimize ultrasonic sandwich transducers for industrial applications.In a future work our aim is to improve the performance of the model in order to obtain a reliable tool for more complex Langevine ultrasonic sandwich transducer design.

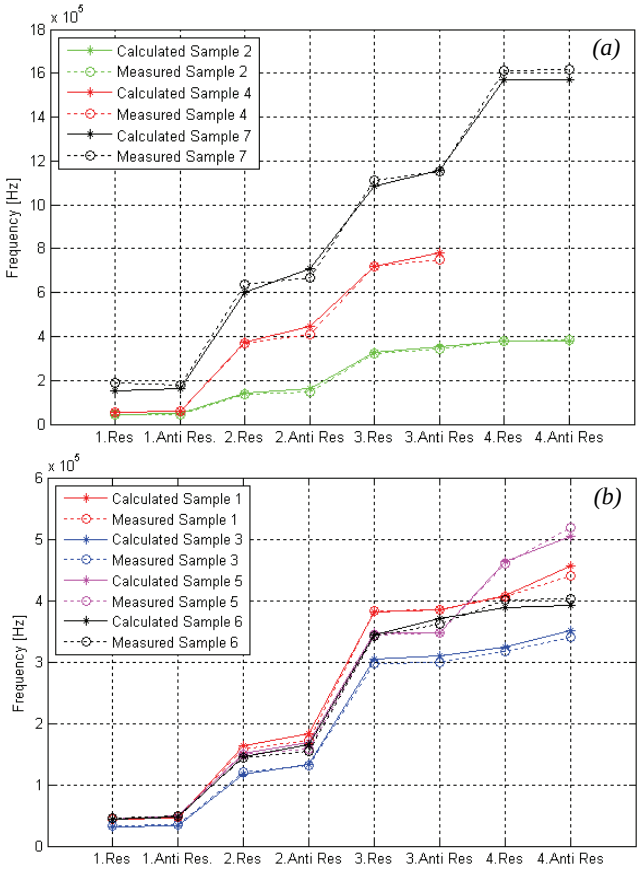


Fig. 9. Comparison between calculated and experimental resonance and antiresonance frequencies for the samples 2, 4 and 7 (a) and samples 1, 3, 5 and 6 (b)

Slightly modified, the realized model is also applicable to piezoceramic disks. Therefore, it is a useful tool for transducer manufacturers and material scientists.

Acknowledgement

This work has been funded by the Serbian Ministry for Science under the projects TR33035, III43014 and OI172057.

References

[1] W.P.Mason, *Electromechanical Transducers and Wave Filters*,Princeton: Van Nostrand, 1948.

[2] A.Iula, N.Lamberti, M.Pappalardo, “A Model for the Theoretical Characterization of Thin Piezoceramic Rings”, IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 43, no. 3, pp. 370-375, May 1996.

[3] D.Mančić, M.Radmanović, “Piezoceramic ring loaded on each face: a three-dimensional approach”,Electronic Journal “Technical acoustics”, vol. 2, pp. 1.1-1.7, 2002.

[4] H.Jalili, H.Goudarzi, “Modeling the hollow cylindrical piezoceramicwith axial polarization using equivalent electro-mechanical admittance matrix”,Sensors and Actuators A: Physical, vol. 149, pp. 266-276, 2009.

[5] M. R. Hatch, *Vibration Simulation Using Matlab and Ansys*,Boca Raton, Chapman & Hall/CRC, 2001.

[6] *Five piezoelectric ceramics*, Bulletin 66011/F, Vernitron Ltd., 1976.

Single-Circuit and Double-Circuit Regulating Apparatus for Gas Discharge Element

Stefan Barudov[1],Rositsa Dimitrova[2], Milena Ivanova[3]

Abstract –**Most purposefully, the usage of electrical discharge is related to its development in a specific, artificially created medium – a discharge element.Due to the specifics of the discharge elements, they are connected as a load to the power supply grid by regulating apparatuses.The work is dedicated to a comparative study between single-circuit and double-circuit regulating apparatuses for control of a direct current discharge with respect to the stability of the discharge current and the power parameters of the regulating apparatuses.**

Keywords – **DC discharge, single-circuit regulating apparatus, double-circuit regulating apparatus**

I. INTRODUCTION

The single-circuit regulating apparatus (SCRA) consists of an uncontrollable rectifier (a dc voltage source with magnitude $U^*_{D\max}$ – Fig.1), which supplies a regulating element (RE) connected in series with a compensation stabilizer (CS) and a discharge element (DE). $U = U^*_{D\max}$ - $U^{'}_{D}$ provides stabilization of the discharge current at a change of the amplitude of the input supply voltage.

The double-circuit regulating apparatus (DCRA) consists of a controllable rectifier which controls the voltage over the regulating element of the compensation stabilizer.

Most often, in operating mode the controllable rectifier realizes the law U_{RE}=const or switches on a regulator of the input supply voltage, which limits the fluctuation of the amplitude of the input supply voltage.

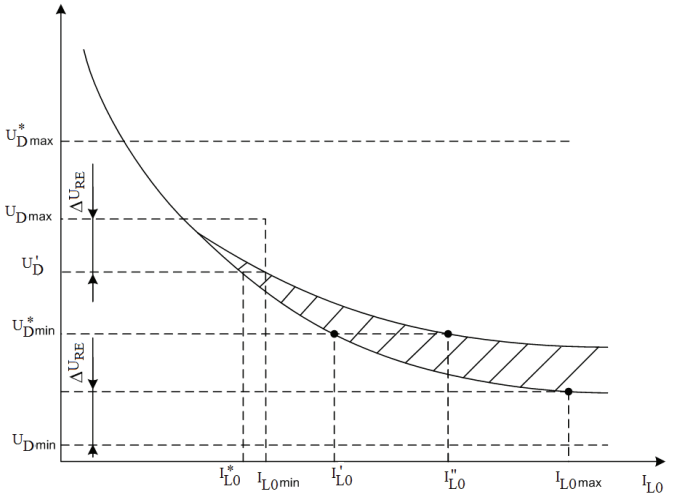


Fig.1. Single-circuit and double-circuit regulating apparatus

The mutual involvement of the input and output parameters, the influence of the destabilizing factors and the control modes can be traced in Fig.2.

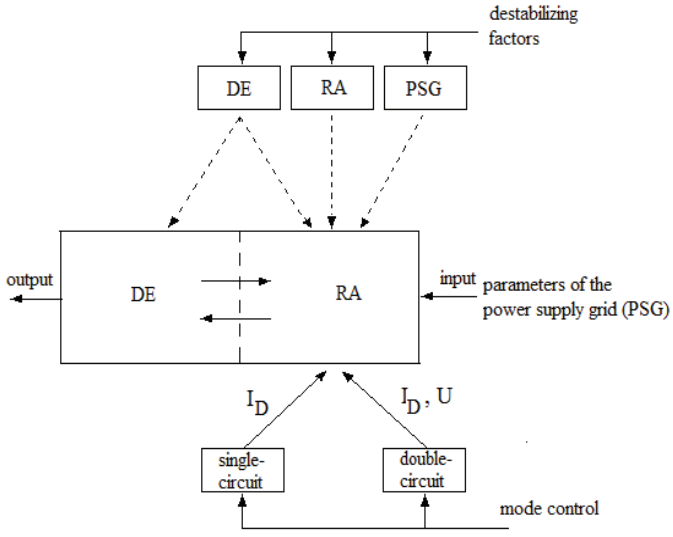


Fig.2. System "regulating apparatus (RA) -discharge element (DE)" – input/output parameters, destabilizing factors, control modes

The purpose of the present work is conducting of comparative analysis with regard to the current stabilization coefficientof the load (discharge current) and dissipated power by the regulating element and efficiency coefficient.

[1]Stefan Barudov is with the Faculty of Electrical Engineering, Technical University of Varna, 1, Studentska Str, Varna 9010, Bulgaria, e-mail: sbarudov@abv.bg

[2]Rositsa Dimitrovais with the Faculty of Electrical Engineering, Technical University of Varna, 1, Studentska Str, Varna 9010, Bulgaria, e-mail: r.dimitrova@tu-varna.bg

[2]Milena Ivanova is with the Faculty of Electrical Engineering, Technical University of Varna, 1, Studentska Str, Varna 9010, Bulgaria, e-mail: m.dicheva@tu-varna.bg

II. ANALYSIS

In Figs.3 and 4 are shown the block diagrams of a SCRA and a DCRA, where:

- DCVS is a DC voltage source which supplies regulating elements connected in series with the RE of the CS and the DE. DCSV also includes sections for providing of the starting mode.
- DE is a discharge element.
- U_m and ΔU_m - the amplitude value and its change of the input supply voltage.
- U_{S0} and ΔU_{S0} - the constant component and its change of the output voltage of DCVS. $U_{S0} = U_{RE} + U_{DE}$, where U_{RE} is the voltage applied over the regulating element and U_{DE} – over the discharge element.
- I_{L0} and ΔI_{L0} – value of the constant component of the discharge current and its change.
- R_{LS} and ΔR_{LS} – static resistance of the load (DE) and its change.

The analysis is performed with the following assumptions:

1. For the specified structures of RA destabilizing factors are the quantities U_m and R_{LS}, and the stabilizing quantity is I_{L0}.
2. DC1RA realizes the law ΔU_{S0}=const, and DC2RA – ΔU_{RE}=const.
3. The coefficient of stabilization K_{ST} can be defined by Eq. (1):

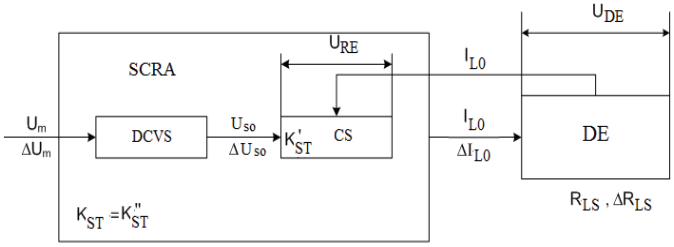


Fig.3 Block diagram of SCRA

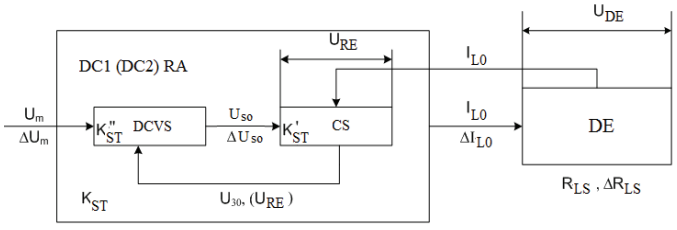


Fig.4 Block diagram of DCRA

$$K_{ST} = \frac{\Delta x}{x} . \frac{y}{\Delta y} \tag{1}$$

In the pointed expression x is a destabilizing factor and y is the stabilized quantity.

For the examined RA can be defined a stabilization coefficient with regard to the voltage K_{STU} and a coefficient with regard to the resistance K_{STR}, defined at different cases as follows:

- For SCRA:

$$K_{STU} = \frac{\Delta U_m}{U_m} . \frac{I_{LO}}{\Delta I_{LO}} = \frac{\Delta U_{S0}}{U_{S0}} . \frac{I_{LO}}{\Delta I_{L0}} \tag{2}$$

- For DC1RA:

$$\left.\begin{aligned} K'_{STU} &= \frac{\Delta U_{30}}{U_{30}} . \frac{I_{L0}}{\Delta I_{L0}} \\ K''_{STU} &= \frac{\Delta U_m}{U_m} . \frac{U_{S0}}{\Delta U_{S0}} \\ K_{STU} &= \frac{\Delta U_m}{U_m} . \frac{I_{L0}}{\Delta I_{L0}} \end{aligned}\right| K_{STU} = K'_{STU} . K''_{STU} \tag{3}$$

- For DC2RA:

$$\left.\begin{aligned} K'_{STU} &= \frac{\Delta U_{S0}}{U_{S0}} . \frac{I_{L0}}{\Delta I_{L0}} \\ K''_{STU} &= \frac{\Delta U_m}{U_m} . \frac{U_{RE}}{\Delta U_{RE}} \\ K_{STU} &= \frac{\Delta U_m}{U_m} . \frac{I_{L0}}{\Delta I_{L0}} \end{aligned}\right| \begin{aligned} K_{STU} &= K'_{STU} . K''_{STU} . A \\ A &= \frac{\Delta U_{RE}}{U_{RE}} . \frac{U_{S0}}{\Delta U_{S0}} \end{aligned} \tag{4}$$

For the three considered RA - $K_{STR} = \frac{\Delta R_{LS}}{R_{LS}} . \frac{I_{LO}}{\Delta I_{LO}}$.

From Eq.(2), Eq.(3) and Eq.(4) it follows that DCRA provides significantly higher K_{STU} than SCRA. The character of the feedback in the controllable rectifier which influences over K''_{STU} leads to a different magnitude of $K_{STU}$for the two types of DCRA. For them $U_{S0} = U_{RE} + I_{L0} R_{LS}$. Assuming that U_{RE}, I_{L0} and R_{LS} are independent variables, for ΔU_{30} is valid Eq.(5):

$$\Delta U_{S0} = \Delta U_{RE} + \Delta I_{L0} R_{LS} + \Delta R_{LS} I_{L0} \tag{5}$$

After transforming Eq.(4) and Eq.(5) for DC2RA, K_{STU} can be defined by Eq.(6).

$$K_{STU} = K'_{STU} . K''_{STU} . \frac{U_{S0}}{U_{RE}} \left(\frac{\Delta I_{L0}}{\Delta U_{RE}} R_{LS} + \frac{\Delta R_{LS}}{\Delta U_{RE}} I_{L0} + 1 \right)^{-1} \tag{6}$$

From Eq.(6) and Eq.(7) it follows that with respect to K_{STU} the choice of DC1RA is advisable since the inequality (Eq.(7)) is always in force:

$$\frac{U_{S0}}{U_{RE}}\left(\frac{\Delta I_{L0}}{\Delta U_{RE}}R_{LS}+\frac{\Delta R_{LS}}{\Delta U_{RE}}I_{L0}+1\right)^{-1}=A<1 \qquad (7)$$

For typical values of I_{L0}, ΔI_{LO}, U_{RE}, ΔU_{RE} and R_{LS} – A=0,2÷0,4.

Hence:

$$K_{STR-SCRA}=K_{STR-DC1RA}=K_{STR-DC2RA}$$
$$K_{STU-SCRA}<K_{STU-DC2RA}=K_{STU-DC1RA}$$

In Figs.5 and 6 are presented respectively $P_{RE}=P_{RE}(I_{L0})$ and $P_{DE}=P_{DE}(I_{L0})$; $\eta=\eta(I_{L0})$ for:

- Curve 1 – DC2RA, controllable rectifier (U_{RE}=600 V);
- Curve 2 – DC2RA, step AC regulator which limits the change of U_{RE} - 600÷2000 V;
- Curve 3 – DC1RA – step AC regulator which limits; $\Delta U_{S0}/ U_{S0}$ in the range up to 5%;
- Curve 4 – SCRA;
- Curve 5 – $P_{DE}=P_{DE}(I_{L0})$;

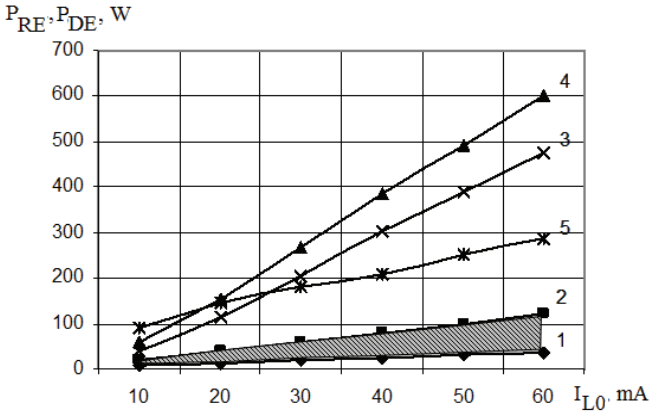


Fig.5.

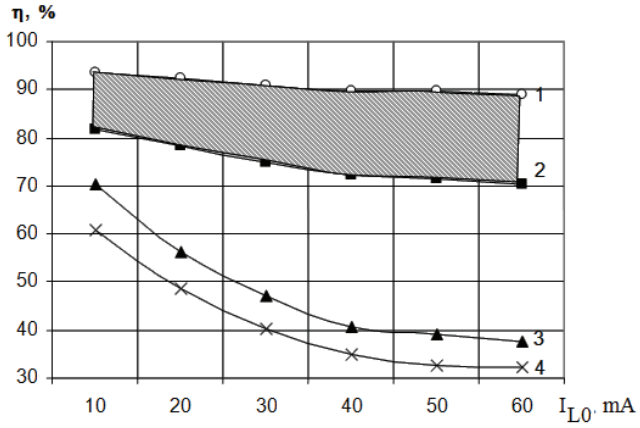


Fig.6.

III. CONCLUSION

The DCRA for discharge elements with comparison to the SCRA provides increased stabilization coefficient of the discharge current only with respect to the change of the input supply voltage. This coefficient quantitatively depends on the character of the feedback of the controllable rectifier.

The choice of DC1RA and DC2RAis a matter of compromise. DC1RA ensures a higher stabilization coefficient of the discharge current related to the change of the input supply voltage but a lower efficiency coefficient and higher dissipated power of the regulating element and vice versa.

The realization of the second control circuit with a controllable rectifier or a step AC regulator also is a question of compromise. The controllable rectifier gives a better efficiency coefficient and lower dissipated power of the regulating element, yet worse power factor and higher level of electromagnetic interference and vice versa.

REFERENCES

[1] Барудов С.,Методы и практика проектирования устройств для управления разрядами в газах, 2004, ПГТУ-Пермь, Россия ISBN954-20-02-79-3.

[2] Sotirov L., S. Barudov, S. Sotirova. Computer mathematical apprcach to discrete modeling of discharge processes in xenon pulse tubes based on optimal singular adaptive M1A6 computer observer. Comptes rendus de L Academie des Sciences tome 60,№ 5,2007, pp. 533-542.

[3] Barudov S., Start-regulating device for pulse xenon lamp. Acta Universitatis Pontica Euxinus, Vol.1, №1, 2002, pp.46-50.

[4] Barudov S., A current regulator in the high –voltage discharge circuit. Ovidius University Annals of Mechanical Engineering, Vol. V, №1, 2003, pp.69-73.

[5] Rashid M., Power Electronics: Circuits, Devices and Applications, Third Edition, Prentice Hall, 2003.

Discharge Element with Transverse High-Frequency Excitation

Stefan Barudov[1], Milena Ivanova[2]

Abstract – **Recently, the usage of high-frequency transverse discharge in technologies for treatment of liquid fluids presents practical interest because: the discharge is characterized by higher stability; reactive elements can be used as a passive load and efficiency coefficient increases; there is symmetric energy dissipation; electrodes with dielectric coating can be used, which solves the problems with their sputtering.The work is devoted to study of the possibilities for application of single circuit generators with automatic pre-voltage and inductive-capacitive matching converter for excitation of a transverse discharge and control of its parameters according to the technological application.**

Keywords – **single circuit generator, generator with automatic pre-voltage, inductive-capacitive converter**

I. Introduction

The effectiveness of the discharge medium excitation highly depends from the ratio between the frequency of the external field and the frequency of the interaction between the electrons and the atoms and more precisely if the electrons energy succeeds to follow the change of the external field during the oscillationperiod. The characteristic velocity of changing of the electron energy with the change of the field is Eq.(1):

$$f_e \cong k_{ei} n_a m_e / m_a \qquad (1)$$

At high frequencies ($f>>f_e$) the energy of the electrons slightly changes in time. Therefore arise conditions for development of an ionization process with constant velocity during the whole time of discharge existence.

Furthermore, the energy of the electron provides such frequency of ionization that electrons have in the so called constant effective electrical field.

$$E_{eff} = \frac{E_a}{\sqrt{2}} \cdot \frac{\nu_{ei}}{\left(\omega^2 + \nu_{ei}^2\right)^{0,5}} \qquad (2)$$

[1]Stefan Barudov is with the Faculty of Electrical Engineering, Technical University of Varna, 1, Studentska Str, Varna 9010, Bulgaria, e-mail: sbarudov@abv.bg

[2]Milena Ivanovais with the Faculty of Electrical Engineering, Technical University of Varna, 1, Studentska Str, Varna 9010, Bulgaria, e-mail: m.dicheva@tu-varna.bg

E_a – amplitude of the field

ω – frequency of the external electrical field

υ_{ei} – frequency of the elastic interactions of the electrons with the neutral atoms;

In case of low frequencies ($f \leq f_e$) the electrons energy "follows" the electrical field. In that case, the discharge in an alternating electrical field remindsconductivity current-dependent discharge with periodic ionization. In force is the exponential character of the dependency of the ionization velocity from the electron energy. Gas ionization is accomplished only in the regions with maximum intensity of the external electrical field, and during the interval between them flows current of the conductivity-dependent discharge inthe decomposing plasma. Typical values of f_e at $n_a \sim 10^{18} cm^{-3}$; $k_{ei} \sim 10^{-7} cm^3.s^{-1}$; $m_e/m_a \sim 10^{-5}$ are $10^6 \div 10^7 Hz$. The amplitude which is necessary for supporting of the high-frequency discharge is defined by the conditions for the balance between generation and recombination of the electrons and it is close to the intensity of the electrical field in direct current discharge.

Near-electrode processes in high-frequency discharge play a relatively minor role than in direct current discharge.

An example realization of a single circuit lamp generator is shown in Fig.1.

The conditions for generator self-excitation and the stability of the steady mode are discussed in [1].

The aim of the present work is study of the possibilities for ensuring of starting and operating modes of the transverse discharge [2], its control and choice of a converter as a matching section for transforming the output voltage of the resonant circuit into current of the discharge element.

II. Analysis

Quadripoles with reactive elements [3] are connectedbetween the load (the discharge element) and the generator. They ensure the following:

- converting of the generator output voltage into constant current through the load according to its specific properties;
- transformation of the active part of the load resistance R_L in resistance of the circuit R_e.

The dependency between the anode voltage U_a and the converter input voltage U_L with a load (the discharge element) is necessary to be found for calculation of the relationship.

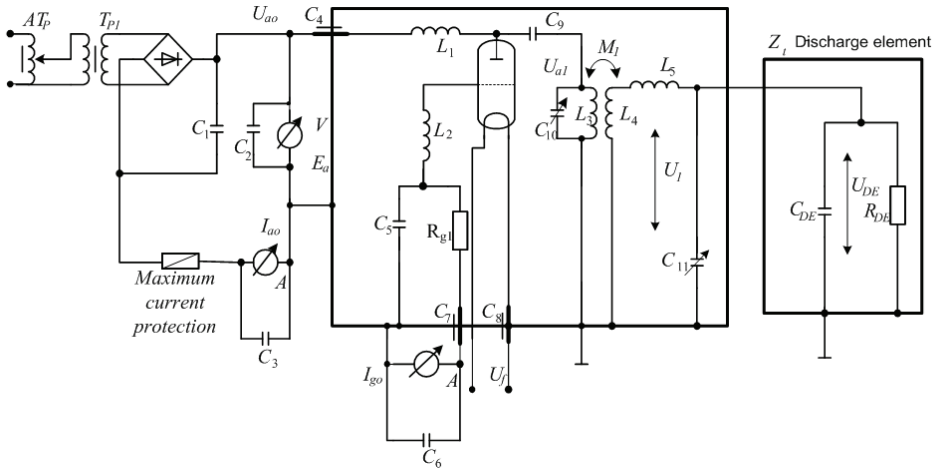


Fig.1. Single-circuit lamp generator

In the presented general circuit of the relationship (Fig.2) Z_a is the impedance of the anode circuit, Z_L – the converter impedance (discharge element),and Z_m is the mutual impedance. C stands for a converter and DE is the discharge element.

Before a discharge occurs in the discharge medium i_L=0, the circuit from Fig.2 could be transformed equivalently into the circuit in Fig.3, where Z_τ is resistance at short circuit, which includes also theresistance introduced from the anode circuit. After transformation according to the Thevenin's theorem the scheme in Fig.3 becomes the one, shown in Fig.4.

$$U_{LOC} = \frac{U_a}{Z_a}.Z_m \tag{3}$$

$$Z_\tau = Z_L + Z_{\text{int}roduced} = Z_L - \frac{Z_m^2}{Z_a} \tag{4}$$

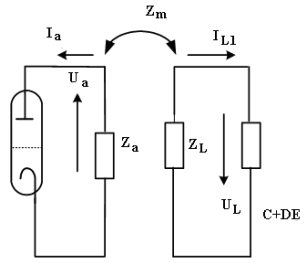
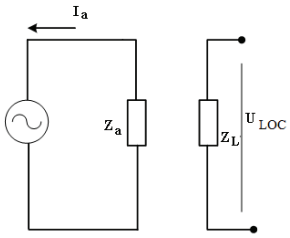


Fig.2



.Fig.3.

Scheme of the relationships and equivalent representation

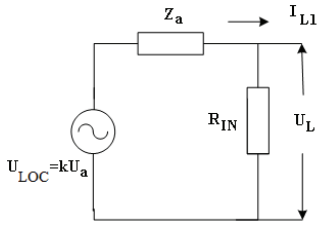


Fig.4.

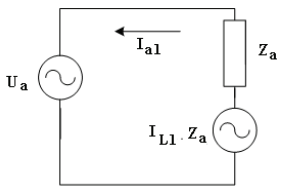


Fig.5.

Equivalent scheme of the anode circuit with consideration of the influence of the converter and the discharge circuit

If we define a transformation coefficient K – Eq.(5), then when load current flows is valid Eq.6:

$$K = \frac{U_{LOC}}{U_a} = \frac{I_{a1}Z_m}{I_{a1}Z_a} = \frac{Z_m}{Z_a} \tag{5}$$

$$U_L = KU_a - I_{L1}Z_a \tag{6}$$

Hence, the influence of the anode circuit on the load is expressed by excitation of electromotive force (EMF) – $k.U_a$ and introducing of additional resistance.

The influence of the load on the anode circuit can be reported by the equivalent circuit in Fig.5 by Eq.(7) and Eq.(8):

$$U_a = I_{a1}Z_a - I_{L1}Z_m \tag{7}$$

$$U_a = Z_a(I_{a1} - KI_{L1}) \tag{8}$$

In steady mode the voltage in the load circuit and the anode voltage of the generator can be given by Eq.(6) and Eq.(8).

In case that K is a real number, Z_a is an imaginary number (Z_a – reactive resistance), where $Z_\tau << Z_a$ and if Z_a is the resistance of a parallel resonance circuit in resonance – $Z_a = R_{EqOC}$at i_L=0, then after transformation of Eq.(6) and Eq.(8) for R_e (at i_L=0) is valid Eq.(9):

$$R_e = \left(1 - \frac{KI_{L1}}{I_{a1}}\right).R_{EqOC} \tag{9}$$

The last equation considers the change of the resonance resistance of the anode resonant circuit from the work of the discharge element.

The discharge element control suggests necessity from converting of the generator output voltage into current i.e. the voltage source is transformed into a current source.

An example variant of such transformation is shown in Fig.6.

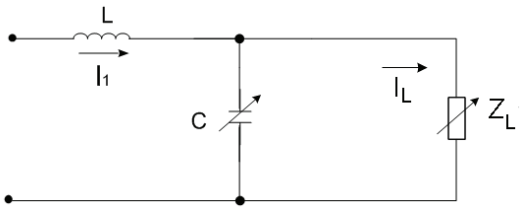


Fig.6. Inductive-capacitive converter

If for L and C is valid $j\omega L = -j\frac{1}{\omega C}$, which is achieved by change of the capacitance C and $u_1 = U_m \sin \omega t$, then the complex value of the load current can be derived from Eq.(10).

$$\overset{*}{I}_L = \frac{-j\overset{*}{U}_1}{\omega L + Z_L(\omega^2 LC - 1)} = \frac{-j\overset{*}{U}_1}{\omega L} \qquad (10)$$

Analyzing the schematic solution from Fig.6, it can be noted that at $Z_L \to \infty$ the system aims at resonance and U_C increases. This favors the rise of a discharge in the discharge medium as the magnitude of U_C is limited by the reduced active resistance, which indicates the losses in L and C as well as the influence of the previous step. At resonance I_1 (Fig.6) increases i.e. according to Eq.(9) R_e decreases, which leads to decrease of U_a and consequently decrease of U_C and I_1.

At Z_L=0, I_1 is limited by the inductive reactance ωL, i.e. this mode does not present any danger.

The linear relationship between the current I_L through the discharge element and U_1 at invariable ω and L without additional compensations means that the instability of I_L doesn't exceed the instability of U_1.

III. Experimental Study

In Fig.7 is shown a prototype of the generator from Fig.1, developed with a lamp (metal-ceramic triode ГИ-7Б) with air cooling. The amplitude of the generated high-frequency oscillation with a frequency of 2MHz can be changed by variation of the anode supply voltage of the lamp generator.

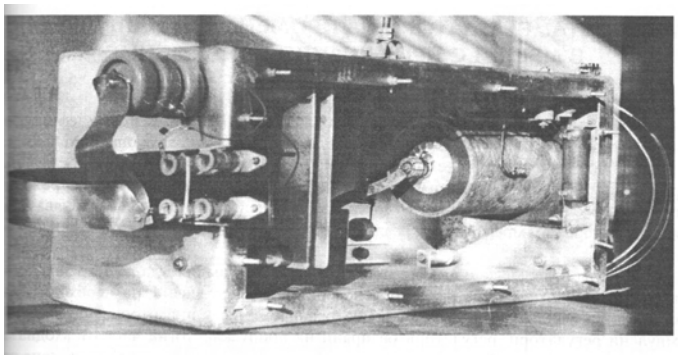


Fig.7. Experimental prototype of a single-circuit lamp generator with a metal-ceramic triode with forced air cooling

The experimental studies have been conducted with an equivalent of the discharge circuit shown in Fig.8.

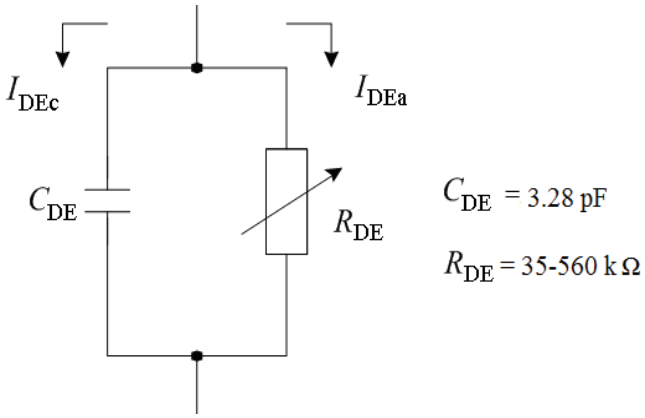


Fig.8. Equivalent representation of the discharge element with transverse high-frequency excitation

At open circuit mode R_{DE}=∞, considering the certain values of L and C in the resonant circuit of the lamp generator and its quality factor, there are achieved values $U_{C_{DE}} = U_{C_{DE}}(2{,}5 \div 4)U_{a1}$ - higher than 14kV.

I_{DEa} (Fig.8) is regulated in the range 25÷100 mA. The achieved maximum power over the load is 350 W.

In the matching impedance circuit the condition $\omega^2 LC = 1$ can be hardly achieved in practice due to a deviation of the generator frequency, the parameters temperature change of L and C and disregard of their parasitic active resistances.

In Fig.9 is shown the change of $\frac{\delta I_{DE}}{I_{DE}} = f_1(A)$ where $A = 1 - \omega^2 LC$ for equivalent R_{DE}=20kΩ and C=3,28pF and R_{DE}=60kΩ and C=3,28pF. The change of A with 20% and R_{DE} – three times – leads to a relative change of I_{DE} not exceeding 20%.

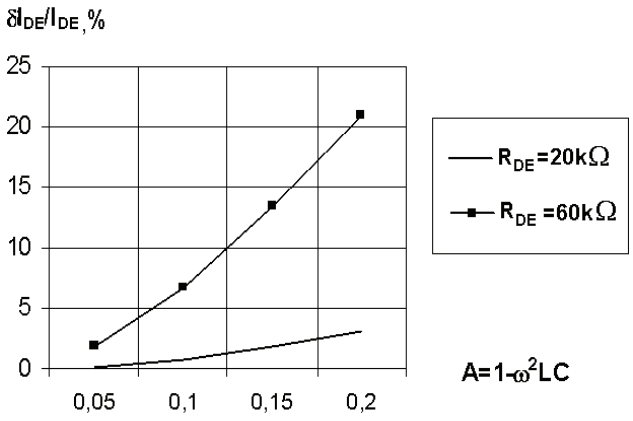


Fig.9. Relative change of the load current as a function of the change of $A = 1 - \omega^2 LC$

IV. Conclusion

It is proposed a variant of a lamp high-frequency generator with a matching section – L-shaped inductive-capacitive converter, which gives the possibility at generated frequency of 2MHz for ensuring starting and operating mode for excitation of a transverse discharge.

The generator provides possibility for fine adjustment of the discharge current in the range of 25÷100 mA at maximum output power 350 W.

The studies of the discharge current stability as a function of the change of the generator parameters and those of the discharge plasma reveal appropriate choice for the matching section.

References

[1] BARUDOV S.T.,Electrical processes and equipment for control of discharges in a gas medium. Baku, Azerbaijan, 2004. ISBN 5-86106-002-9.

[2] YouS.J., T.T. Hai, M. Park, D.W. Kim, J.H. Kim, D.J. Seong , Y.H. Shin, S.H. Lee, G.Y. Park, J.K. Lee, H.Y. Chang, Role of transverse magnetic field in the capacitive discharge, Thin Solid Films (2011), Elsevier.

[3] S. Y. Ron Huia and Henry S.H. Chung, Power Electronics Handbook (Second Edition), Devices, Circuits, and Applications, 2007

Investigation and Analysis of Organic Electroluminescent Heterostructures

Mariya Aleksandrova[1], Georgy Dobrikov[1], Milka Rassovska[1]

Abstract – **In this study organic electroluminescent structures using heteroelectrode have been prepared to increase current density and injection efficiency. Current-voltage characteristics are measured to evaluate the improvement of the electrical behavior. For clarifying and further understanding of the processes in the bulk and at the layer interfaces, additional simulation analysis has been carried out. The charge carriers spatial distribution and electrical field distribution depending on the structure configuration has been investigated.**

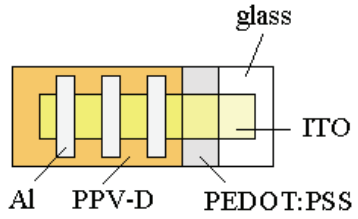
Keywords – **Organic semiconductors, Electroluminescent devices, Heteroelectrode, Heterojunctions.**

I. Introduction

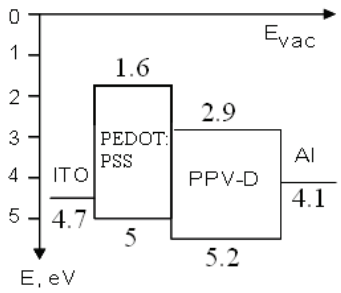
Organic electroluminescence (EL) is electrically stimulated emission of light from organic semiconductors and finds application in organic light-emitting display (OLED) [1]. There is rapid progress in improvement of the stability of these devices, even for the materials, which produce blue light and require higher energies for charge carrier transitions in the structure [2]. However, for the practice there is one more important parameter except exploitation time. This is device's efficiency and in this case is quantum efficiency, or ratio between numbers of generated photons to numbers of injected charge carriers, which recombine inside the structure. There are many efforts concentrated in this direction [3,4], but most of the explanations about the physical processes in OLED are not full or the knowledge is still poor. Many materials are incorporated in the EL structure between the electrodes and the light emitting organic semiconductor to decrease the height of injection barrier at the layer's boundary [5,6]. Despite of the energy level alignment at the interfaces the charge balance is destroyed after inserting of additional layers, so deeper investigation of this problem is necessary.

In this paper we suggest combine approach of experimental measurements and numerical simulations to get comprehensive information about the behavior and electrophysical properties of OLED structure, which consists of heterojunction for higher efficiency. The investigation aims to reconstruct processes at the layer interfaces and bulks, where it is not possible to receive simple picture about the device only from direct experimental measurements.

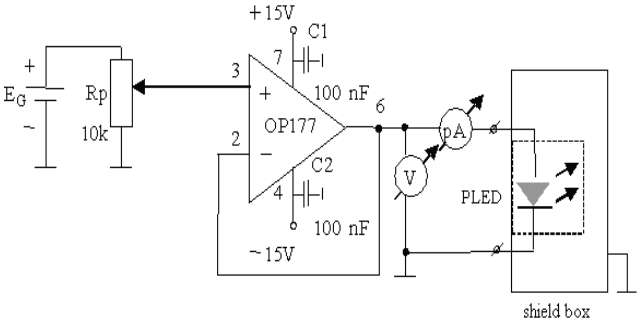
II. Experimental Section



a)



b)



c)

Fig. 1. Prepared multilayer EL structure (a), energy levels diagram (b) and block diagram of the test circuit for I-V characteristics measurement (c).

[1]Mariya Petrova Aleksandrova is with Faculty of Electronic Engineering and Technologies, Department of Microelectronics, Technical University of Sofia, Bulgaria, "Kliment Ohridski" blvd, 8 E-mail: meri_7@abv.bg.

[1]Georgy Dobrikov is with the Faculty of Electronic Engineering and Technologies, Department of Microelectronics, Technical University of Sofia, Bulgaria, "Kliment Ohridski" blvd, 8 E-mail:georgy_ddobrikov@abv.bg

[1]Milka Rassovska is with the Faculty of Electronic Engineering and Technologies, Department of Microelectronics, Technical University of Sofia, Bulgaria, "Kliment Ohridski" blvd, 8 E-mail: rasovska_m@mail.bg

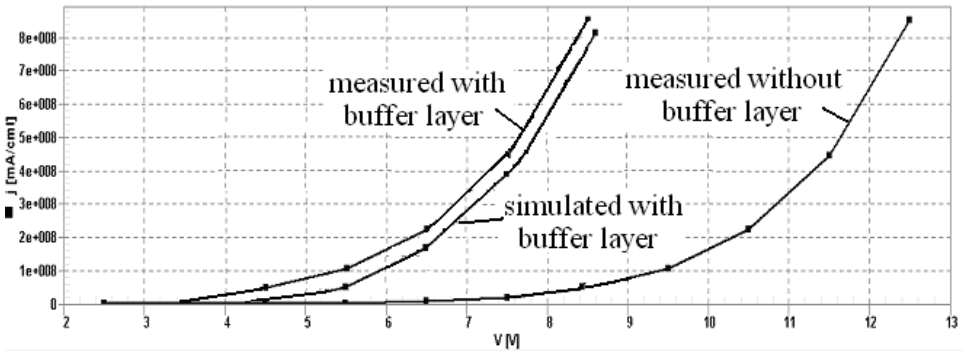


Fig. 2. Current-voltage characteristics of the organic structure with and without heterojunction together with the simulation curve.

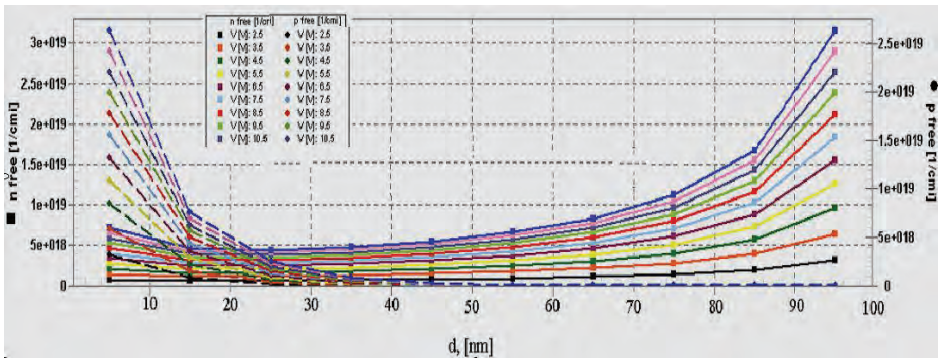


Fig. 3. Spatial distribution of the charge carriers in the bulk of EL structure.

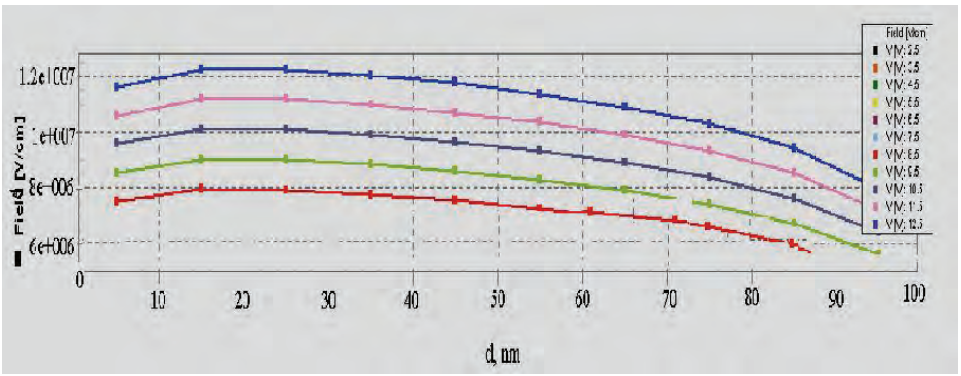


Fig. 4.Electricalfield distribution in depth in the EL structure.

For the fabrication of electroluminescent heterostructure, indium tin oxide (ITO) covered glasses obtained by RF reactive sputtering were used as substrates with anodes. The target was metallic indium tin disk, mounted in vacuum installation model A400VL. Rotary and turbomolecular pumps were used to achieve the desired vacuum level in the system, which is less than 1×10^{-5} Torr. The obtaining of conductive and transparent ITO thin films is described elsewhere [7]. For preparation of hole transporting layer solution of 1 ml poly(3,4-ethylenedioxythiophene) dissolved in 10 ml poly(styrenesulfonate) (PEDOT:PSS) is deposited by spin coating with rotation speed of 1000 rpm for 40s to produce 60 nm thin film. The emissive layer made from poly((9,9-dihexyl-9H-fluorene-2,7-vinylene)-co-(1-methoxy-4-(2-ethylhexyloxy)-2,5-phenylenevinylene)) (PPV-D) is deposited also by spin coating at the same conditions. Important heterojunction is formed between both polymer layers. Finally Al cathode was deposited by vacuum thermal evaporation. Prepared multilayer EL structure is shown on Fig. 1.

The current-voltage characteristics were measured by precise Keythley 6485 picoampermeter.Etalon structure without buffer layer is prepared and its current-voltage characteristics are compared to that with hole transporting layer.

The analysis in this article is based on simulation results achieved with OLED simulation software based onOLEDWin,which consider processes like charge carrier injection and transport, radiative recombination, non-radiative decay, ratio between singlet and triplet states, etc. The program uses experimental measured data as input parameters, so the simulation results complement the data from the real measurement.

III. Results and Discussion

Measured current-voltage characteristics of the etalon structure and the structure with buffer layer are compared and shown on Fig. 2. In the same plot is presented the result from simulation of the same structure, consisting of the same materials as thin films, with the same thickness and consequence of deposition. As could be seen there is great match between measured and simulated curves (average error of 1.1 %), which give us reason to consider the further simulations as exact representation of the real structure's behavior.

Following assumptions are accepted for the simulated device: trap depth 0.25 eV; singlet and triplet lifetime respectively 1.10^{-8} s and 1.10^{-6} s; efficiency of radiative decay 0.3; charge carriers mobility 10^{-4} $cm^2/V.s$ for holes and 10^{-2} $cm^2/V.s$ for electrons. These acceptances are based on the typical reported values in the scientific literature for the organic semiconductors and concretely for organic electroluminescent devices. One of the main reasons for bad efficiency is the meeting point of both types of charge carriers, namely near one of the electrodes, where high defect concentration exists. This happens, because of the different mobility for electrons and holes and the different injection barriers at the contacts. The defects are caused by electrode deposition processes and they act as luminescence quenching centers. That's why it is necessary the recombination zone to be shifted away from the electrode interfaces and toward the bulk of the organic light emitting layer. Fig. 3 shows spatial distribution of the charges in the bulk and the recombination zone in the case of inserting PEDOT:PSS as buffer layer between ITO and PPV-D. Starting zero value of the "x" axis means the interface between PEDOT:PSS and PPV-D. There is intermediate zone of approximately 5 nm where both layers are partly penetrated to each other. If the whole thickness of the electroluminescent layer is 100 nm the meeting zone of the opposite charge carriers is in the range 25-35 nm, which is not exactly in the middle of the film, but the recombination zone is shifted in comparison with the case without PEDOT:PSS. Charge distribution is independent of the applied voltage in the operational range of 2.5 to 10.5 V, which proves the stability of the prepared device.

Because of the uniform distribution of charges there is no space charge formation and therefore current limitations and field distortions are missing. This can be observed on Fig. 4 where is shown the electrical field distribution inside the EL layer. There are some deflections at the interface with the aluminum electrode (near 100 nm) because in this place there is no cathode buffer layer to improve the conditions for electron injection as this is made for holes. This tendency is preserve with the change of the applied voltage intensity from 2.5 to 12.5 V/cm.

IV. Conclusion

In summary hole injection efficiency in organic EL device is found to be enhanced by adding thin PEDOT:PSS layer between the ITO anode and light-emitting layer. Theadvantage of such heteroelectrod (ITO/PEDOT:PSS) and heterojunction (PEDOT:PSS/PPV-D) is energy level alignment at the interfaces and injection barrier decreasing, as well as shifting of the recombination zone inside the structure's bulk and better distribution of the electrical field. It was established that the place of the recombination zone is not sensitive to the voltage applied. We think that this kind of combined research (experimental and simulations) explains well electrophysical behaviour of organic EL devices.

Acknowledgement

This work was financially supported by grant D002-358/2008 of Ministry of Education and Science, Bulgaria.

References

[1] M. Eritt, C. May, K. Leo, M. Toerker, C. Radehaus, "OLED manufacturing for large area lighting applications", Thin Solid Films, vol. 518, no. 11, pp. 3042-3045, 2010.

[2] S. O. Kim, K. H. Lee, G. Y. Kim, J. H. Seo, Y. K. Kim, S. S. Yoon, "A highly efficient deep blue fluorescent OLED based on diphenylaminofluorenylstyrene-containing emitting materials", Synthetic Metals, vol. 160, no. 11-12,pp. 1259-1265, 2010.

[3] H. Mu, W. Li, R. Jones, A. Steckl, D. Klotzkin, "A comparative study of electrode effects on the electrical and luminescent characteristics of Alq_3/TPD OLED: Improvements due to conductive polymer (PEDOT) anode", Journal of Luminescence, vol. 126, no. 1, pp. 225-229, 2007.

[4] J. W. Ma, Z. Liang, C. Jin, X. Y. Jiang, Z. L. Zhang, "Enhanced power efficiency for white OLED with MoO3 as hole injection layer and optimized charge balance", Solid State Communications, vol. 149, no. 5-6, pp. 214-217, 2009.

[5] H. P. Lin, D. B. Yu, X. W. Zhang, J. Li, L. Zhang, X. Y. Jiang, Z. L. Zhang, "Enhanced hole injection and improved performance in organic light-emitting devices by utilizing a novel composite hole injection layer", Solid State Communications, vol. 150, no. 35-36, pp. 1601-1604, 2010.

[6] X. W. Zhang, M. A. Khan, X. Y. Jiang, J. Cao, W. Q. Zhu, Z. L. Zhang, "Electron injection property at the organic–metal interface in organic light-emitting devices revealed by current–voltage characteristics", Physica B: Condensed Matter, vol. 404, no. 8-11, pp. 1247-1250, 2009.

[7] G. Dobrikov, M. Rassovska, E. Kashchieva, "Physical Properties and Microstructure of Indium-Tin Oxide Films", Electronics&Electrotechnics, vol. 5-6, pp 170-172, 2006.

Session PO8:

PO8 - POWER TRANSMISSION, DISTRIBUTION SYSTEM AND ELECTRICAL MACHINES

Possibilities to Manage Burning Process at the Conditions of Cement Kiln

Neven Krystev[1]

Abstract **- In the work are shown results of numerical modeling of burning process in the combustion chamber of cement kiln. A program for numerical modeling in the MATLAB platform, created in Faculty of Engineering and Pedagogy Sliven at Technical University Sofia is used. Received results show different ways to change the shape, position of the torch and temperature level in the combustion chamber.**

Key words **– numerical modeling, degree of rotation, gas fuel**

I. INTRODUCTION

Rotary kiln is one of the most important units in the technological process for obtaining of clinker in the cement industry. It is got by firing to the boundary of melting of the finely ground particles and completely mixed raw materials taken in appropriate proportions. Firing takes place at about 1400 ° C to convert into a solid product.

In these kilns is organized diffusion torch over a moving composite material. Its geometric shape and heat-exchange performance have to ensure optimal temperature regime in different zones, thermal efficiency and operating reliability of the furnace [1].

The raw material is given in the form of well-prepared sludge. Humidity of the sludge is in the limits 30-40%. Falling in the rotating kiln, the raw material consequently passes through different zones of physical-chemical conversion [2].

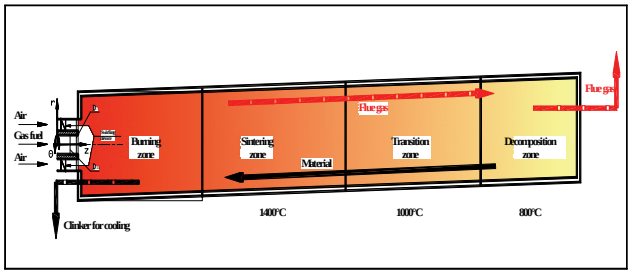


Fig. 1. Rotary furnace scheme of cement clinker formation.

[1]Neven Jordanov Krystev – Faculty of Engineering and Pedagogy – Sliven at Technical University Sofia, No 59 "Burgasko shosse" Bld., 8800 Sliven, Bulgaria, E-mail: nkrystev@yahoo.com

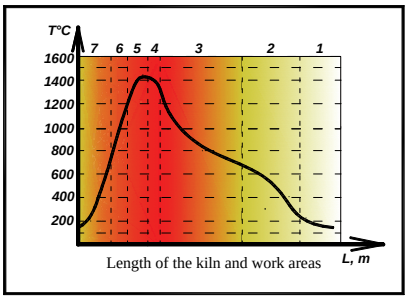


Fig. 2 Specific temperature zones.

Specific temperature zones along the furnace are following:

- 1 – drying zone (20%L);
- 2 – initial heating zone (25%L);
- 3 – zone of calcination (35%L) – dehydration and dissociation of carbonates;
- 4 – zone of exothermic reactions (5%L). - formation of silicates and compounds of aluminum and calcium;
- 5 – sintering zone (8%L);
- 6 и 7 – zones of cooling of the material (7%L) where, the combustion air simultaneously is heated to 250-300 °C (≈500K).

To create optimal technological regimes, well-grounded from economical point of view, is necessary to predict opportunities for active influence on combustion process in rotating furnace, allowing temperature changes along its length [3].

The purpose of this work is with numerical simulation of burning process in diffusion torch to obtain temperature regimes along the furnace.

The obtained results would be allowed to analyze the temperature conditions in the combustion chamber for appropriate management of the combustion process at constant heat load.

In numerical experiments are used geometric dimensions and thermal load of a real cement kiln. The main dimensions are:

- Diameter of the gas hole D_{11} = 60 mm;
- Diameter of the air hole D_7 = 316 mm;
- Diameter of the combustion chamber $D_5 = D_3$ = 3000 mm;
- Length of the burner L_{75} = 1400 mm;
- Length of the combustion chamber L_{54} = 30000 mm.

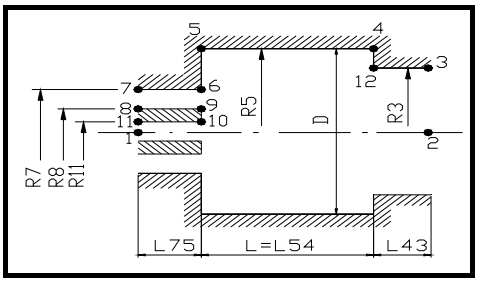


Fig. 3 Computational scheme of the combustion chamber with typical geometric dimensions.

The combustion process is simulated by fuel - natural gas, with a coefficient of excess air α = 1,05 and temperature of air heating - 300 °C.

On the basis of passport data is set nominal heat load at gas velocity -120 m/s and air velocity - 86,5 m/s (300°C) and 45,27 m/s (20°C) according with dimensions of the geometric model.

The study is done at adiabatic conditions. In the different experiments is changed only the degree of rotation, determined by [4] in the range from 0 to 2.0, which corresponds to the possibilities of the swirling device of the burner.

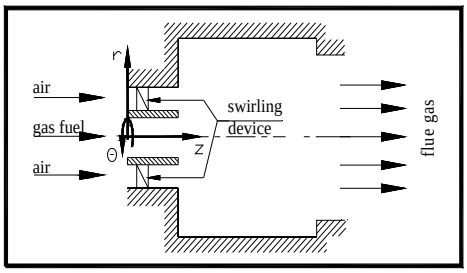


Fig. 4. Scheme of the combustion chamber with swirling device.

As a measure to define the degree of rotation of the jets is used the relationship: $s = \frac{\overline{w}}{\overline{u}}$, where $\overline{w}$ and $\overline{u}$ are average integrated values of tangential and axial components of the velocity vector.

II. Numerical Model - Features

The mathematical model is presented with a system of equations for the cylindrical coordinate system [5] and it is built on the basis of equations for the quantity of the motion, for energy and mass saving, and relationships describing the turbulent model, chemical reactions, boundary conditions, thermodynamic, the source and transmission properties of the fluid. The authors are used the JB Spalding's algorithm from London University, as the mathematical model is modified for non-isometric conditions. The differential equations are transformed into cylindrical coordinates.

They are as follows:

- Equation about velocity vortex:

$$r^2\cdot\left\{\frac{\partial}{\partial z}\left(\frac{\Omega}{r}\cdot\frac{\partial\psi}{\partial r}\right)-\frac{\partial}{\partial r}\left(\frac{\Omega}{r}\cdot\frac{\partial\psi}{\partial z}\right)\right\}-\frac{\partial}{\partial z}\left\{r^3\cdot\frac{\partial}{\partial z}\left(\mu_{e\phi}\cdot\frac{\Omega}{r}\right)\right\}-$$
$$-\frac{\partial}{\partial r}\left\{r^3\cdot\frac{\partial}{\partial r}\cdot\left(\mu_{e\phi}\cdot\frac{\Omega}{r}\right)\right\}-r\cdot\frac{\partial}{\partial z}\left(\rho\cdot w^2\right)- \quad (1)$$
$$-r\cdot\frac{\partial}{\partial z}\left(\frac{u^2+v^2}{2}\right)\cdot\frac{\partial\rho}{\partial r}+r^2\cdot\frac{\partial}{\partial r}\left(\frac{u^2+v^2}{2}\right)\cdot\frac{\partial\rho}{\partial z}=0$$

- Burning equation:

$$\frac{\partial}{\partial z}\left(\varphi_{zo}\cdot\frac{\partial\psi}{\partial r}\right)-\frac{\partial}{\partial r}\left(\varphi_{zo}\cdot\frac{\partial\psi}{\partial z}\right)-\frac{\partial}{\partial z}\left(\frac{\mu_{e\phi}}{\mathrm{Pr}}\cdot r\cdot\frac{\partial\varphi_{zo}}{\partial z}\right)-$$
$$-\frac{\partial}{\partial r}\left(\frac{\mu_{e\phi}}{\mathrm{Pr}}\cdot r\cdot\frac{\partial\varphi_{zo}}{\partial r}\right)=0 \quad (2)$$

- Equation for saving of the quantity of the motion on the direction Θ:

$$\frac{\partial}{\partial z}\left(r\cdot w\cdot\frac{\partial\psi}{\partial r}\right)-\frac{\partial}{\partial r}\left(r\cdot w\cdot\frac{\partial\psi}{\partial z}\right)--\frac{\partial}{\partial z}\left\{r^3\cdot\mu_{e\phi}\cdot\frac{\partial}{\partial z}\left(\frac{w}{r}\right)\right\}- \quad (3)$$
$$-\frac{\partial}{\partial r}\left\{r^3\cdot\mu_{e\phi}\cdot\frac{\partial}{\partial r}\left(\frac{w}{r}\right)\right\}=0$$

The differential equations in the numerical procedure for convenience are represented by general characteristic equation of elliptic type:

$$a_\varphi\cdot\left\{\frac{\partial}{\partial z}\left(\varphi.\frac{\partial\psi}{\partial r}\right)-\frac{\partial}{\partial r}\left(\varphi.\frac{\partial\psi}{\partial z}\right)\right\}-\frac{\partial}{\partial z}\left\{b_\varphi.r.\frac{\partial\left(c_\varphi.\varphi\right)}{\partial z}\right\}- \quad (4)$$
$$-\frac{\partial}{\partial r}\left\{b_\varphi.r.\frac{\partial\left(c_\varphi.\varphi\right)}{\partial r}\right\}+r.d_\varphi=0$$

The coefficients of the general characteristic equation are given in Table 1. By substituting the coefficients from the table in the general elliptic equation are obtained, the above differential equations.

Table 1 Coefficients of the characteristic equation.

φ	aφ	bφ	cφ	dφ
ψ	0	$\frac{1}{\rho\cdot r^2}$	1	$-\frac{\Omega}{r}$
$\frac{\Omega}{r}$	r^2	r^2	$\mu_{e\phi}$	$-\frac{\partial}{\partial z}(\rho.w^2)-$ $-r\cdot\left[\frac{\partial}{\partial z}\left(\frac{u^2+v^2}{2}\right)\frac{\partial\rho}{\partial r}-\frac{\partial}{\partial r}\left(\frac{u^2+v^2}{2}\right)\cdot\frac{\partial\rho}{\partial z}\right]$
r.w	1	$\mu_{e\phi}.r^2$	$\frac{1}{r^2}$	0
φ_{ro}	1	$\frac{\mu_{e\phi}}{\mathrm{Pr}}$	1	0

In the program is introduced turbulent viscosity [7], based on the hypothesis of Kolmogorov-Prandtl for "way of mixing":

$$\mu_{e\phi}=K.D^{\frac{2}{3}}.L^{-\frac{1}{3}}.\rho^{\frac{2}{3}}.\left(G_f.V_f{}^2-G_a.V_a{}^2\right) \quad (1)$$

where: K is constant;
D and L dimensions of the combustion chamber;
G – mass flow;

Indexes **f** и **a** regarding fuel and air;

V_f и V_a are real velocity.

V_a in the calculation procedure is presented as a geometric sum of axial and tangential component of absolute velocity.

III. Numerical Study

The main series of experiments are done at a temperature of air heating - 300 ° C (573.15 K).

An experiment without air heating - 20 ° C (≈ 300K) was carried out. For contrast reflecting of the ability to control the combustion process by changing the parameter - degree rotation of the flow, the regime conditions are same.

They include - thermal load, heat exchange with the environment, coefficient of excess air.

Results of the detailed study in the range of variation of the degree of rotation are shown in figures 5, 6 and 7.

They are charts of the vortex lines or contours, where the mass flow is constant in the half of the combustion chamber.

X axis is axial direction of the combustion chamber and Y axis is radial direction.

The units on both axes are number of integration steps.

The size of the integration steps are: axis z - d_z = 250mm, and axis r - d_r = 15mm.

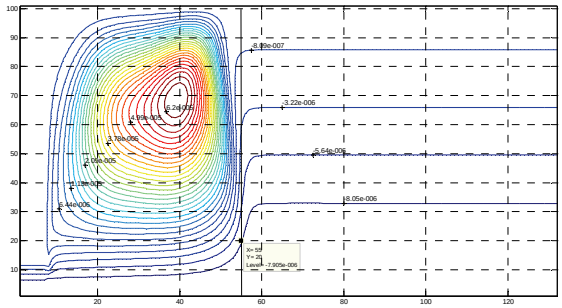


Fig. 5a Contours of the function ψ at S=0

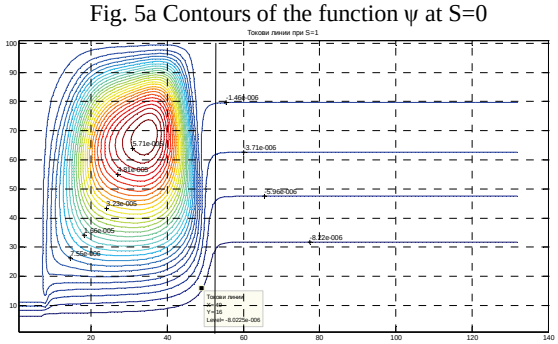


Fig. 5b Contours of the function ψ at S=1

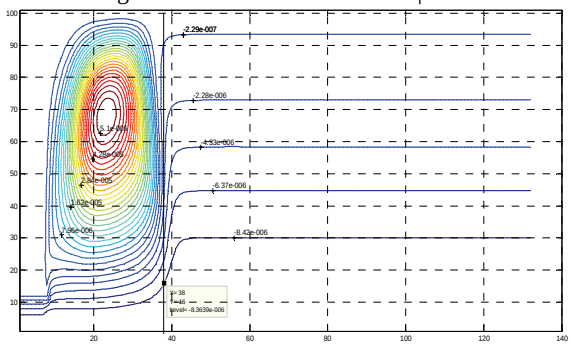


Fig. 5c Contours of the function ψ at S=2

Temperature fields in the combustion chamber at various cases of rotation of the burning gas torch are also displayed.

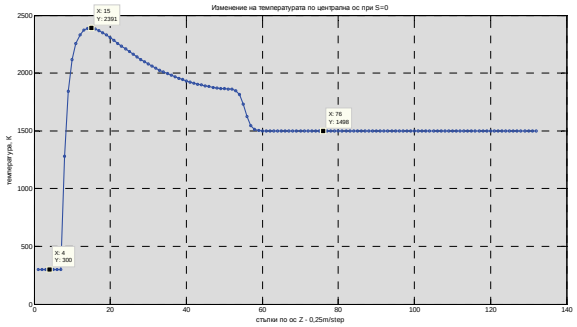


Fig. 6a. A temperature change in the central axis of the combustion chamber at S=0.

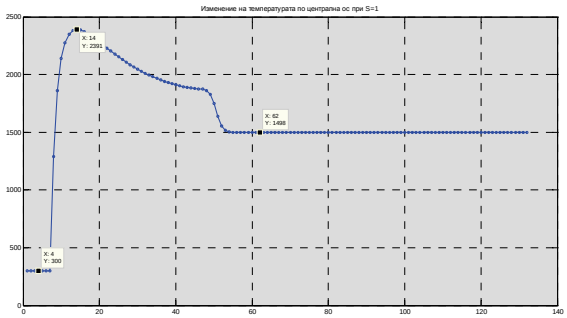


Fig. 6b. A temperature change in the central axis of the combustion chamber at S=1.

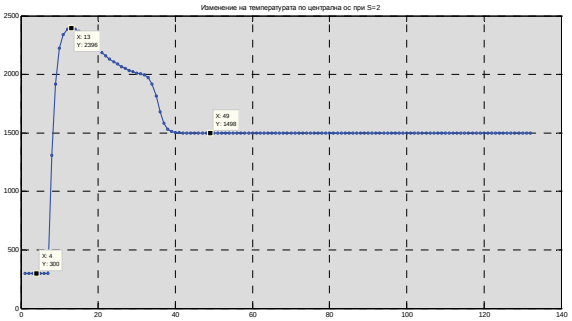


Fig. 6c. A temperature change in the central axis of the combustion chamber at S=2.

The following figures show us the distribution of isothermal surfaces in longitudinal section of the half of the combustion chamber.

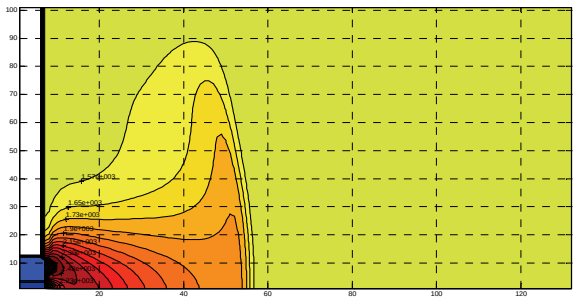


Fig. 7a Contours of the temperature change along the length of the combustion chamber at S= 0, T_a=573,15K.

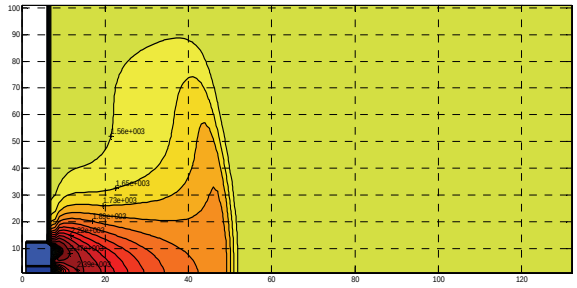


Fig. 7b Contours of the temperature change along the length of the combustion chamber at S= 1, T_a=573,15K.

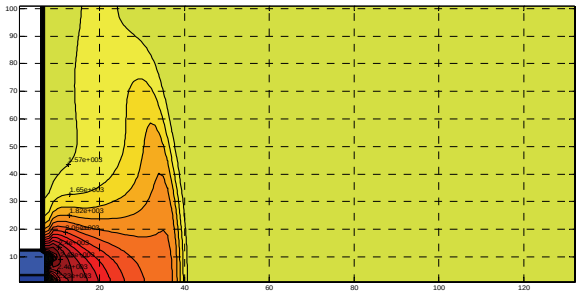


Fig. 7c Contours of the temperature change along the length of the combustion chamber at S= 2, T_a=573,15K.

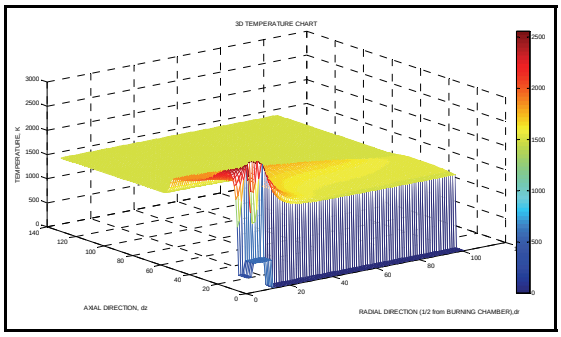


Fig. 8 3D view of the temperature distribution in the chamber.

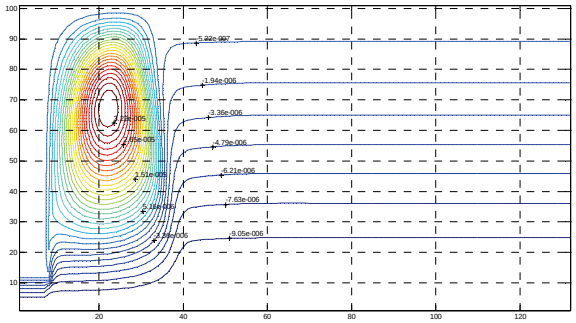


Fig. 9a Contours of the function ψ at S=0 and T_a=300K.

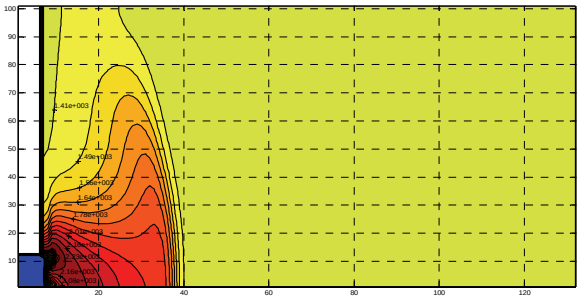


Fig. 9b Contours of the temperature change along the length of the combustion chamber at S=0 and T_a=300K.

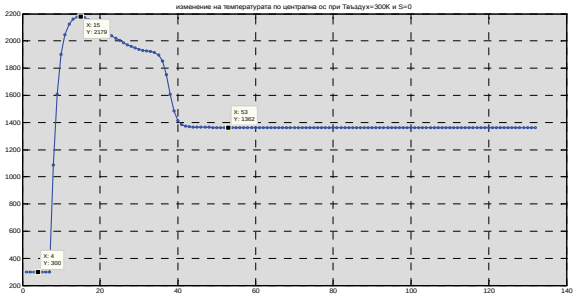


Fig. 9c A temperature change in the central axis of the combustion chamber at S=0 and T_a=300K.

IV. Conclusion

1. With the help of numerical simulation temperature fields in the combustion zone are obtained.

2. By increasing the degree of rotation the length of the torch is reduced. There is also a change of the position of the high-temperature core.

3. The degree of rotation except the intensification of the combustion process can successfully be used for its management, according to the requirements of the technological process.

4. By changing the air temperature, fed into the combustion chamber, additionally on the geometrical and thermal characteristics of the torch are affected.

References

[1] Bhad T.P, etc, CFD modeling of a cement kiln with multi channel burner for optimization of flame profile. Seventh International Conference on CFD in the Minerals and Process Industries, CSIRO, Melbourne, Australia 9-11 December 2009.

[2] Вознесенский А. А., Тепловыеустановки в производствестроительных материалов и изделии. ГИСАС, Москва, 1978.

[3] Гиндбург Д. Б., Печи и и сушки силикатной промышлености. ГИЛСМ, Москва, 1969.

[4] Антонов, Ив. С., Изследване на завъртени турбулентни струи, Автореферат на дисертация за получаване на степен “доктор”, С., 1976.

[5] Kostov P., Atanasov K., Computer Simulation of the Combustion of Gas Fuel in a Limited Turned Flow, “Energetics” Magazine, issues 6-7, 2002

[6] Locher, G., “Mathmatical models for the cement clinker burning process. Part 2: Rotary kiln”, ZKG Int., 55 (3), 68-80,2002.

[7] Lin Bao, Modeling, Identification & Control of Cement Kiln, Technical University of Denmark, DK-2800 Lyngby, Denmark.

[8] Patankar, S.V. Numerical heat transfer and fluid flow, Hemisphere Publishing Corporation, 1980.

Damping Low-frequency Oscillations by Three-channel Power System Stabilizer PSS4B

Nikolay Nikolaev[1], Yulian Rangelov[2], Konstantin Gerasimov[3]

***Abstract* – It is a well-known fact the power oscillations in electric power systems consist of many frequencies. The classic type power system stabilizers PSS2A and 2B each have one phase shift block and thus their optimal settings are around a certain frequency meaning that they cannot damp the local and the inter-area oscillation at the same time. In regard to this issue the multiband power system stabilizers, like PSS4B, were developed. This paper reviews the capabilities of the modern three-channel PSS. A comparative analysis in cases with PSS2A either PSS4B is made and graphical results are present.**

***Keywords* – Power system stabilizer, input filters, multiband power system stabilizers, electric power system.**

I. INTRODUCTION

Because the automatic voltage regulator (AVR) takes into account only the generator terminal voltage, it is possible that they have bad influence on the generator damping. This drawback can be compensated if other input quantities are considered. This approach can not only neutralize this negative AVR influence but even increase the synchronous generator damping coefficient and therefore improve the stability. This is the main idea of the power system stabilizer. Additional signals can be obtained from quantities for instance the change in the rotor speed ($\Delta\omega$), in the generator voltage frequency (Δf), or in the electrical power (ΔP_E) [2,4,5,6].

The block diagram of the main PSS elements is shown in Fig. 1. Special sensors for speed, frequency or power transform the measured quantities into controlling voltage. After that its phase is shifted in such a way that it can appropriately compensate the time delay of the generator and the excitation system. The obtained signal is amplified to the required level and is being limited by the end module, if it is necessary.

[1]Nikolay Nikolaev, Faculty of Electrical Engineering, Electric Power Engineering department, Technical University of Varna, Bulgaria, E-mail: n.nikolaev86@gmail.com

[2]Yulian Rangelov, Faculty of Electrical Engineering, Electric Power Engineering department, Technical University of Varna, Bulgaria, E-mail: j.rangelov@tu-varna.bg

[3]Konstantin Gerasimov, Faculty of Electrical Engineering, Electric Power Engineering department, Technical University of Varna, Bulgaria, E-mail: kosio_gerasimov@abv.bg

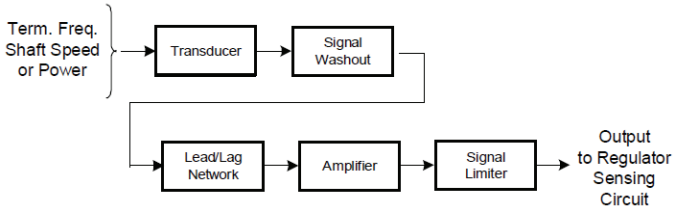


Fig. 1. Main PSS blocks

System stabilizers of type PSS4B are structurally based on three separate frequency bands – for mode oscillations with low, medium and high frequencies.

The low frequency band usually is specific for global system oscillations, the medium frequency band – for inter-area oscillations, and the high frequency band – for local oscillations. Each of the three channels is consisted of differential filter, amplifier and limiter. Their outputs are summed and are entered into the end limiter V_{STMIN}/V_{STMAX}, forming the final output signal V_{ST} of the stabilizer. Its structural scheme is shown in Fig. 2.

PSS4B measures the speed deviation in two different ways. $\Delta\omega_{L\text{-}I}$ acts as input signal in the low and medium frequency channels while $\Delta\omega_H$ is entered into the high frequency channel. The equivalent model of those two speed sensors is shown in Fig. 3. Additional notch filters with transfer function $N_i(s)$ for regulating the level, can be used for the torsional modes of the steam turbine generators. They are described with the following equation [5,6]:

$$N_i(s) = \frac{s^2 + \omega_{ni}^2}{s^2 + B_{wi}.s + \omega_{ni}^2} \qquad (1)$$

where ω_{ni} is the frequency of the filter, and B_{wi} is its frequency band at 3 dB.

Although the parameters of the differential filters can be tuned in different ways, a simple method for tuning is based on three symmetrical band filters, correspondently set to frequencies F_L, F_I and F_H. The time constants and gains of the separate channels are obtained from equations (2), (3), (4) and the equation (5) for low frequency band [7]. It is necessary that only six parameters are known – F_L, F_I, F_H, K_L, K_I, K_H.

$$T_{L2} = T_{L7} = \frac{1}{2\pi F_L \sqrt{R}} \qquad (2)$$

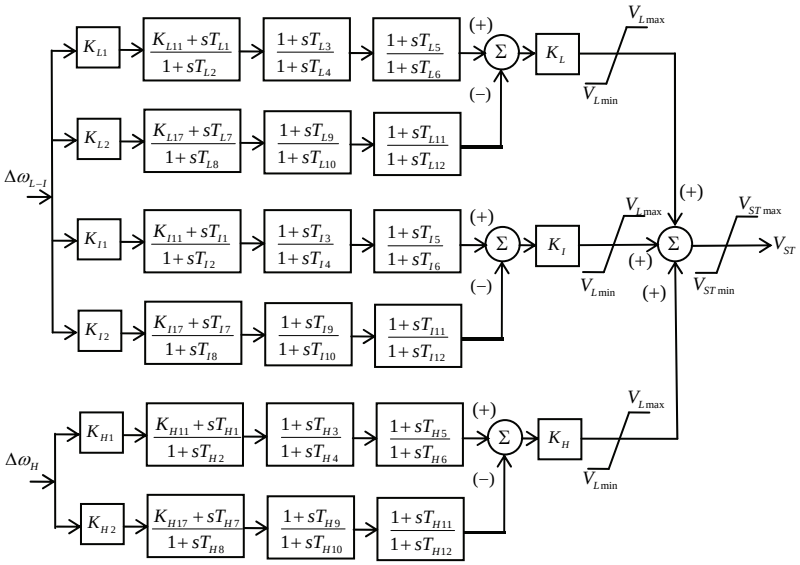


Fig. 2. Block diagram of PSS4B

$$T_{L1} = T_{L2} / R \tag{3}$$

$$T_{L8} = T_{L7} \cdot R \tag{4}$$

$$K_{L1} = K_{L2} = (R^2 + R) / (R^2 - 2R + 1) \tag{5}$$

where R is a constant. Such relations are also valid for the other two channels. A total of 24 parameters are selected.

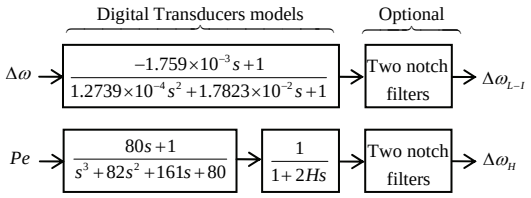


Fig. 3. Block diagram of the sensors for change of the speed

II. Problem Formulation

In Fig. 4 is shown the structure of the analyzed united electric power system (EPS) [2,8]. The first EPS is represented in more details with two zones, and the second – generalized. The connection between the two zones is made by long 400 kV power line (W5). The task is to analyze and compare the electromechanical oscillations of G4 without PSS, with PSS2A and with PSS4B.

The data for the steady state and the circuit parameters are as follows:

1) Generators (G_1 - G_4): $U_н$=15,75 kV; $P_н$=280 MW; x_d=2,19; x_q=2,1; x'_d=0,34; x'_q=0,54; x''_d=0,25; x''_q=0,25; x_l=0,2; r_a=0,003; T'_{d0}=4,54s; T'_{q0}=0,38s; T''_{d0}=0,031s; T''_{q0}=0,068s; T_J=11,12s.
Generator G_5: $U_н$=400 kV; $P_н$=2000 MW; x''_d=0,05; r_a=0; T_J=6s; D=1.

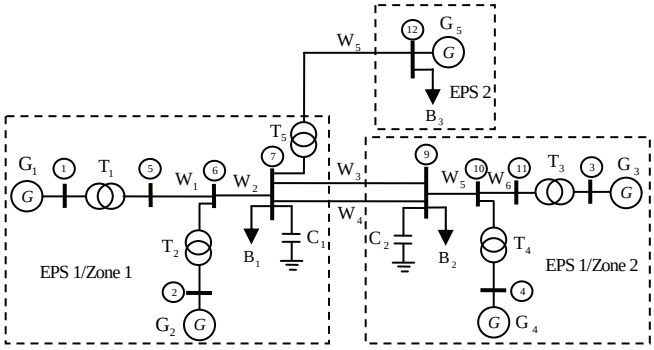


Fig. 4. Structure of the analyzed united EPS

2) Power lines

TABLE I. POWER LINES' PARAMETERS

Nodes	$U_н$, kV	R, p.u.	X, p.u.	B, p.u.
5-6; 10-11	400	0,0025	0,025	0,04375
6-7; 9-10	400	0,001	0,01	0,0175
2x(7-9)	400	0,037	0,37	0,385
7-12	400	0,0025	0,25	0,044

3) Transformers

TABLE II. TRANSFORMERS' PARAMETERS

Nodes	$U_{н1}$, kV	$U_{н2}$, kV	$S_н$, MVA	R, p.u.	X, p.u.
5-1; 6-2; 10-4; 11-3	400	15,75	295	0	0,13

4) Excitation system with regulator type UNITROL from ABB (Fig. 5): T_r=0,02s; T_S=0,003s; K_R=350; T_{c2}=1s; T_{b2}=1s; T_{c1}=5s; T_{b1}=6,6s.
5) PSS – IEEE PSS2A (Fig. 6): TW_1=TW_2=TW_3=2s; TW_4=0s; T_6=0s; T_7=2s; KS_2=0,18; KS_3=KS_4=1; T_8=0,5s; T_9=0,1s; N=1; M=5; KS_1=10; T_{L1}=0,2s; T_{L2}=0,03s; T_{L3}=0,1s; T_{L4}=0,02s;

6) PSS - IEEE PSS4B:
K_{L1}=7,7; K_{L2}=0; K_{L11}=0; K_{L17}=0; T_{L1}=5s; T_{L2}=5s; T_{L3}=0,36s; T_{L4}=1,57s; T_{L5}=0,37s; T_{L2}=1,21s; K_L=11,14s;
K_{I1}=6,5; K_{I2}=0; K_{I11}=0; K_{I17}=0; T_{I1}=0,4s; T_{I2}=0,4s; T_{I3}=0,19s; T_{I4}=0,38s; T_{I5}=0,1s; T_{I2}=0,09s; K_I=25,6;
K_{L1}=10,5; K_{L2}=0; K_{L11}=0; K_{L17}=0; T_{L1}=0,03s; T_{L2}=0,03s; T_{L3}=0,096s; T_{L4}=0,021s; T_{L5}=0,09s; T_{L2}=0,01s; K_L=49,2.

Regime parameters of the analyzed EPS

1) Complex loads and capacitive power of the shunt condensers: P_{B1}=320 MW; Q_{B1}=133 MVAr; Q_{C1}=60 MVAr; P_{B2}=600 MW; Q_{B2}=133 MVAr; Q_{C2}=50 MVAr.

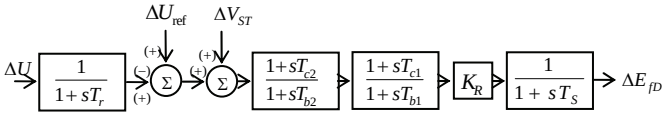


Fig. 5. AVR type UNITROL

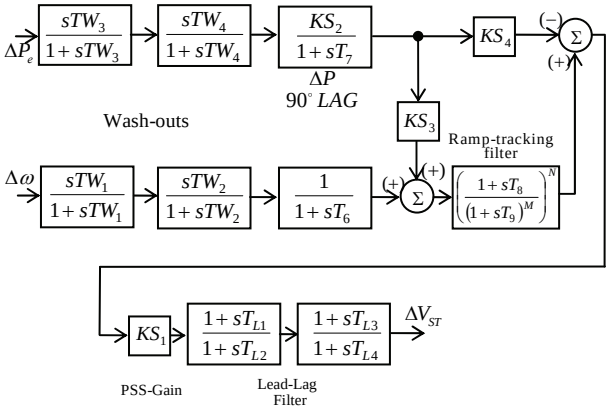


Fig. 6. System stabilizers type PSS2A

2) Generator regime

TABLE III. GENERATOR REGIME PARAMETERS

Generators	*P*, MW	*Q*, MVAr	*U*, p.u.
G_1	245,00	77,479	1,03
G_2	245,00	29,644	1,01
G_3	719	74,224	1,03
G_4	700	65,590	1,01
G_5	-50,452	24,304	1,00

III. RESULTS

In order to assess the PSS effect, the specialized software tool NASAVR [1,2] is used. It is developed by a team from department Electric Power Engineering in Technical University of Varna. NASAVR operates in the MATLAB and Simulink environment and is capable of tuning automatic voltage regulators and power system stabilizers of synchronous generators in large electric power unions considering the influence of the specific generator unit parameters and of the EPS, to which it is connected.

In Fig. 7 is shown the generalized assessment for the transient processes quality by the means of the H_∞ norm for the three different cases.

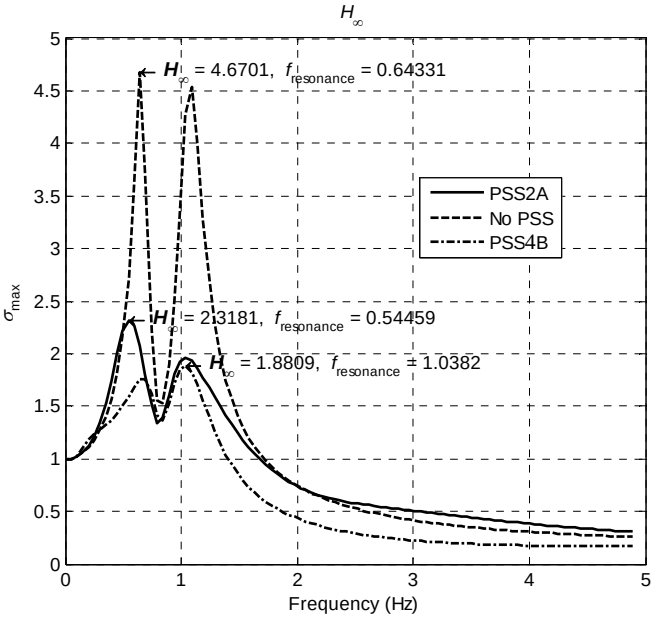


Fig. 7. Generalized assessment of the transient processes quality - H_∞

The analyzed system is constructed in such a way so that distinct oscillations between the generators in the separate zones are observed with frequency around 0,63 Hz. An oscillation with frequency of around 1 Hz (the local mode) is also observed. The installation of PSS leads to successful damping of those oscillations and PSS4B copes equally well with the local and the inter-area oscillations because of its specific design. Another important aspect is that its settings can be optimized for damping inter-area oscillations without worsening its influence on the local oscillations. For PSS2A this is possible but only to a certain extent.

Fig. 8 shows the frequency response of the change of the voltage, the active power and the rotor angular speed for unit step disturbance of the AVR reference. As expected, good damping of the electromechanical oscillations is observed in case of PSS presence.

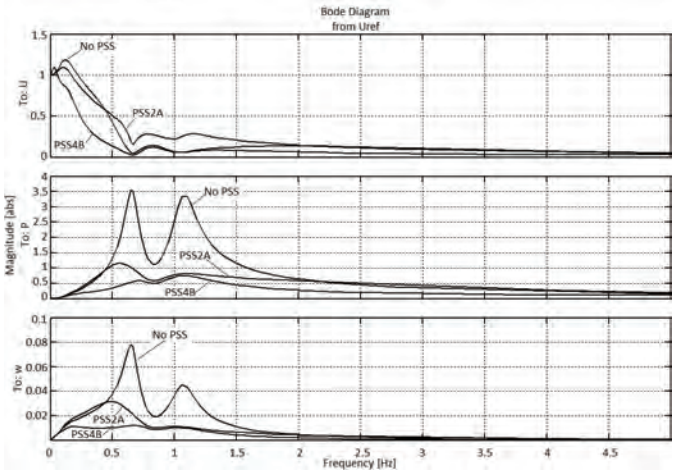


Fig. 8. Frequency response of the change of *U*, *P* and ω for unit step disturbance in the AVR reference

The step response, shown in Fig. 9, demonstrates the transient processes progress in time of U, P and ω for change of the AVR reference (ΔU_{ref}) with one percent.

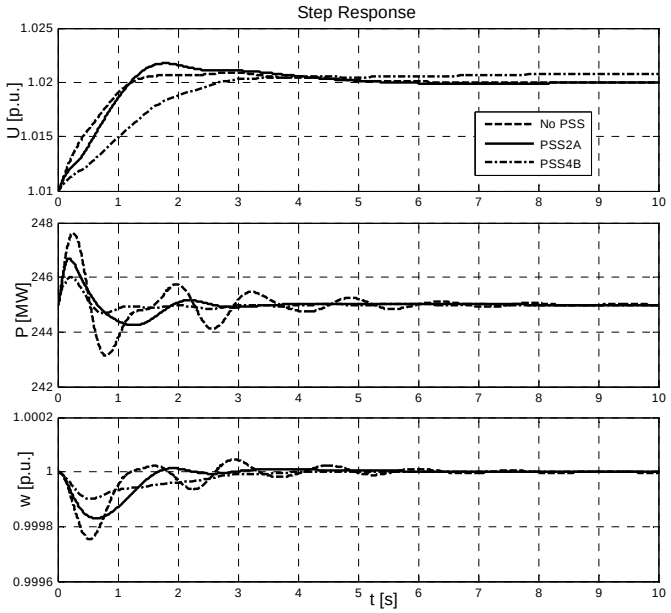


Fig. 9. Step response of U, P and ω for ΔU_{ref} =1%

The clearly expressed oscillations of the regime parameters of the generator without PSS are successfully damped when system stabilizer is activated. The best results are obtained with the use of PSS4B.

IV. Conclusion

From the conducted comparative analysis the following conclusions can be made:

1. The use of modern system stabilizers significantly improves the transient processes quality at normal parallel operation of the generators in EPS and even makes them obligatory for generating units with bigger power;

2. The creation of large electric power system unions favors the occurrence of low frequency inter-area and inter-zone oscillations, which can be successfully damped with PSS4B without decrease of the local oscillations damping in the specific machine;

3. In all conducted tests PSS4B behaves better and is far more flexible in terms of tuning;

4. The use of special software for modal analysis of the processes in EPS and for PSS tuning enables the accurate determination of problematic synchronous generators in EPS and appropriate tuning of their stabilizers.

References

[1] Gerasimov, K., Y. Rangelov, Ch. Ivanov, Y. Kamenov. MATLAB Based Software for AVR and PSS Tuning. Acta Universitatis Pontica Euxinus, Constanta, Romania, Vol. II, №2, 2005, pp. 145-150.

[2] Герасимов, К. Математическо описание на електромеханичните движения в електроенергийната система и алгоритми за изследване на устойчивостта им при малки смущения. Дисертационен труд за придобиване на научна степен доктор на науките. ТУ-Варна, 2006.
(Gerasimov, K. Mathematical modeling of electromechanical transients in electric power systems and algorithms for analysis of stability at small disturbances. Dissertation to acquire the academic degree of Doctor of Science, TU-Varna, 2006)

[3] Нотов, П., К. Герасимов. Преходни процеси в електроенергийните системи. С. 1997.

[4] Нотов, П., Кр. Герасимов. Стандартни IEEE модели на възбудителни системи на синхронни генератори. сп. Енергетика - 8', 2007 год., 34 – 44 стр.

[5] “IEEE Tutorial Course Power System Stabilization Via Excitation Control” Sponsored by IEEE Power Engineering Society Life Long Learning Committee and the Excitation Systems Subcommittee of the Energy Development and Power Generation Committee

[6] “IEEE Recommended Practice for Excitation System Models for Power System Stability Studies”, IEEE Standard 421.5 - 2005.

[7] B. Sumanbabu, S. Mishra, B. K. Panigrahi, and G.K. Venayagamoorthy. Robust Tuning of Modern Power System Stabilizers Using Bacterial Foraging Algorithm

[8] Kundur P. Power system stability and control. McGraw-Hill Inc. 1993.

Influence of the Settings of PSS2A and 2B Input Filters over the Damping of Low-frequency Power Oscillations

Yulian Rangelov[1], Konstantin Gerasimov[2], Joncho Kamenov[3], Krum Gerasimov[4]

***Abstract* – In the electric power system of Bulgaria power system stabilizer type PSS2A and 2B are most common. They were installed during the preparation period for connection with the European electric power system. Usually the settings of these stabilizers refer to the gain and time constants of the phase shift block. This paper proves that the settings of the input filters can influence the damping of local and mostly of inter-area oscillations, which is a real problem after the actual connection of the power systems. An algorithm for calculation of these settings is proposed and results for single-machine and multi-machine sample system are discussed.**

***Keywords* – Power system stabilizer, input filters, electric power system.**

I. INTRODUCTION

In order to highlight the influence of the input filters (Fig. 1) of PSS-2A and 2B on the mechanical oscillations damping, an algorithm for calculation of their settings is developed.

The blocks with time constants from TW_1 to TW_4 form two input filters (wash-outs) – one for the power channel and one for the frequency channel. The purpose of these filters is the prevent the PSS activation in cases of normal continuous deviations of generator's active power (P) and rotor speed (ω). The blocks with time constants T_7 and T_6 shift the signal phase with lag up to 90°. The purpose of this is to obtain at the end of the power channel a signal which is proportional to the rotor speed deviation $\Delta\omega$. The electric unbalance power ΔP which enters the channel can approximately be described with the relation [2]:

$$\Delta P \approx -T_J \cdot \frac{d\Delta\omega}{dt}, \text{ т.е. } \Delta\omega \approx \frac{-\Delta P}{s \cdot T_J}, \quad (1)$$

from which for $T_7 \approx T_J$ the output signal of the block is proportional to minus $\Delta\omega$. This signal is then added to the incoming signal for ω. The resultant signal enters in the filter for torsional oscillation (ramp-tracking filter). For steam turbine generators, due to torsion forces in the shafts, rotor oscillations occur in the sub harmonic range of few tens Hz. Depending on the place of measurement of ω deviations, these oscillations superpose with the low-frequency machine's rotor oscillations (the rotor is treated as a single mass object). If these oscillations are amplified and get into the excitation, high-frequency oscillations will occur and as a result the rotor can be damaged. The filter passes only inter-area oscillations. The filter output is summed with the output of ΔP channel and as a result a signal proportional to the inter-area deviation of ω is formed. Precisely this signal is entered input of the PSS gain, i.e. the block with KS_1. This way, regardless that there is also ΔP channel, the PSS reacts to the inter-area deviation of ω.

The literature most often discusses only the filters settings for passing a certain frequency range. Here, besides this requirement, also is required that the filters output signal $\Delta\omega_e$ is in phase the $\Delta\omega$ of the unit. This is a very important requirement, especially when in PSS2A is entered only generator bus voltage frequency deviation Δf. If the output signal $\Delta\omega_e$ is not in phase with $\Delta\omega$, then a wrong phase compensation will be obtained, even though the phase shifting PSS block are tuned correctly.

II. ALGORITHM FOR CALCULATION OF APPROPRIATE FILTER SETTINGS

In order to meet the requirement for close phases of $\Delta\omega_e$ and $\Delta\omega$, the following optimization task is formed.

$$\min_{\mathbf{\Pi}} \gamma_f(t), \text{ for } \begin{vmatrix} \gamma_f(t) = (\Delta\omega_e(\mathbf{\Pi}, t) - \Delta\omega(t)) - g_f(t); \\ \gamma_f(t) \le 0; \\ \mathbf{\Pi}_Д \le \mathbf{\Pi} \le \mathbf{\Pi}_Г, \end{vmatrix} \quad (2)$$

where $\mathbf{\Pi}_Д$ and $\mathbf{\Pi}_Г$ are respectively the lower and upper limits of deviation the parameters to be $\mathbf{\Pi}$={TW_1, TW_2, TW_3, TW_4, T_6, T_7, T_8, T_9, KS_2, KS_3, KS_4, M, N}; $\Delta\omega_e(\mathbf{\Pi},t)$ and $\Delta\omega(t)$ – respectively the change in time of the filters output signal $\Delta\omega_e$ and generators rotor speed $\Delta\omega$ caused by disturbances of the AVR - ΔU_{ref}; $g_f(t)$ – limit function which theoretically converges to zero if there is complete phase and amplitude overlapping of $\Delta\omega_e$ and $\Delta\omega$.

To solve the optimization task (2), a computing scheme is constructed in the MATLAB application Simulink, whose structure is shown in Fig. 1. The solution is carried out by the application Simulink Design Optimization (SDO).

[1]Yulian Rangelov, Faculty of Electrical Engineering, Electric Power Engineering department, Technical University of Varna, Bulgaria, E-mail: j.rangelov@tu-varna.bg

[2]Konstantin Gerasimov, Faculty of Electrical Engineering, Electric Power Engineering department, Technical University of Varna, Bulgaria, E-mail: kosio_gerasimov@abv.bg

[3]Joncho Kamenov, Faculty of Electrical Engineering, Electric Power Engineering department, Technical University of Varna, Bulgaria, E-mail: kosio_gerasimov@abv.bg

[4]Krum Gerasimov, Faculty of Electrical Engineering, Electric Power Engineering department, Technical University of Varna, Bulgaria, E-mail: k.gerasimov@tu-varna.bg

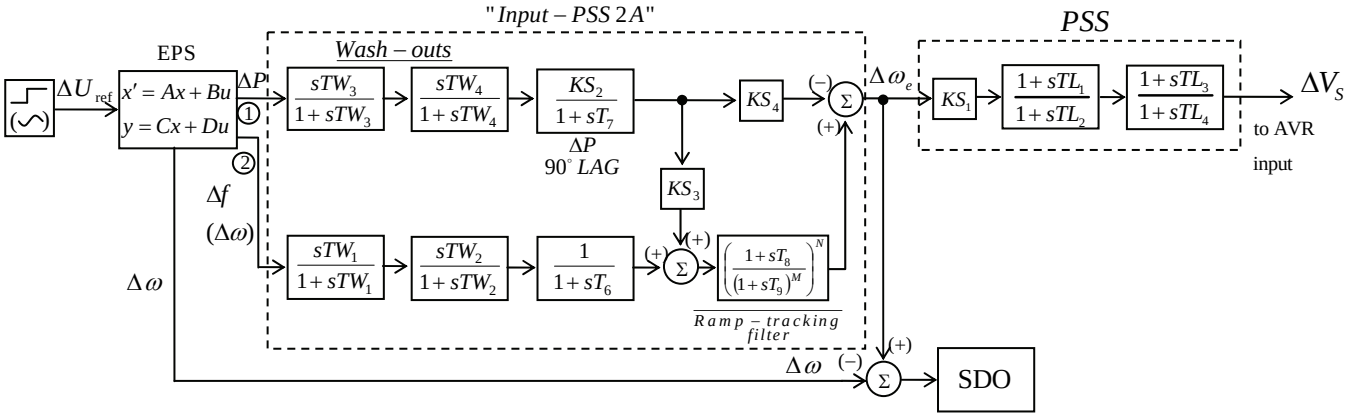


Fig. 1. Structure of the Simulink model used to calculation of PSS2A filters settings

The EPS (Electric Power System) block represents the model the electric power system in state space [1]. The disturbances are introduced through change of the reference ΔU_{ref} of the AVR in the generator whose PSS is being tuned. The disturbance can be step or if the dominant modes of mechanical oscillations for the specific generator are known the disturbance can be generated as a sum of sinusoids with the frequencies of the dominant modes. At first the limits $g_f(t)$ are set to about ± 0,005. In the process of the optimization procedure they are changed until the possibilities for their satisfaction are exhausted. As a result the quantities of **Π** which get $\Delta\omega_e$ maximally close to $\Delta\omega$ are obtained.

The stated above is basis for the formulation of the following algorithm for tuning of "Input – PSS 2A":

1. Construction of the computing scheme (Fig. 1) in Simulink.
2. Use of SDO to solve the optimization task (2) for the smallest limits $g_f(t)$. As a result the quantities of **Π** which get $\Delta\omega_e$ as close as possible to $\Delta\omega$ are obtained.
3. The calculated **Π** settings are used to generate the frequency response characteristics of the input filters "Wash - outs" and the torsional filter "Ramp - tracking". Their frequency bandwidths are checked if they meet the requirements. The input filters must not pass oscillations with frequencies under 0,01 Hz, i.e. they must not pass through the slow and continuous deviations in the regime parameters. The torsional filter must not pass frequencies in the 5÷50 Hz range, where the torsional oscillations are. If these requirements are fulfilled, the calculated settings are treated as final. Otherwise step 4.
4. The filters are tuned in respect to the frequency band requirements. After that step 2 is carried out again but with reduced number of parameters to be tuned **Π**={T_6, T_7, KS_2, KS_3, KS_4}. The obtained parameters settings are final.

III. Test Results

For illustration of the proposed algorithm results are shown for the calculation of the settings of "Input-PSS 2A" of generator G_1 of the tested in [1] EPS. The calculated optimal settings are: TW_1=3,66 s, TW_2=3,66 s, TW_3=2,07 s, TW_4=0 s, T_6=0 s, T_7=2,07 s, T_8=0,122 s, T_9=0,193 s, KS_2=0,16 o.e., KS_3=1,03 o.e., KS_4=1,1 o.e., N=1,M=5. In Fig. 2 are shown the frequency response of $\Delta\omega_e$ and $\Delta\omega$, and in Fig. 3 – their step response. It is clearly seen goal phase agreement of the two signals is achieved.

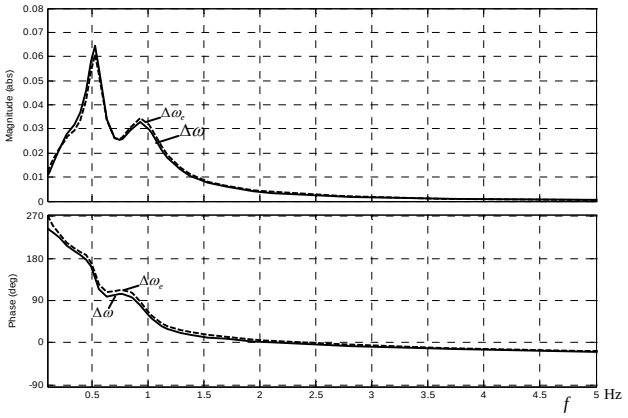


Fig. 2. Bode diagram of $\Delta\omega_e/\Delta U_{ref}$ and $\Delta\omega/\Delta U_{ref}$

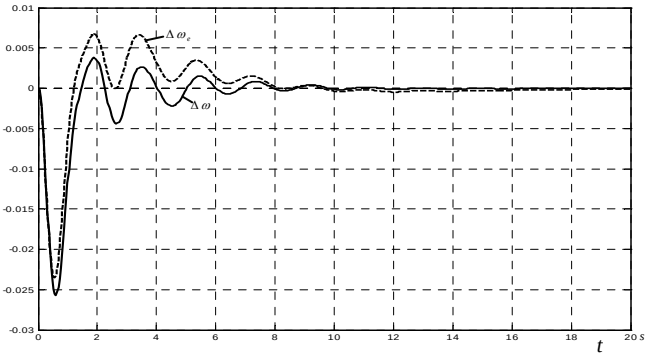


Fig. 3. $\Delta\omega_e$ and $\Delta\omega(t)$ for step change of ΔU_{ref} with 2 %

In Fig. 4, Fig. 5 and Fig. 6 are shown the frequency responses of the input filters of inputs 1 and 2 and of the torsional filter. It can clearly be seen the requirements for the pass-through frequency bands of the filters are satisfied. Therefore the calculated settings in step 2 of the algorithm are final.

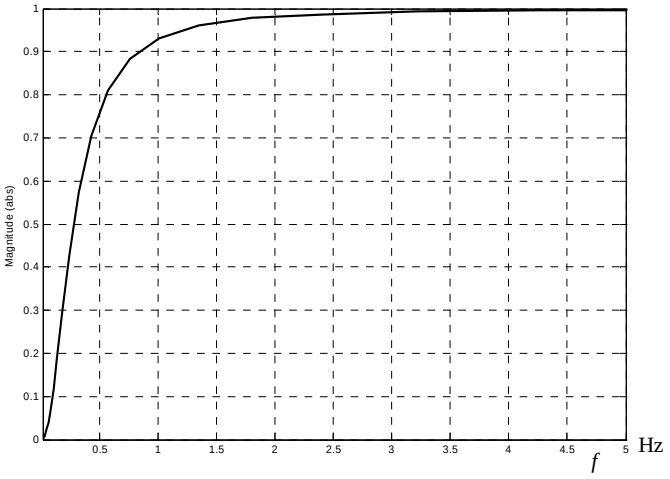


Fig. 4. Frequency response of the input filter of PSS 2A input 1

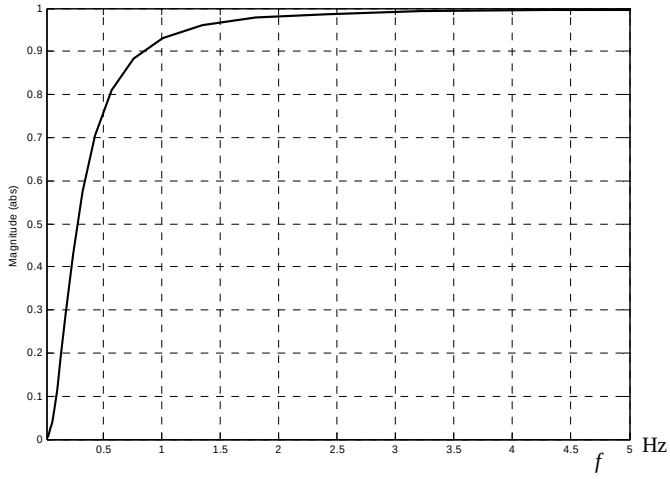


Fig. 5. Frequency response of the input filter of PSS 2A input 2

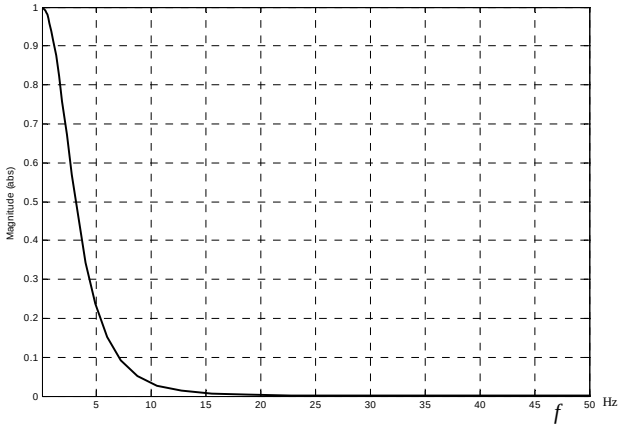


Fig. 6. Frequency response of the PSS 2A torsional filter

In Fig. 7 and 8 are shown results revealing the difference in the transient processes quality for different values of the input filters parameters. With dashed line are depicted the results for the settings: TW_1=2 s, TW_2=2 s, TW_3=2 s, TW_4=0 s, T_6=0 s, T_7=2 s, KS_2=0,18 p.u., and with solid line – the settings: TW_1=10 s, TW_2=10 s, TW_3=10 s, TW_4=0 s, T_6=0 s, T_7=10 s, KS_2=0,9 p.u.

The generalized assessment of the transient processes quality [1,3], given by the H_∞ norms for both cases, are shown in Fig. 7. They, together with the step response of the regime parameters U, P and ω for step change with 1% of the AVR reference (Fig. 8) shows how opening the filters for passing through frequencies under 1 Hz enables the PSS to damp them.

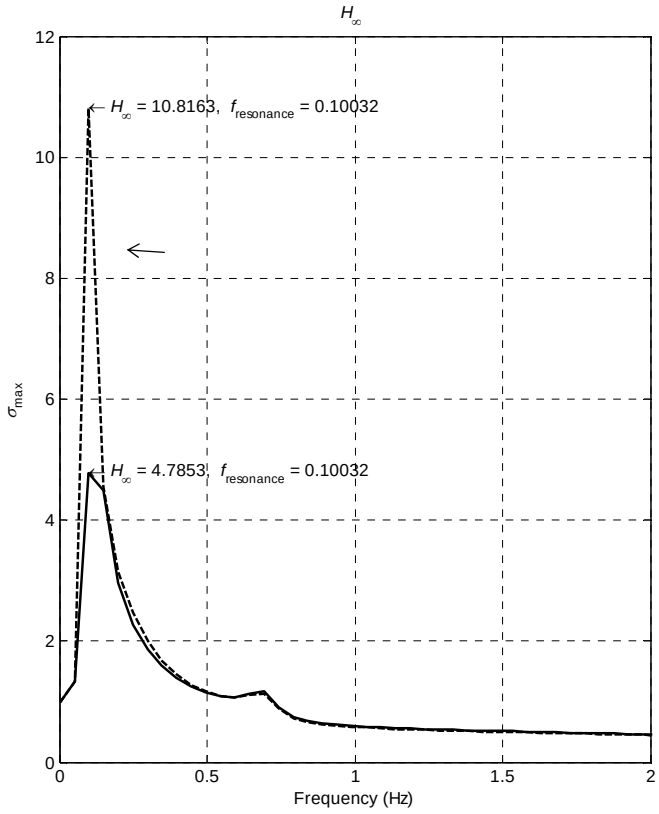


Fig. 7. Transient processes quality generalized assessment by the means of H_∞ norm

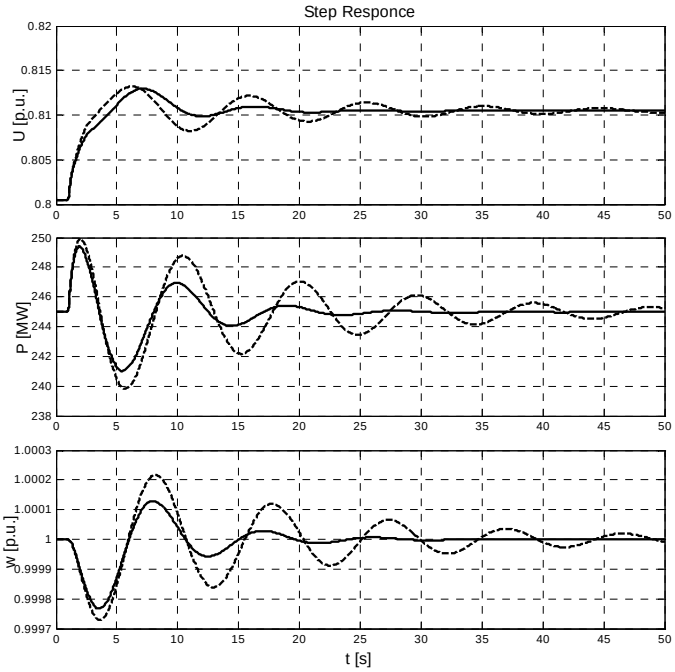


Fig. 8. Step response of the regime parameters U, P and ω for step change with 1% of the AVR reference

IV. CONCLUSION

The power system stabilizers type PSS2A and 2B successfully damp the local low-frequency electro-mechanical oscillations. The creation of large electric power systems consolidations lead to higher requirement for these devices, namely to damp successfully simultaneously the local and inter-area oscillations. Because the latter and in the frequency range under 1 Hz it is a must to tune appropriately the input filters so that they can pass them through to the PSS.

REFERENCES

[1] Герасимов, К. Математическо описание на електромеханичните движения в електроенергийната система и алгоритми за изследване на устойчивостта им при малки смущения. Дисертационен труд за придобиване на научна степен доктор на науките. ТУ-Варна, 2006.
(Gerasimov, K. Mathematical modeling of electromechanical transients in electric power systems and algorithms for analysis of stability at small disturbances. Dissertation to acquire the academic degree of Doctor of Science, TU-Varna, 2006)

[2] Петков, П., М, Константинов. Робастни системи за управление. Анализ и синтез с MATLAB. АВС Техника, София, 2002.
(Petkov, P., Konstantinov, M., Robust control systems. Analysis and synthesis with MATLAB. ABC Tehnika, Sofia, 2002)

[3] "IEEE Tutorial Course Power System Stabilization Via Excitation Control" Sponsored by IEEE Power Engineering Society Life Long Learning Committee and the Excitation Systems Subcommittee of the Energy Development and Power Generation Committee

[4] "IEEE Recommended Practice for Excitation System Models for Power System Stability Studies", IEEE Standard 421.5 - 2005.

Modeling of Electromagnetic and Thermal Processes of High-frequency Induction Heating of Internal Cylindrical Surfaces of Ferromagnetic Details

Bohos Aprahamian[1] and Maik Streblau[2]

Abstract **- Currently increasingly widespread the application of the induction hardening of ferromagnetic details, due to the high efficiency and universality of this kind of heat treatment.**

For the most part, the companies which realize this method have limited power capacities and limitations in the frequency range. For these reasons, certain difficulties arise in the induction hardening of cylindrical internal surfaces and achievement of even hardened layer along the detail providing concrete depth.

Optimization of the technological regimes can be effectively done using computer models.

For this purpose two-dimensional model was developed, simultaneously analyzing the electromagnetic and thermal processes, having taken into account their influence on the properties of the detail and with his help we have optimized the parameters of the inducers.

Keywords **- Induction hardening, Hardened layer, Inducer, Computer model, Optimization**

I. INTRODUCTION

The annealing of the internal cylindrical surfaces of ferromagnetic details require the achievement of appropriate speed of heating providing uniform hardened layer depth [1].

Since induction heating machines with lamp generators (440kHz) are still used, often problems related to a shortage of capacity for implementation of the technological regime arise, especially in hardening of internal cylindrical surfaces, due to the ring and proximity effects [2], leading to dissipation of the electromagnetic field around the treated surface of the workpiece.

In regard to the reasons maintioned above interesting is the optimization process for maximum utilization of the power entering the generator through reduction of the electromagnetic dissipation. This paper presents a model of inductor-workpiece system to solve the defined problem.

The Figure 1 presents the inductor-workpiece system configuration. The workpiece is made of structural steel grade C45 [1], [2].

[1]Bohos Aprahamian is with the Faculty of Electrical Engineering, Technical University – Varna, 1 Studentska str., Varna, Bulgaria, e-mail: bohos@abv.bg.

[2]Maik Streblau is with the Faculty of Electrical Engineering, Technical University – Varna, 1 Studentska str., Varna, Bulgaria, e-mail: streblau@yahoo.com.

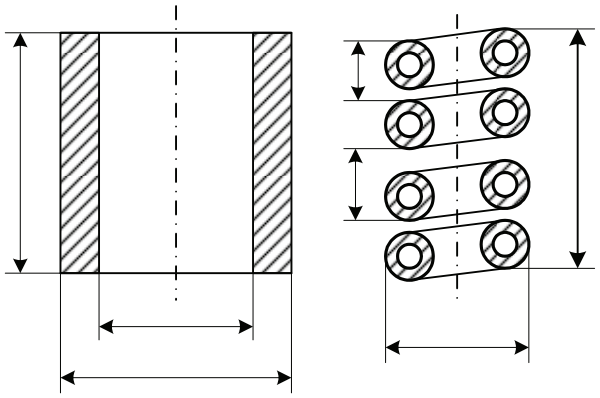


Fig. 1. Dimensions of the workpiece and the inductor

The variety of configurations of the processed details define as uneconomic the process of experimental investigation. Currently the large opportunities of process modeling are widely used [3], [4].

II. THEORETICAL MODEL

To carry out the theoretical investigation two-dimensional axial symmetric model is used, analyzing both the electromagnetic and thermal problem [5].

The simulation of the electromagnetic processes is done by harmonic electromagnetic analysis, described by the following differential equation and boundary condition - zero magnetic potential within the model.

$$\nabla \times \left(\frac{1}{\mu(B,T)} \times \nabla \times \dot{A} \right) + j \cdot \omega \cdot \gamma(T) \cdot \dot{A} = \frac{\gamma \cdot \dot{V}_{coil}}{2 \cdot \pi \cdot r} \qquad (1)$$

As a source is set the voltage, determinant the current density inside the inductor.

The thermal problem is simulated by transient thermal analysis, described by the following differential equation:

$$\rho \cdot c \cdot \frac{\partial T}{\partial t} = \nabla\left(\lambda(T) \cdot \nabla T\right) + q_V \qquad (2)$$

where:

$$q_V = \frac{1}{2} \cdot \frac{\dot{J}^2}{\gamma(T)} = \frac{1}{2} \cdot \omega^2 \cdot \gamma(T) \cdot \dot{A}^2 \qquad (3)$$

The boundary condition, ensuring coherence of equation (2) is the requirement of Dirichlet, specified on the boundary of the model. On the bordering surface of the inducer to the environment a natural convective heat transfer as a function of the temperature and radiant heat transfer are set [6].

In the description of the model the perceived assumptions are as follows:

- the power supply of the inductor-workpiece system is via sinusoidal voltage with constant frequency - 440kHz;
- the inductor is with water cooling and therefore the changes of the electrical conductivity during the heating process are not taken into account;
- the changes in the density and specific heat capacity of the workpiece, depending on the temperature changes do not account in the model building.

III. Theoretical Investigation

With the above model theoretical invstigations were conducted at constant voltage and frequency. The time to realize the heating process was set to 4.5 seconds.

In the figures below are represented graphically the distribution of the magnetic vector potential and magnetic induction in the starting and ending point of time, and the distribution of temperature field in the section and on the surface of the workpiece.

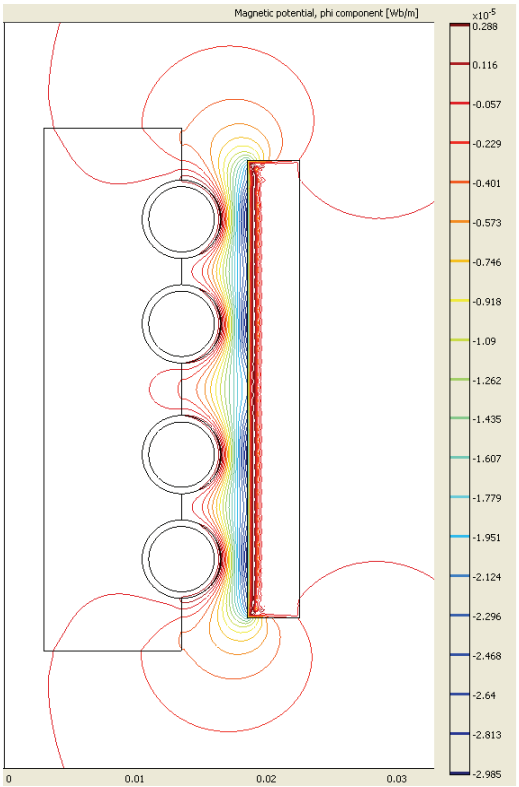


Fig. 2. Distribution of the magnetic vector potential in the beginning of the process

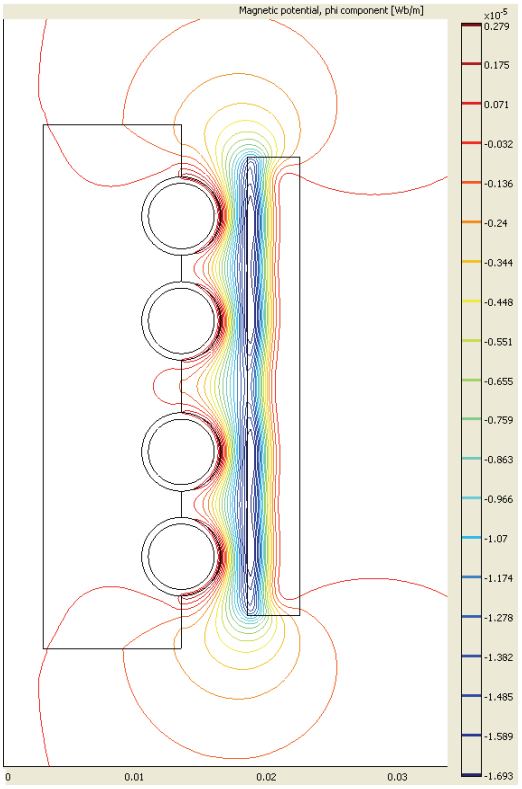


Fig. 3. Distribution of the magnetic vector potential at the end of the process

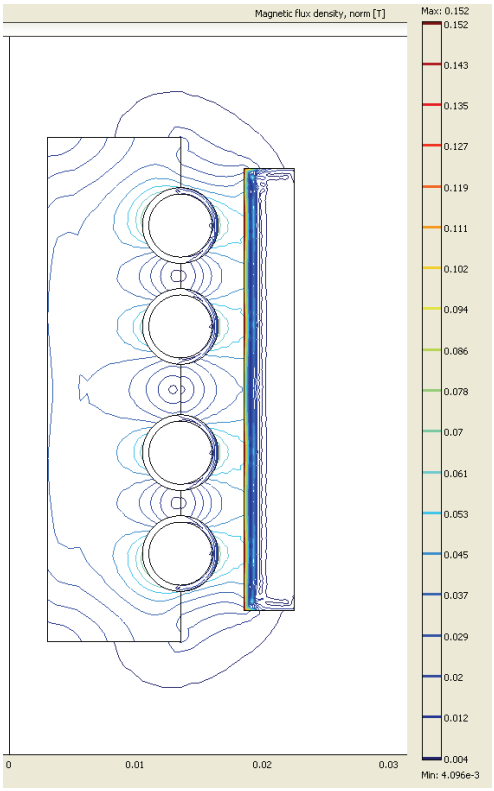


Fig. 4. Distribution of the magnetic induction at the beginning of the process

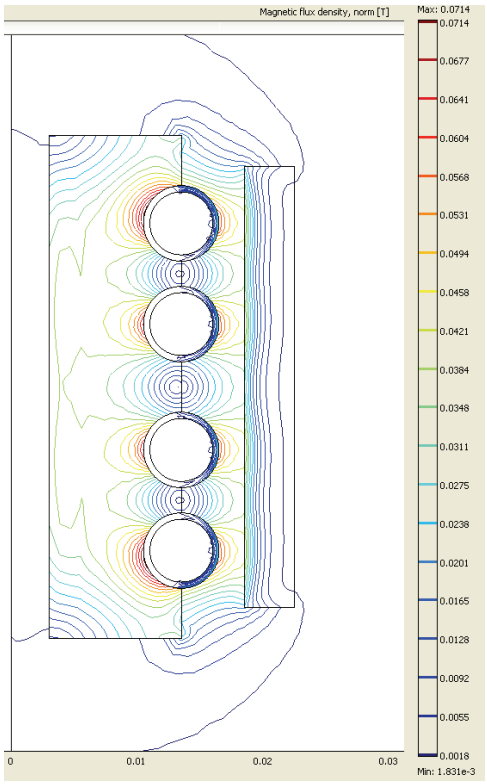


Fig. 5. Distribution of the magnetic induction at the end of the process

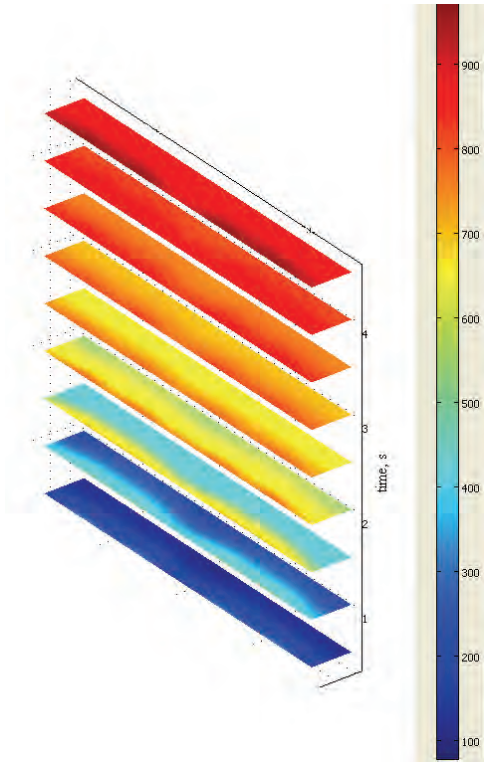


Fig. 6. Temperature distribution along the detail section during the work time

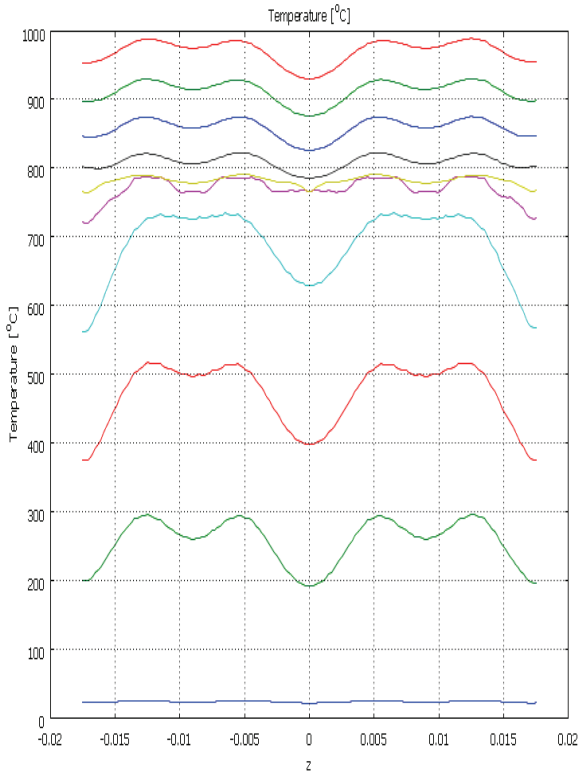


Fig. 7. Distribution of the temperature field on the surface of the workpiece

IV. Conclusions of the Theoretical Investigation

The distribution of the vector magnetic potential and magnetic induction shown in Figures 2 to 5 demonstrates the existence of double layer (with two highly different permeability zones) heat penetration medium under section of the workpiece until in the end of the process the temperature of the workpiece has passed the Curie point and the distribution of the electromagnetic field cover greater layer then the layer of the heat penetration.

Another important point is the achievement of uniform heating of the workpiece at a depth of the set hardening layer - Figures 6 and 7. The temperature difference is less than 50°C throughout the volume of the heated layer.

To ensure high efficiency it is necessary to use ferromagnetic cores in the system inducer - detail.

The ferromagnetic core shape must follow the configuration of the inducer. For this purpose, a better option is the preparation of magnetic core by molding rather than applying standard core configurations.

The used approach makes possible to develop adequate theoretical models for studying the process of induction heating for different configurations of the details and the inductors.

V. APPLICATIONS

The investigated technological process has been applied in the production of steel sleeves of train carriage braking systems. In Bulgaria the main manufacturer of such sleeves is Pomorie PLC, furnishing with annealed steel sleeves the railway companies of Germany, Bulgaria and other countries of the European Union.

In Figure 8 a typical system inductor-detail used in Pomorie PLC is presented and in Figure 9 an inductor used for sleeves hardening is shown.

In our research we aim to optimize this production process.

Fig. 8. A typical system inductor-detail used in Pomorie PLC

Fig. 9. An inductor used for hardening of internal cylindrical surfaces of steel sleeves

REFERENCES

[1] Kuvaldin A.B., Induktsionnyi nagrev ferromagnitnoi stali, Énergoatomizdat, Moskva, 1988.

[2] Todorov T., Mechev I., Induktsionno nagryavane s visokochestotni tokove, Tehnika, Sofia 1979

[3] Chaboudez C., Clain S., Glardon R., Rappaz J., Swierkosz M., Touzani R., Numerical Modeling of Induction Heating of Long Workpieces, IEEE Transactions on magnetic, vol.30, № 6, November 1994

[4] Hoemberg D., Induction heat treatments – modeling, analysis and optiomal design of inductor coils, Habilitationsschrift, TU-Berlin, 2002

[5] Meeker D., Finite element method magnetic, ver. 4.2, User's manual, September 2009

[6] H.Uong, Osnovnye formuly i dannye po teploobmenu dlya inzhenerov, Atomizdat, Moskva 1979.

Model Research of Atmospheric Electric Effects in Electrical Low Voltage Network with Local Photovoltaic System

Margreta Vassileva[1] and Dimitar Dimitrov[2]

***Abstract* – The problem solved in this paper is to conduct variant research on the occurrence and limit atmospheric overvoltages in low voltage electrical networks with local photovoltaic system.**

The options are determined by place of installation of protective devices and cable length between inverter and photovoltaic panels.

***Keywords* – overvoltage, protective devices, low voltage networks.**

I. Introduction

Overvoltages in low voltage power supply networks threaten electrical facilities. They can not be designed with sufficiently high dielectric strength for economic reasons. Economical and secure network operation requires adequate protection of the equipment of the unacceptable impact of overvoltages. This applies to networks of high and low voltage networks.

II. Research Variants and Results

The task of this study is to conduct variant research on the occurrence and limit atmospheric overvoltages in low voltage electrical networks with local photovoltaic system. It is necessary find technically and economically viable solution to install surge protective devices in a residential building [1].

Protective devices must meet the requirements of standard IEC 61643-11. They are mounted between each lightning protection zones and must have an appropriate protective level.

According to standard IEC 60364-4-44, endurance levels of overvoltages of the equipment are classified into four categories. Protective devices must limit overvoltages under these levels.

Figure 1 shows one scheme of the research grid. The model integrate the following structural elements: power system (S); power lines 0,4 kV - air and cable, main switchboard (GRT); surge arresters - type metaloxide (MOSA), low voltage installation, consumers with different power (C_1, C_2, C_3). Subsystem of DC consists of a source of direct voltage (PV), which are modeled by photovoltaic panels, inverter, grounding resistors. Voltage of photovoltaic panels is 240 V. Parameters of the protective devices are [3,4]: MOSA1 between inverter and the photovoltaic panel – U_c=600 V DC , I_{max}=70 kA; MOSA2 before inverter - U_c=600 V AC, I_{max}=40 kA; U_p= 1,5 kV; MOSA3 in main switchboard - U_c=255 V AC, I_{max}=25 kA; U_p= 1,5 kV; MOSA4 - U_p= 1190 V.

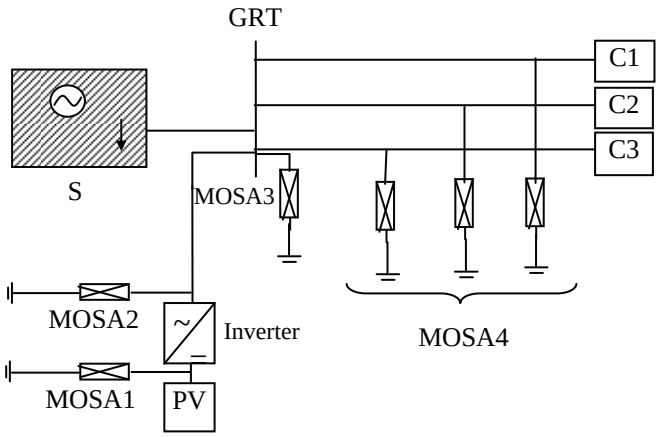


Fig. 1. One-line diagram of the research grid

Examined a case of direct hit of lightning in phase conductor of the power line in the system S. Observe the effects of lightning with parameters 40 kA and 1 /10 μs.

The voltage of the incoming wave and residual voltages of protective devices are controlled. Model scheme of low voltage network is presented in [2].

Studies have been made for the following cases:
1) Presence of MOSA in main switchboard, in the AC and DC parts of inverter and before consumers (fig. 1);
2) Presence of MOVO in main switchboard, in the AC and DC parts of inverter, without protective devices to consumers;
3) Presence of MOVO only in main switchboard;
4) Protective devices is not included.
5) Different cable length (10, 20, 50 m) between inverter and photovoltaic panels for case 1.

[1] Margreta Vassileva, Technical University of Varna, Department of Electric Power Engineering, Studentska 1, Varna, 9010, Bulgaria, m.vasileva@tu-varna.acad.bg.

[2] Dimitar Dimitrov, Technical University of Varna, Studentska 1, Varna, 9010, Bulgaria.

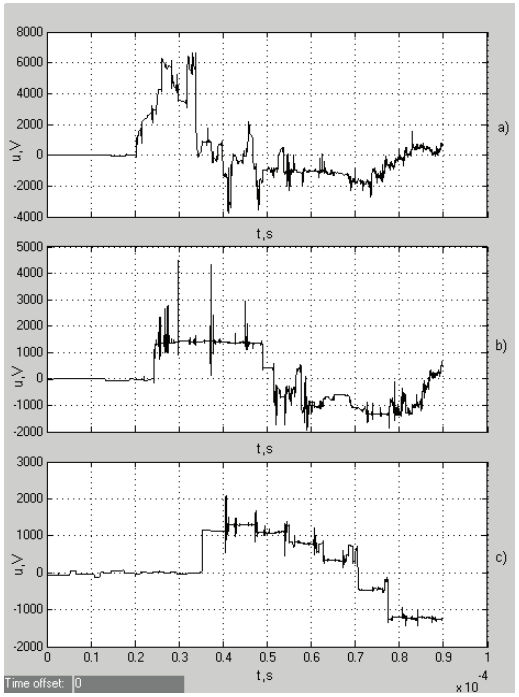


Fig. 2 Voltage of the incoming wave (a), residual voltage of MOSA3 (b) and of MOSA2 (c) (20 m cable length between inverter and photovoltaic panels)

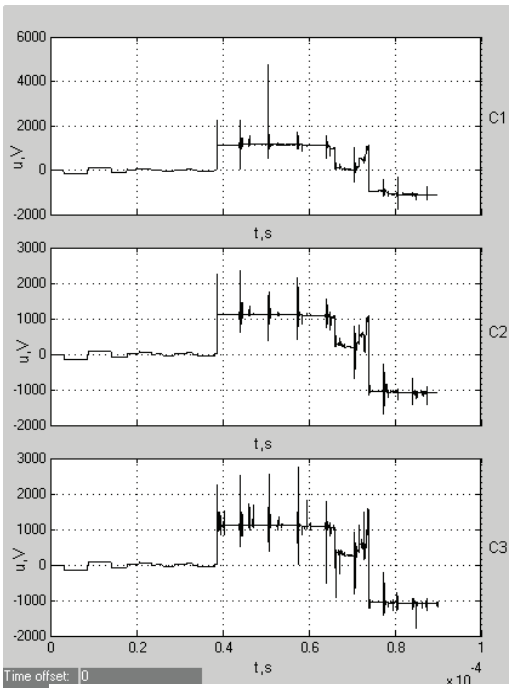


Fig. 3 Residual voltage of MOSA4 (20 m cable length between inverter and photovoltaic panels)

Figures 2 and 3 show the results for case 1.

Common pattern in the resulting time dependencies of voltages is that they develop as harmony fading fluctuations, which in steady pass to form the working voltages. All MOSA begin to work, limiting overvoltages to the corresponding protection levels (Fig. 2 and 3). The results for case 2 and 3 have the same character development as well as for case 1. Voltages of the consumers in cases 2 and 3 exceed the insulation level (1500 V) for first class facilities.

In case 3 voltage to the AC side of inverter also increase and exceeds the permissible level of insulation. MOSA 3 in GRT limited the overvoltages to the corresponding protection level, but its operation becomes unstable.

The results of controlled voltages when not using the protective devices (case 4) show that the insulation levels for all facilities are exceeded.

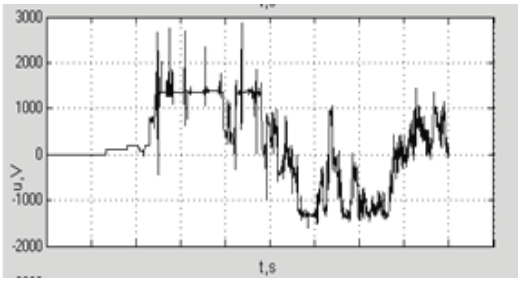


Fig. 4 Residual voltage of MOSA3 (50 m cable length between inverter and photovoltaic panels)

Concerning study on the influence of cable length between inverter and photovoltaic panels can be concluded that the frequency of the process in the network increases with increasing cable length. Protective devices restrict overvoltages to their set security levels, as in the 50 m length the action of MOVO in GRT is highly unstable (Figure 4).

III. CONCLUSION

From analytical studies can be made the following major conclusions:

1) The presence of MOSA only in main switchboard can not provide protection from overvoltages of equipment in lower installation category.

2) The presence of a DC circuit in the scheme and in particular the inverter brings additional disturbances of the operation of the protective devices.

3) It is necessary to include protective devices for all types of facilities to ensure their protection.

REFERENCES

[1] Stojanov A., V. Valchev and other. “Technical task of finding technically and economically viable solution to install the protective devices of overvoltages in residential building”, Engineering Review, pp. 42-56, march 2007.

[2] Vasileva M.P., K.K.Gerasimov, M.J.Yordanova, “A study of the function of surge protective devices for networks TN”, Akta Universitatis Pontika Euxinus, Russia, volume 4, number 1, pp. 99-102., 2005.

[3] Protective devices, Schneider Electric.

[4] Surge protection, ABB.

Risk Assessment of Lightning Damages

Marinela Yordanova[1], Mediha Mehmed-Hamza[2], Margreta Vasileva[3]

Abstract – **Need and economic benefit of introducing lightning protection and choice of protection measures is related to assessment and managing the risk of lightning damages. The article describes a computer program developed to assess the risk of lightning activity based on the European standards for lightning protection system.**

Keywords – **Risk Assessment, Lightning Protection**.

I. INTRODUCTION

European standards for lightning protection system EN 62305 introduces the risk associated with the lightning influences, need and economic convenience for lightning protection and the choice of protection measures. The comprehensive and complex risk assessment takes into account the structure to be protected and the services to which the structure is connected.

Assessment and risk management is subject to EN 62305-2 and purposes the choice of appropriate protection level, providing risk reduction to a value less than or equal to the limit.

Tolerable risk R_T [1] is the maximum acceptable value, which varies from 10^{-5} to risk loss of life to 10^{-3} on the risk of loss of public service networks or cultural heritage.

The authors have developed a computer program to determine the risk of lightning [2]. As a final result the program provides the appropriate lightning protection level. The program uses simplified assessment procedure for the following risks: R_1 - loss of human life or cause permanent damage; R_2 - loss of public service networks; R_3 - loss of cultural heritage.

In the present paper the authors offer a developed new computer program taking into account additional risk of loss of economic value R4 and additional components of all types risks R_1 to R_4- injury to living beings due to touch and step voltages as well as component related to physical damage caused by dangerous sparking inside the structure triggering fire or explosion. Those elements are assessed for following cases: lightning flash to a structure or lightning flash near the structure; lightning flash to an incoming service or lightning flash near the service.

[1]Assoc. Prof. PhD Marinela Yordanova – Department of Electrical Power Engineering, TU- Varna, e-mail: mary_2000@abv.bg

[2]Assist Prof. PhD. Mediha Mehmed-Hamza – Department of Electrical Power Engineering, TU-Varna, e-mail: mediha.hamza@mail.bg

[1]Assoc. Prof. PhD Margreta Vasileva – Department of Electrical Power Engineering, TU-Varna, e-mail: greta_w@mail.bg

II. PROCEDURE FOR RISK ASSESSMENT

To develop the new computer program methodological guidance for the assessment of damage from lightning [3] based on [4] is used and it includes:

– Identification of the object to be protected and its characteristics;

– Identification of all the types of loss in the object and the relevant corresponding risk R (R_1 to R_4);

– Evaluation of risk R for each type of loss (R_1 to R_4);

– Evaluation of need of protection, by comparison of risk R_1, R_2 and R_3 for a structure with the tolerable risk R_T;

– Evaluation of cost effectiveness of protection by comparison of the costs of total loss with and without protection measures.

Table 1 shows the risk components.

All types of risks are defined as:

$R_1 = R_A + R_B + R_C + R_M + R_U + R_V + R_W + R_Z$

$R_2 = R_B + R_C + R_M + R_V + R_W + R_Z$

$R_3 = R_B + R_V$

$R_4 = R_A + R_B + R_C + R_M + R_U + R_V + R_W + R_Z$

Any risk R is the sum of the components of risk R_D and R_I. R_D is a component of risk relating to physical injuries due to flashes to the structure, leading to fire or total or partial destruction of buildings and external facilities. R_I is a component of risk relating to physical damage (fire or total or partial destruction due to dangerous flashes between the internal fittings, and metal parts, which are usually located at the entry point of the line in buildings and outdoor facilities) arising from currents carried by lightning over or input into the buildings and external facilities for public service networks.

Risk components for a structure due to flashes to the structure are related to:

R_A: injury to living beings caused by touch and step voltages in the zones up to 3 m outside the structure.

R_B: physical damage caused by dangerous sparking inside the structure triggering fire or explosion, which may also endanger the environment.

R_C: failure of internal systems caused by LEMP.

Risk component for a structure due to flashes near the structure are related to:

R_M: Component related to failure of internal systems caused by LEMP.

Risk components for a structure due to flashes to a service connected to the structure are related to:

R_U: Component related to injury to living beings caused by touch voltage inside the structure, due to lightning current injected in a line entering the structure.

R_V: Component related to physical damage (fire or explosion triggered by dangerous sparking between external installation and metallic parts generally at the entrance point

TABLE I

RISK COMPONENTS FOR A STRUCTURE FOR DIFFERENT TYPES OF DAMAGE CAUSED BY DIFFERENT SOURCES

Source of damage Damage	S1 Lightning flash to a structure	S2 Lightning flash near a structure	S3 Lightning flash to an incoming service	S4 Lightning flash near a service	Resulting risk according to type of damage
D1 Injury to living beings	$R_A=N_D.P_A.r_a.L_t$		$R_U = (N_L + N_{Da}).P_U.r_U.L_t$		$R_S=R_A+R_U$
D2 Physical damage	$R_B = N_D.P_B.r_p.h_z.r_f.L_f$		$R_V = (N_L + N_{Da}).P_V.r_p.h_z.r_f.L_f$		$R_F=R_B+R_V$
D3 Failure of electrical and electronic systems	$R_C = N_D.P_c.L_0$	$R_M = N_M.P_M.L_0$	$R_W = (N_L + N_{Da}).P_W.L_0$	$R_Z = (N_I + N_L).P_Z.L_0$	$R_0=R_C+R_M+R_W+R_Z$
Resulting risk according to the source of damage	$R_D = R_A+R_B+R_C$	$R_I = R_M+R_U+R_V+R_W+R_Z$			

Note: N_i ,P_i , L_i , r_i , h_i are according the standart [1].

of the line into the structure) due to lightning current transmitted through or along incoming services.

R_W: Component related to failure of internal systems caused by overvoltage induced on incoming lines and transmitted to the structure.

Risk component for a structure due to flashes near a service connected to the structure are related to:

R_Z: Component related to failure of internal systems caused by overvoltage induced on incoming lines and transmitted to the structure.

III. COMPUTER PROGRAM FOR RISK ASSESSMENT

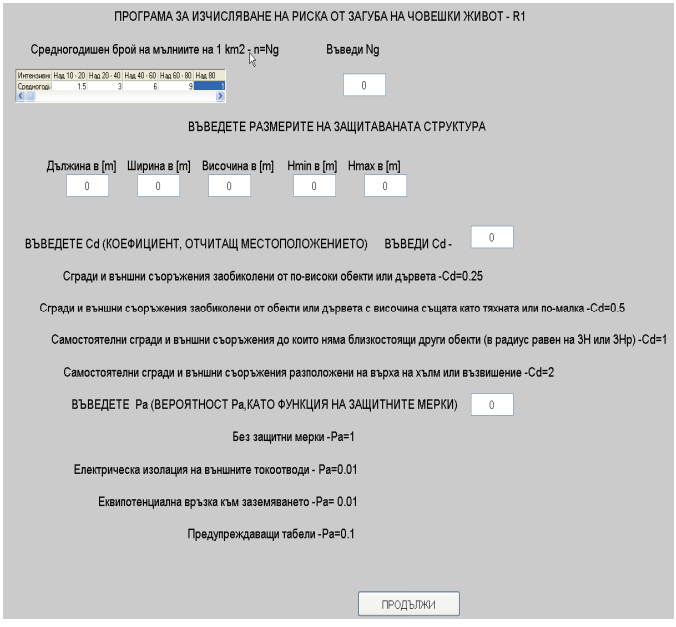


Fig.2. Initial dialog box for input data for risk component R_A related to injury to living beings caused by touch and step voltages.

The authors have developed a computer program based on the procedure [2], Table 1 and [3]. The user can set the calculation of each risk separately - R1 to R4. The program incorporated all the tabular data for determining the risk that the consumer operates in interactive mode, as shown in the sample dialog screens in Fig. 2. The program offers the appropriate level of lightning protection for each risk separately.

At the top of box from fig.2 it is introduced length, width, height and Ng. The coefficient Cd takes into account the location and it is equal to:

Cd =0,25 when the object is surrounded by higher objects or trees;

Cd =0,5 when the object is surrounded by objects or trees of the same heights or smaller;

Cd = 1 at isolated object: no other objects in the vicinity;

Cd = 2 at isolated object on a hilltop or a knoll.

Probability Pa of injury to living beings is:

Pa= 1 when there is not protection measures;

Pa= 0,01 - there is an electrical insulation of exposed down-conductor (e.g. at least 3 mm cross-linked polyethylene);

Pa= 0,01 - there is an effective soil equipotentialization;

Pa= 0,1 - there is the warning notices.

Similarly, the user continues to enter data and / or use the built-in program data. The program assesses risk and provides an appropriate level of protection.

REFERENCES

[1] EN 62305 Protection against lightning.

[2] Йорданова, М., М. Василева, М. Мехмед- Хамза. Компютърна програма за оценка на риска от атмосферни електрически явления, International Scientific and Technical Conference „Electrical Power Engineering 2010”, 2010q Varna, Bulgaria

[3] http://euroengineeringbg.com/

[4] БДС EN 62305-2 **“**Оценка на риска”

Active Front End Converter in Common DC Bus Multidrive Application

Nebojsa Mitrovic[1], Vojkan Kostic[2], Milutin Petronijevic[3] and Bojan Bankovic[4]

***Abstract* – In this paper, methods of energy recovery for AC adjustable speed drive in braking mode are presented. The analysis includes different front end converter topologies in terms of energy regeneration capabilities. Three solutions are found in practice: diode rectifier with braking modul, line-commutated rectifier-regenerative feedback unit and self-commutated pulsed rectifier-regenerative feedback unit. Results of practical application of the active front end unit on common dc bus are shown for the multimotor crane drives.**

***Keywords* – Multi-motor drives, Active front end, Common DC bus, Crane drives.**

I. Introduction

Adjustable speed drives (ASD) in industrial applications are usually characterized by a power flow direction from the AC distribution system to the load. This is, for example, the case of an ASD operating in the motoring mode. In this instance, the active power flows from the DC side to the AC side of the inverter. However, there are an important number of applications in which the load may supply power to the system. When the rotating element of an AC motor turns faster than the AC drive's speed command, the motor begins to act as a generator and regenerates energy back into the DC bus of the drive. Moreover, this could be an transient condition as well a normal operating condition. This is known as the regenerative operating mode. For example, these regenerative conditions can occur when quickly decelerating a high inertia load (flywheel) and this can be considered as transient condition. The speed control of a load moving vertically downward (hoist) or declining conveyor in minning application can be considered as normal operating condition, [1]. If the drive cannot absorb this excess energy, the DC bus voltage will continue to climb until the drive trips on a high bus fault.

This paper presents the characteristics of the three most commonly used topology for the front end rectifier: diode rectifier with brakig modul, line commutated regenerative unit and pulse comtutated pulse regenerative unit. Finally, the paper addresses the main characteristics and performance of the crane drives with active front end rectifier on common DC bus.

[1]Nebojsa Mitrovic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nebojsa.mitrovic@elfak.ni.ac.rs.

[2]Vojkan Kostic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: tijana.dimitrijevic@elfak.ni.ac.rs.

[3]Milutin Petronijevic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nebojsa.mitrovic@elfak.ni.ac.rs.

[4]Bojan Bankovic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nebojsa.mitrovic@elfak.ni.ac.rs.

II. Regenerative Operating Mode in ASD

The typical pulse width modulated AC drive is not designed for regenerating power back into the three phase supply lines, so all energy absorbed from the motor goes into the capacitor bank, resulting in increased DC bus voltage inside the drive. When equipped with a standard duty braking module and resistor, the drive is capable of dissipating only short-term energy, typically a few seconds at a time. It is possible to specify a braking module and resistor to dissipate this energy continuously, by taking into consideration the maximum current capacity of the brake switch, the duty cycle and the resulting power rating of the resistor. In either case, energy dissipated in a braking resistor is energy wasted.

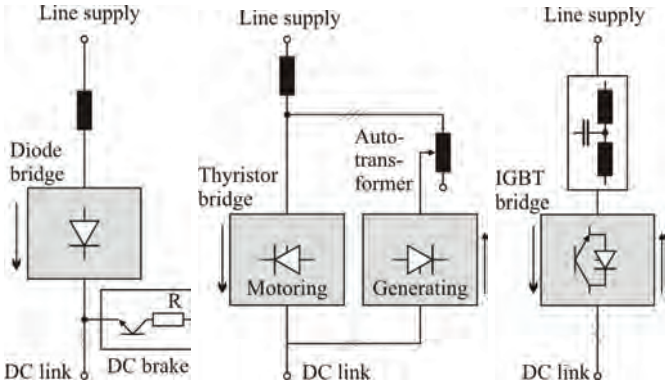


Fig.1. Front-end rectifier unit: a) Diode bridge with braking modul, b) Line-commutated rectifier-regenerative feedback unit, c) AFE-self-commutated pulsed rectifier-regenerative feedback unit

Besides the wasted energy during braking, there also is the disadvantage of not having any control over the DC link voltage. Also, the drives have to use large capacitor banks to "smooth" the DC link voltage. That's because the ripple frequency due to the rectifier tends to be relatively low and depends on the line frequency and number of diode bridge phases. For example, if the line frequency is three-phase 50 Hz, and if the system uses a three-phase diode bridge rectifier, then the ripple frequency will be 300 Hz.

A line-regenerative drive can improve on the diode rectifier front end by employing a 6-pulse SCR/thyristor bridge for rectification and another antiparallel thyristor bridge for regenerative feedback through a transformer. Such a system will allow power to flow back to the line (full four-quadrant operation), eliminating or greatly reducing the need for the brake resistors and improving system efficiency. This

topology provides several advantages when compared to the diode bridge rectifier front end, including four quadrant operation and control over the DC link voltage. However, this approach has disadvantages such as the addition of the antiparallel thyristor bridge and the requirement for filter components due to the use of line-commutating devices. Another drawback is the need for expensive compensation equipment to maintain power quality of the regenerative feedback. The large size and weight of the system can pose a challenge in some applications, [2-4].

Replacing the SCR/thyristor converters with IGBT-based active converters, also known as active front ends (AFE), will provide all the advantages of a four-quadrant regenerative drive while eliminating harmonic currents and improving the power quality without the need for expensive compensation equipment. AFEs also improve system efficiency and dynamic behavior of the load, as well as eliminate the need for the antiparallel converter or brake resistors. The active converters also significantly reduce the size and weight of the system when compared to systems based on line-commutated devices. The ability to switch the devices independent of the line frequency, typically between 15 kHz and 20 kHz, makes the filtering components smaller, lighter and less expensive.

When there are multiple drives in one location the common DC bus system is usually the most efficient way to operate and can incorporate the energy savings and recover concepts that have been previously discussed. However, if there is a regenerative drive and motor section in the system it is ideally suited for maximizing energy recovery and cost savings. The reason is that losses are generated when power is converted from the AC supply to the DC bus or from the DC bus to the AC supply. For multiple standalone (compact) drives, the power must go through two or more AC to DC conversions and two DC to AC conversions.

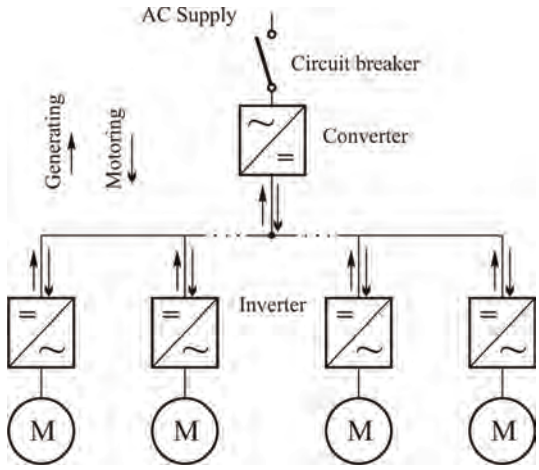


Fig 2. Common DC bus example

In a common DC bus configuration power only goes through one AC to DC conversion in the motoring direction, Fig.2. When an inverter section of the drive regenerates power to the DC bus, the power goes straight to another inverter, which is motoring via the common DC bus link, and does not have to travel through a converter at all. This method eliminates two conversion points where energy would be lost which increases efficiency by few percent for each regenerative section. The more sections there are which are regenerative, the more energy savings are accumulated. In addition, when a common bus solution is used with an AFE, it will have the ability to do power factor correction, which further increases the savings of a common bus system.

III. Case stady - Active Front End Converter in Multimotor Crane Drives

A. Torque and power requirements for crane drives

Speed control is an essential feature in crane drives. It is required for allowing soft starting and stopping of the travel motions for enabling its correct positioning of load. The torque and power that have to be delivered by the drive may be obtained from the torque versus speed characteristic from the load, [5,6].

If no wind influence has to be taken into account, the load characteristic is given in Fig.3a for travel motion. Apart from the zone around zero, the torque is constant. The available torque is used for accelerating the system.

The crane driver supplies the speed reference signal. For a travel in one direction, braking and reversing to full speed in other direction, the speed reference signal is given by top curve of Fig.3b. The torque reference signal is generated (second curve), leading to the machine actual speed. Myltiplying the actual and torque reference, yields the actual power (third curve). The peak power is found at the end of the acceleration period.

If wind forces are taken into consideration, the torque vs speed curve is shifted horizontally as shown in Fig.3c. The torque and speed reference remain the same, as well as the actual speed. However, the torque reference and the actual power differ, as shown on Fig.3d.

During acceleration, kinetic energy is stored in the system. To stop the crane, this energy must be absorbed by the drive. In the indoor situation, this energy is well known and only present for a short period of time. For outdoor applications, the wind forces may become very important. When travelling in the same direction as the wind, the wind drives the crane and a situation may occur, where a continuous electrical braking is required. The drive must be capable of handling this inverse power direction either by consuming the power in a resistor or preferably by feeding it back to the supply.

The hoist torque vs speed characteristic is shown in Fig.4a for an unloaded hook. The characteristic resembles the one for the travel motion. However, it is always asymmetric with respect to the vertical axis, due to the gravitation force. This asymmetry becomes more pronounced when the hook is loaded (Fig.4c). For both unloaded and loaded situation, the speed, torque and power are given in Fig.4b and Fig.4d. Again the amount of braking power is indicated. The worst braking case with a hoist motion, is when sinking a loaded hook. It should be noted that the weight of the hook may be considerable. The hook may be simple, or may consist of several parts to handle the load.

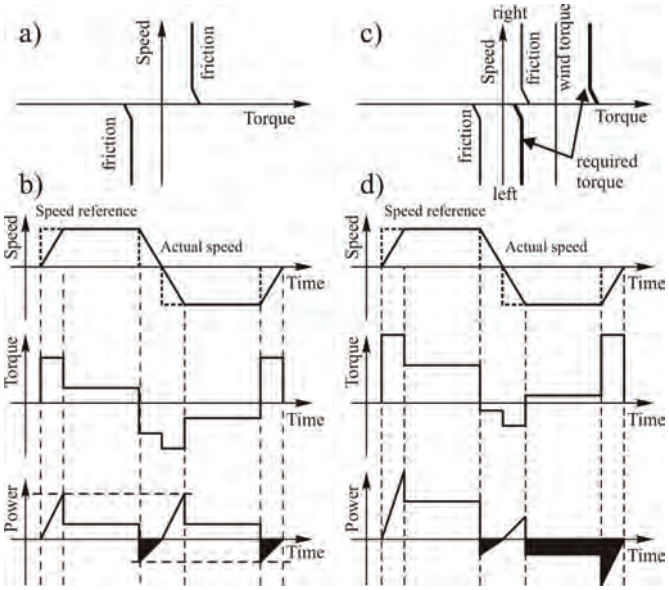


Fig. 3. Power and torque requirements for travel motion
a) and b) without wind influence, c) and d) with wind influence

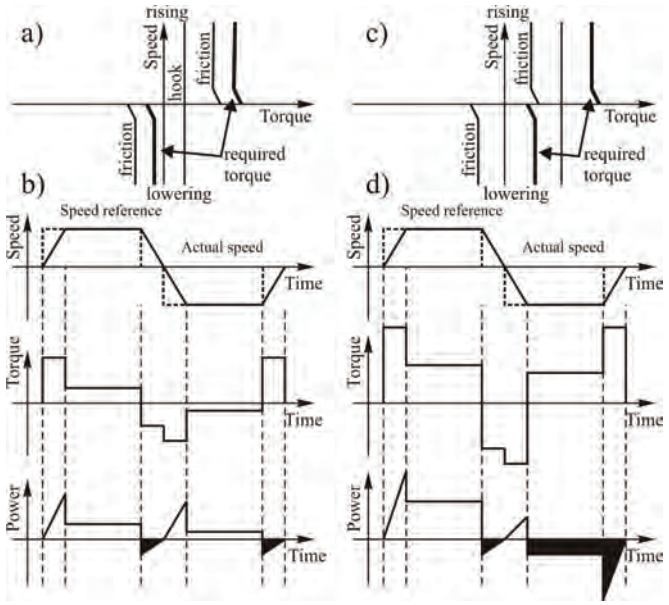


Fig. 4. Power and torque requirements for hoist
a) and b) without load, c and d) with load

A. Experimental results

The experimental behavior analysis of some drives is considered in a derrick crane, which serves for load handling i many industry branches. The main task in adjustable speed drives design is a safe, multi-axis movement that allows material handling throughout the working area, [7].

Using AFE rectifir/regenerative unit on common DC bus, five groups of invertor-motor combinations are supplied:
- hoist motion with 2x55 kW vector control inverter, the motor is a six pole, 2x45 kW,
- auxiliary hoist motion with a 55 kW vector control inverter, supplying a four pole, 45 kW motor,
- jib-boom motion with 2x55 kW vector control inverter, the motor is a six pole, 2x45 kW,
- travel motion with 3x7.5 kW vector control inverter , the motor , six pole 3x7.5 kW,
- auxiliary drives with 8x1.1 kW and motors (8x1.1 kW).

The rating of the AFE rectifir/regenerative unit output at cosφ=1 and 400 V supply voltage is 177 kW. This is far less than the sum of the ratings of the individual invertors, being 300 kW.

A number of experimental curves were recorded for the derrick crane. Fig.5 shows a hoist movement with the 30% of full load. Curve 1 gives the actual speed signal (reference speed signal is given at 100% from the crane driver joystick command). Curves 2, 3 and 4 shows the torque, power and rms motor current, respectively . After an acceleration period (ending at 5 sec), a constant torque is delivered. This transition in torque level coincides with reaching the prescribed speed. At 17.5 sec, the speed reference signal is made zero (stop command). The driving torque becomes zero and the system decelerates due to gravity. After 20 sec, zero speed is reached, and the drive has to deliver the torque required for holding the load before closing the mechanical brake. After that the same the same measurements were performed during lowering. Due to high friction losses, torque was required to start the decent, After short initial period, only the dynamic friction is present, yielding a small driving torque. After the acceleration, the power flow is reversed and the drive lowers the load at a constant speed. From 43 sec on, the system is braked with maximum torque, until standstill.

The same measurements were performed during lowering and hoist movement with the 30% of full load, Fig. 6. The reference speed was 25, 50, 75 and 100% of rated speed.

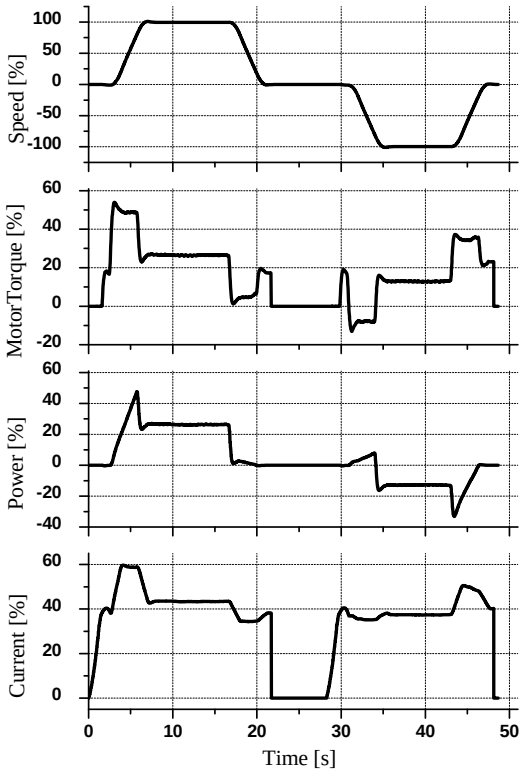


Fig.5. Measured pattern of the hoist motion

As on the Fig.5 actual speed, motor torque, power and rms current are shown. After that for jib-boom motion the same signal were record as shown on Fig.7. In both figure regenerative periods during the lowering, at any reference speed, can be seen.

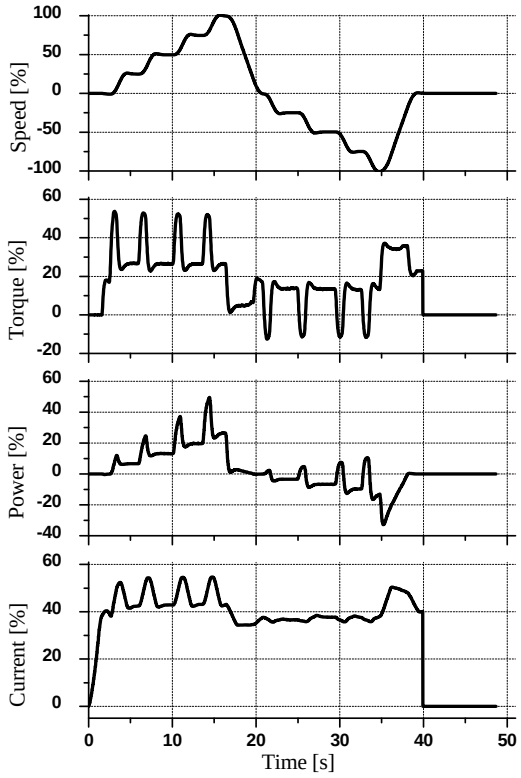


Fig.6. Measured pattern of the hoist motion at 25, 50, 75 and 100% of rated speed

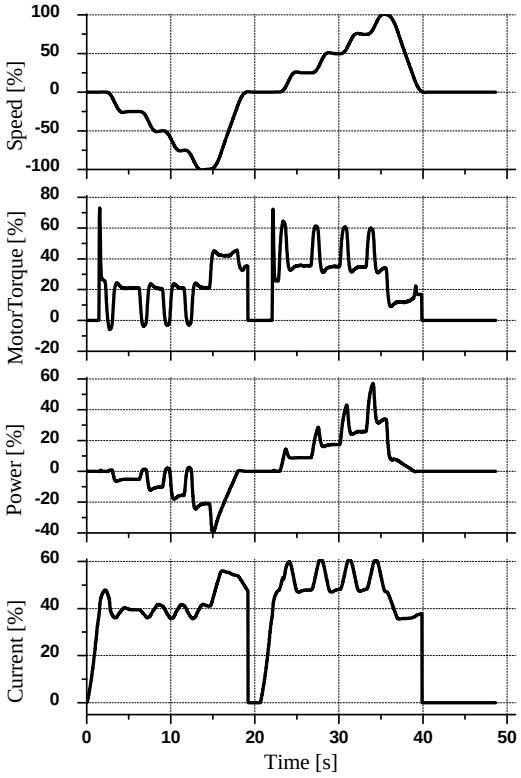


Fig.7. Measured pattern of the jib-boom motion at 25, 50, 75 and 100% of rated speed

From the Figs. 5 to 7, can be seen periods when the energy recovery occurs at the point of load lowering. It is very important to point out that the AFE topology allows for fully regenerative operation, which is quite important for crane application.

IV. Conclusion

This paper describes different drive technologies used in large power application. Each topology has advantages and disadvantages, especially from the power system point of view. The ability of handling regeneration and to suppress current harmonics is one of the basic features that must be included during the technical specification of such schemes. In multimotor drives the common DC bus system is usually the most efficient way to operate and can incorporate the energy savings and many other benefits such as: modular configuration, compact design, reduced installation costs, redundancy. In addition, common bus solution realised with Active Front End rectifier keeps the network current sinusoidal and a unity power factor by controlling the drive input to produce sinusoidal current without the harmonic components associated with conventional rectifiers.

Finally, it was shown that the use of AFE rectifiers in modern crane drives is a major technological advance and they produce an important improvement in the behavior of the drive system: higher power factor, reduced input current harmonics, and inherent regenerative operation.

Acknowledgement

This paper is supported by Project Grant III44006 (2011-2014) financed by Ministry of Education and Science, Republic of Serbia."

References

[1] H. Rashid, *Power Electronics Handbook*, Academic Press, San Diego, 2001.

[2] C. Klumpner, M. Liserre, F. Blaabjerg, "Improved Control of an Active-Front-End Adjustable Speed Drive with a Small DC-Link Capacitor Under Real Grid Conditions,". PESC'04, vol.2, pp. 1156-1162, June 2004.

[3] G. Skibinski, D. Dahl, K. Pierce, R. Freed, D. Gilbert, "Installation Considerations for Multi-Motor AC Drives and Filters Used in Metal Industry Applications," Conf. Rec. IEEE-IAS, vol.3, pp. 2270-2278, Oct 1998.

[4] Jeftenic B., Bebic M., Statkic S., "Controlled Multi-Motor Drives", SPEEDAM 2006, Taormina (Sicily) - ITALY, pp. 1392-1398, May 2006.

[5] R. Belmans, F.Busschots, and R.Timmer, "Practical Design Considerations for Braking Problems in Overhead Crane Drives," Conf. Rec. IEEE-IAS, vol.1, pp. 473-479, Toronto, Oct. 1993.

[6] J. Rodriguez, L. Moran, J. Pontt, J. Espinoza, R. Diaz, E. Silva, "Operating Experience of Shovel Drives for Mining Applications," Conf. Rec. IEEE-IAS, vol.1, pp. 705-711, 2002.

[7] N. Mitrovic, V. Kostic, M. Petronijevic, B. Jeftenić, "Practical Implementation of Load Sharing and Anti Skew Controllers for Wide Span Gantry Crane Drives", Journal of Mechanical Engineering, Vol. 56, no. 3, pp. 207-216, 2010.

Comparison of Two Methods for Estimation of a Single-Phase Transformer's Magnetization Curve

Milan M. Radic[1], Zoran P. Stajic[2]

***Abstract* – This paper presents analysis of differences between shapes of a single-phase transformer's magnetization curves obtained through two different approaches. The first approach has been based on the standard no-load test performed at several different values of the applied voltage, using the laboratory instrumentation capable to register true RMS electrical quantities (voltage, current, active power and reactive power). In the second approach, analysis has been performed considering recorded waveforms of no-load currents and corresponding induced voltages. Using the MATLAB/Simulink software the way explained in the paper, for any pair of recorded induced voltage and no-load current waveforms it has been possible to obtain dynamic hysteresis loop. Further analysis has shown that both methods give similar final results, if magnetic core is not heavily saturated. However, if the applied voltage is higher than the rated value and the no-load current is highly distorted due to non-linearity of the magnetic core, a significant difference between estimated magnetization curves occurs.**

Keywords **– Single-phase transformer, hysteresis loop, main magnetization curve.**

I. INTRODUCTION

There are numerous methods for measuring magnetization characteristics of electrical machines and magnetic materials in general. Most of them are based on direct current method and use specialized instrumentation that is not always available ([1]). However, there are situations in engineering and scientific practice when it is necessary to estimate nonlinear magnetization curve of an electrical machine, without use of specialized measuring equipment. In such cases, the only solution is to use an alternative approach and to exploit some of available AC measuring methods ([1-4]). Some of them are simple, and do not demand much equipment and time for analysis. On the other hand, some of AC measuring methods are more complex and have to be supported by sophisticated analytical process.

The aim of this paper is to perform comparison between two AC experimental methods in order to make conclusion, which one is more appropriate for use.

II. COMPARED EXPERIMENTAL METHODS

The simplest experimental procedure for quick estimation of magnetization curve is to perform no-load test at different values of applied voltage. According to this, single-phase transformer that was subject of experimental investigation in the paper has been powered from the secondary side, using an autotransformer as variable voltage source. In order to perform no-load test, terminals of the primary winding were opened, while the applied voltage U_0'', no load current I_0'' and consumed active power P_0'' were measured at terminals of the secondary winding. Since no-load current of the transformer always contains higher harmonics due to nonlinearity of the magnetic core, it is very important to use measuring equipment that retains rated accuracy even when non-sinusoidal electrical quantities are measured. For this purpose, a digital laboratory power analyzer, capable to register true RMS values has been used.

At any experimental point, the power factor can be estimated as

$$\cos\varphi_0'' = \frac{P_0''}{U_0'' I_0''}, \tag{1}$$

and that enables reactive part of no-load current to be calculated as

$$I_\mu'' = I_0'' \sin\varphi_0'' = I_0'' \sqrt{1-\cos\varphi_0''^2}\,. \tag{2}$$

Current I_μ'' is often called „magnetizing current", since it establishes magnetic flux in a core of a transformer, but does not take into account active power dissipated in the core. Knowing the effective value (Eq. 2), maximum value of magnetizing current can be easily calculated, using the well-known relation between maximum and effective value

$$I_{\mu m}'' = \sqrt{2} I_\mu''. \tag{3}$$

However, in such approach mistake is consciously made, because no-load current is not sinusoidal and contains harmonics of higher order (especially third and fifth harmonic). In fact, a realistic, more or less distorted current waveform is being supplemented by fictitious sinusoidal current that produces equivalent sinusoidal magnetic flux in a core of a transformer.

If resistance and leakage reactance of secondary winding, R'' and X_γ'', are known, for any effective value of the applied voltage, corresponding effective value of induced voltage in secondary winding can be calculated using Eq. 4.

$$E'' = \Big[\big(U_0'' - R'' I_0'' \cos\varphi_0'' - X_\gamma'' I_0'' \sin\varphi_0''\big)^2 + \\ + \big(R'' I_0'' \sin\varphi_0'' - X_\gamma'' I_0'' \cos\varphi_0''\big)^2\Big]^{1/2}. \tag{4}$$

Finally, maximum value of magnetic flux is obtained from

[1]Milan M. Radic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia
E-mail: milan.radic@elfak.ni.ac.rs

[2]Zoran P. Stajic is with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia
E-mail: zoran.stajic@elfak.ni.ac.rs

$$\Phi_m = \frac{E''}{4.44\, fN''}, \tag{5}$$

where N'' denotes number of turns in secondary winding and f denotes frequency.

Pair of values ($I''_{\mu m}, \Phi_m$) defines coordinates of the point that should belong to the main magnetization curve. If applied voltage is varied in reasonable small steps (e.g. one step could be about 10% of rated voltage), points obtained through previously described analysis should depict the nonlinear shape of the main magnetization curve. Results of such analysis performed on the real laboratory transformer are presented in Section 3.

The other method that can be used in order to identify nonlinear magnetization curve of a single-phase transformer is based on knowing of no-load current $i_0(t)$ (i.e. exciting current), and corresponding core flux $\Phi(t)$ waveforms, under different values of applied voltage. Basics of this method have been frequently described in the literature considering electrical machines and electromagnetism in general ([5-7]). The main idea is that, if several dynamic hysteresis loops have been successfully identified by plotting core flux $\Phi(t)$ versus exciting current $i_0(t)$, one can further obtain main magnetization curve by interpolating points whose coordinates are determined by peaks of hysteresis loops in the first quadrant. Sometimes, it is more appropriate to plot flux density $B(t)$ versus magnetic field $H(t)$, what can be easily done by scaling waveforms for $\Phi(t)$ and $i_0(t)$ with constant coefficients, depending on geometry of a magnetic core and construction parameters.

However, details of the experimental method are usually unexplained, which can be understood, because there is not only one and unique set of steps to perform in order to reach the final goal. The second author of the paper has already considered problem of identifying dynamic hysteresis loops of a single phase-transformer in his previous work ([1]), but the methodology used in that reference was significantly different.

The essential step in method used in this paper is to record accurate waveform of the no-load current in secondary winding $i''_0(t)$ and corresponding induced voltage in primary winding $e'(t)$. During experimental work, real current $i''_0(t)$ has been transformed to a voltage signal of appropriate amplitude, using LEM current module. Induced voltage $e'(t)$ has also been conditioned to an adequate voltage level, using linear isolating attenuator. These two signals, carrying all necessary information describing $i''_0(t)$ and $e'(t)$ waveforms, were recorded on the hard disc, using National Instruments PCI 6036E data acquisition card and LabView software. Both waveforms have been recorded with 400 samples per one cycle (sampling time was $5 \cdot 10^{-5}$ seconds).

After initial computations, performed in order to rescale recorded values to those equal with real electrical quantities, data can be used as input for a simple MATLAB/ Simulink model, based on use of Discrete Fourier blocks from SimPowerSystems Extra Library. This model was created in order to calculate magnitude (I_m or E_m) and phase angle (φ_i or φ_e) of odd harmonics in analyzed waveforms, up to 29-th.

Following this briefly described procedure, for any voltage applied to the secondary winding, no-load current in secondary and induced voltage in primary winding can be presented as:

$$i''_0(t) = \sum_{k=1}^{15} I_{m_{2k-1}} \sin\left((2k-1)\omega_1 t + \varphi_{i_{2k-1}}\right) \tag{6}$$

and

$$e'(t) = \sum_{k=1}^{15} E_{m_{2k-1}} \sin\left((2k-1)\omega_1 t + \varphi_{e_{2k-1}}\right), \tag{7}$$

where $2k-1$ is the order of the harmonic and ω_1 angular frequency of the main harmonic.

Regarding Eqs. 6 and 7, it is possible to create another simple Simulink model, proposed to generate smooth and accurate waveforms of current $i''_0(t)$ and induced voltage $e'(t)$ for each specific case. If the transformer was in steady state operation during the recording of mentioned waveforms (i.e. applied voltage had constant effective value), it is enough to generate only one cycle of no-load current and induced voltage.

Since induced voltage in primary winding and core flux are connected through

$$e'(t) = -N' \frac{d\Phi(t)}{dt} \tag{8}$$

waveform of the core flux $\Phi(t)$ can be further obtained as

$$\Phi(t) = -\frac{1}{N'} \int e'(t)dt + C, \tag{9}$$

where N' denotes number of turns in primary winding. Constant C is the initial value of core flux at the beginning of the analyzed cycle, and it is very important to be accurately determined, otherwise waveform $\Phi(t)$ will have an offset. Calculation of this constant can be easily done if simulation is once performed assuming that $C = 0$, since whole waveform will be shifted along flux axis for exactly $-C$ in that case.

Finally, plotting obtained waveform $\Phi(t)$ versus $i''_0(t)$ defines dynamic hysteresis loop that is valid for the applied voltage. If the value of the applied voltage is varied, a family of concentric hysteresis loops will be identified. In this case, points characterized by maximum value of flux and corresponding no-load current from any of obtained hysteresis loops, will be points that define shape of main magnetization curve.

III. Experimental Results and Discussion

Both of methods explained in Section 2 were used in order to estimate main magnetization curve of a real single-phase

transformer with rated values $S_n = 2\,kVA$, $U'_n = 380V$, $U''_n = 190V$, $I'_n = 5.3\,A$, $I''_n = 10.6\,A$, and number of turns per winding $N' = 360$ and $N'' = 179$.

During the investigation conducted according to the first described method, no-load test has been performed for 17 different values of the applied voltage U''_0, the highest being almost 1.5 times greater than rated voltage U''_n. Measured values are presented in first three columns of the Table I. Fourth and fifth column of the same table represent corresponding values of maximum magnetizing current $I''_{\mu m}$ and maximum core flux Φ_m. These values were calculated using Eqs. (1)-(5), regarding previously determined parameters of the secondary winding $R'' = 0.91\Omega$ and $X''_\gamma = 2.48\Omega$.

TABLE I
VALUES FROM STANDARD NO-LOAD TESTS

measured			calculated	
U''_0 [V]	I''_0 [A]	P''_0 [W]	$I''_{\mu m}$ [A]	Φ_m [Wb]
33.2	0.083	1.54	0.097	0.00083
55.8	0.111	4.19	0.115	0.00140
82.3	0.142	8.55	0.137	0.00206
106.7	0.174	13.57	0.168	0.00267
144.1	0.247	23.56	0.261	0.00361
167.8	0.337	31.5	0.395	0.00420
185.7	0.453	38.02	0.571	0.00464
193.7	0.526	41.06	0.681	0.00484
201.9	0.619	44.62	0.818	0.00504
213.9	0.788	50.06	1.064	0.00533
225	0.994	55.91	1.361	0.00560
235.6	1.229	62.54	1.698	0.00585
245.3	1.488	68.26	2.068	0.00608
253.8	1.735	74.86	2.419	0.00627
264.7	2.11	82.1	2.953	0.00652
276.1	2.581	94.07	3.621	0.00678
287.3	3.172	106.62	4.459	0.00702

Calculated values are graphically presented on Fig. 3. White circles represent points with coordinates ($I''_{\mu m}$, Φ_m), while the dashed line presents estimated basic magnetization curve, obtained by interpolation between experimental points. This curve has also been extrapolated above last experimental point.

Experimental investigation according to the second described method has been performed for 13 different values of the voltage applied to terminals of secondary winding. However, these voltages were not exactly measured, since their values were meaningless in this investigation. As it had been explained in Section 2, waveform of voltage induced in primary winding $e'(t)$ was recorded instead. It had only been important to get the range of flux core variation that could be compared to the range from the first experiment. In order to obtain uniform distribution of experimental points, RMS value of voltage induced in primary winding was measured as an orientation.

The lowest measured value was $E'_{\min} = 51V$, while the highest value was $E'_{\max} = 580V$. Knowing that transformer's turn ratio is $N'/N'' \approx 2$, it is obvious that the range of flux variation during the second experiment was somewhat wider, compared to the experiment whose results are presented in Table I. Following the procedure described in Section 2, 13 different dynamic hysteresis loops for investigated single-phase transformer have been plotted (Fig. 1).

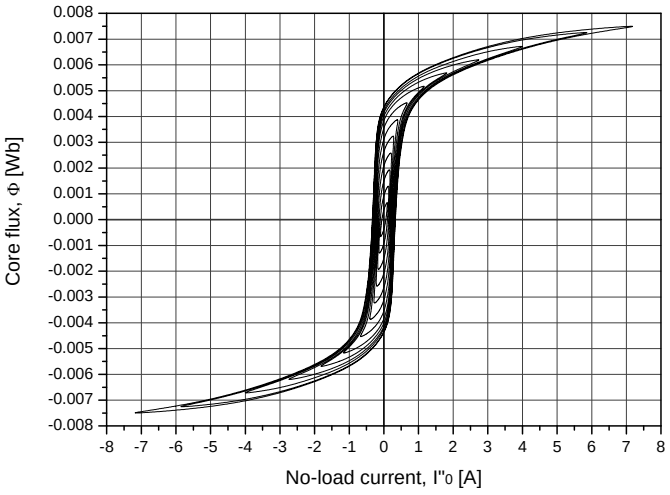


Fig. 1. Obtained hysteresis loops

In order to obtain basic magnetization curve through this approach, it was necessary to identify coordinates of points characterized by maximum value of flux, on each of 13 plotted hysteresis loops.

Fig. 2 shows only the part of hysteresis loops presented on Fig. 1, but for the desired analysis, this part is the most important. On any of presented loops, one can notice the point where flux reaches its maximum Φ_m, and consequently, magnetizing current also has maximum value $I''_{\mu m}$. These points are presented by black circles, and after they had been identified, it was possible to perform interpolation between them and to estimate main magnetization curve (thick line on Fig. 2).

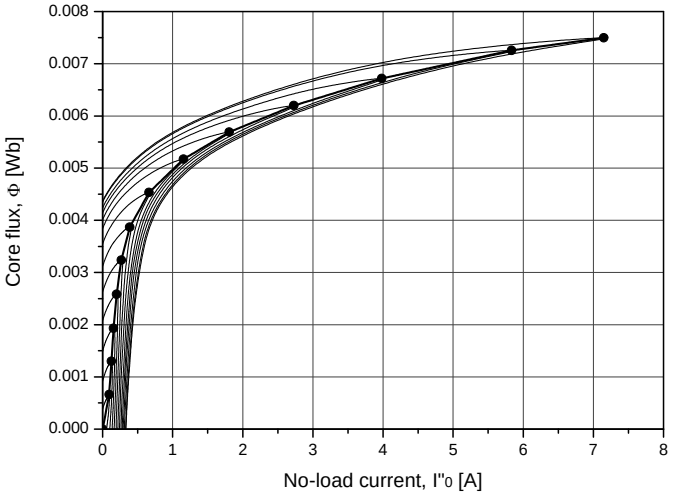


Fig. 2. Peaks of hysteresis loops and interpolation between them

Results obtained through two different experimental methods are finally presented on the same graph (Fig. 3), in order to enable easy comparison. From Fig. 3, it is obvious that experimental points obtained through two different approaches lay on two different nonlinear curves. It is not appropriate to compare results using point-by-point method, since their number is not equal, and they were not supposed to be comparable, as it has been already mentioned. However, nonlinear curves defined by those points should be the subject of consideration. It has to be mentioned that interpolations shown on Fig. 3 surely are not the best nonlinear approximations that could have been made, but even with points roughly connected by straight-line segments, significant difference can be noticed.

Complete main magnetization curve obtained through the first experimental method is placed above the curve obtained by the second method. If those curves are compared using constant flux as a criterion, it can be said that first curve underestimates maximum value of magnetizing current necessary to create desired magnetic flux in transformer's core.

Perhaps somebody might consider this situation in opposite direction, saying that second curve overestimates maximum value of the magnetizing current, however, it would not be correct. Previous statement is clear if one keeps on mind that first experimental method deals with fictitious, equivalent sinusoidal waveforms of no-load current and core flux, while the second one is established on analysis of real no-load current and core flux waveforms.

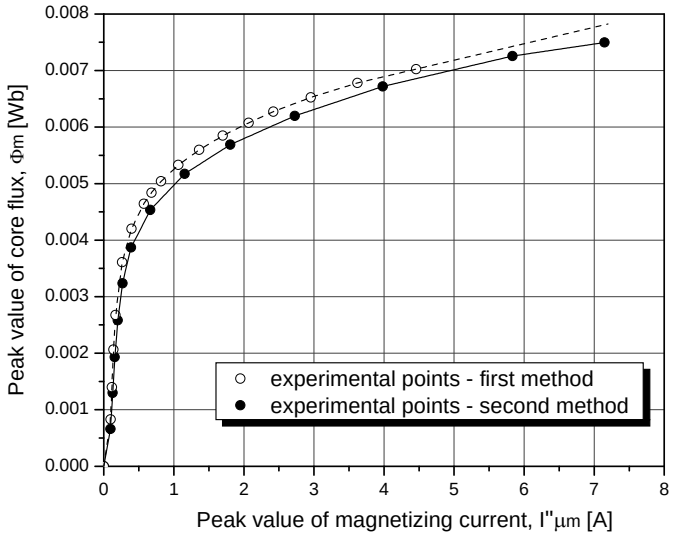


Fig. 3. Comparison of obtained magnetization curves

It can be noticed that small difference exists even between those segments of estimated magnetizing curves that are usually treated as „linear part of magnetization curve" in the technical literature ([4], [6]). In fact, there is no part of magnetization curve that can be considered as absolutely linear, however, an usual engineering approach is to treat the first segment of the curve in that way. As the value of the applied voltage increases, transformer's magnetic core becomes more and more saturated, and in this segment greater difference between estimated magnetization curves can be observed. This can be explained by the fact that when magnetic circuit is heavily saturated, no-load current becomes highly distorted, due to harmonics of higher order emerging in it's reactive component (i.e. in magnetizing current).

IV. Conclusions

Based on previous analysis, it can be said that considered experimental methods for estimation of a single-phase transformer's magnetization curve give similar, but not identical results. If magnetic core is not saturated and operating point remains on linear segment of magnetization curve, difference between obtained curves can be neglected. However, even at rated operating point, which is usually on the knee of the curve, slight difference is notable. Finally, if transformer's operating point enters the saturated region of magnetization curve for any reason, basic magnetization curve obtained through classical no-load test will underestimate maximum values of magnetizing current.

The final conclusion is that before any experimental activity, one should have clear idea what is the real purpose of identification of magnetizing curve. If estimated curve will be used for analysis of non-saturated or slightly saturated regimes, it is convenient to use the first method, which is less complex and demands less equipment and time. Otherwise, the advice is to use more complex, but also more accurate experimental method, based on analysis of no-load current and core flux waveforms.

Acknowledgment

The work presented here was supported by the Serbian Ministry of Education and Science (project III44006).

References

[1] Z. P. Stajić, S. Aleksandrov, Đ. R. Vukić and N. Rajaković, „Značaj poznavanja karakteristika magnećenja magnetskih kola pri remontu transformatora" (in Serbian), Zbornik referata III Simpozijuma Industrijska Elektronika INDEL 2000, pp. 60-65, Banjaluka, 2000.

[2] Z. Wlodarski, "Extraction of Hysteresis Loops From Main Magnetization Curves", Journal of Magnetism and Magnetic Materials No 308, pp. 15-19, 2006.

[3] S. Bogarra, A. Font, I. Candela and J. Pedra, "Parameter Estimation of a Transformer With Saturation Using Inrush Measurements", Electric Power System Research No 79, pp, 417-425, 2009.

[4] R. Wolf, *Ispitivanje električnih strojeva* (in Croatian), Zagreb, Sveučilište u Zagrebu, 1964.

[5] J. Hindmarsh, *Electrical Machines and Their Applications*, Oxford, Pergamon Press, 1977.

[6] G. Slemon, A. Straughen, *Electric Machines*, Reading, Massachusetts, Adison-Wesley Publishing Company, 1982.

[7] A. Fitzgerald, C. Kingsley, S. Umans, *Electric Machinery*, Tokyo, McGraw-Hill International Book Company, 1983.

Experimental Analysis of Direct Torque Control Methods for Electric Drive Application

Vojkan Kostić[1], Nebojša Mitrović[2], Milutin Petronijević[3] and Bojan Banković[4]

***Abstract* – In this paper different direct torque and flux control of induction motor schemes (DTC) are presented. A control techniques, analysed in this paper, related to voltage inverters and their solutions are essential diverse. Classical DTC method, its modifications for torque and flux ripple reduction, as well as modified DTC method with PI controllers (PI-DTC) based on space vector modulation (SVPWM), are considered. For each of methods, theoretical principles and experimental results, at laboratory condition using dSPACE development tool realised, are presented.**

***Keywords* – induction motor drive, space vector modulation, direct torque and flux control.**

I. INTRODUCTION

Direct Torque Control (DTC) was proposed by Takahashi and Depenbrock [1-2]. This method presents the advantage of a very simple control scheme of stator flux and torque by two hysteresis controllers, which give the input voltage of the motor by selecting the appropriate voltage vectors of the inverter through a look-up-table in order to keep stator flux and torque within the limits of two hysteresis bands. Different voltage vector selection criteria can be employed to control the torque according to whether the flux has to be reduced or increased, leading to different switching tables. Very high dynamic performance can be achieved by DTC, however, the presence of hysteresis controllers leads to a variable inverter switching frequency operation. In addition, the time discretization, due to digital implementation, plus the limited number of available voltage vectors is source of large current and torque ripple, causing the deterioration of the steady performance especially in low speed range. In order to improve the steady performance, different DTC strategies have been proposed to perform constant switching frequency operation and to decrease the torque ripple.

This paper presents the theoretical principles of conventional DTC method, its modification in order to reduce the torque and stator flux pulsations, and constant switching frequency DTC method with PI controllers (PI-DTC), which solved some of the above-mentioned shortcomings [3-4]. Each of the considered method was implemented and experimentally verified in the dSPACE development system in the laboratory.

[1]Vojkan Kostić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vojkan.kostic@elfak.ni.ac.rs.

[2]Nebojša Mitrović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nebojsa.mitrovic@elfak.ni.ac.rs.

II. PRINCIPLES OF DIRECT TORQUE CONTROL

A. DTC with Classical Switching Technique (c_DTC)

A uniform rotating stator flux is desirable, and it occupies one of the sectors at any time, Fig. 1. The stator flux vector has a magnitude of ψ_s with instantaneous angle $\theta_{\psi s}$.

If the stator flux vector is in sector 2, Fig. 1, the left influencing voltage vector has to be either $\mathbf{U}_1$ or $\mathbf{U}_6$. As seen from vector diagram, in case of applying voltage vector $\mathbf{U}_1$, the flux vector increases in magnitude. In case of vector $\mathbf{U}_6$, it decreases. This implies that the closer voltage vector applying increase the flux and the farther voltage vector decreases the flux and both of them change the flux vector magnitude and orientation. Similarly for all other sectors, the switching logic can be developed. A flux error (ψ_s^*-ψ_s), thus determines which voltage vector has to be called, is converted to the error state signal S_ψ using hysteresis flux controller with $\Delta\psi_s$ hysteresis band. The digitized output signals of the two level flux controller are given in Table I.

Torque error is processed through hysteresis controller to produce error state signal, S_m as shown in Table II. Interpretation is as follows: S_m=1 requires increasing the voltage angle, 0 means to keep it at zero, and S_m= -1 requires decreasing the voltage angle.

Combining the flux error output S_ψ, the torque error output S_m, and the sector number of the flux vector, a switching table can be realized to obtain the switching states of the inverter.

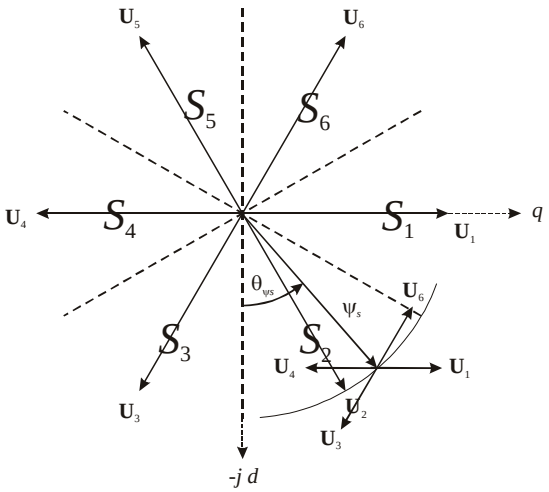


Fig. 1. Classical DTC and its sectors

In the classical DTC, there are several drawbacks [5-6]. Some of them can be summarized as follows:

- large and small errors in flux and torque are not distinguished. In other words, the same vectors are used during start up and step changes and during steady state;
- high torque pulsation especially at low speed.

TABLE I SWITCHING LOGIC FOR FLUX ERROR

State	S_ψ
$\psi_s^* - \psi_s > \Delta\psi_s / 2$	1
$\psi_s^* - \psi_s < -\Delta\psi_s / 2$	-1

TABLE II SWITCHING LOGIC FOR TORQUE ERROR

State	S_m
$m_e^* - m_e > \Delta m_e / 2$	1
$-\Delta m_e / 2 \le m_e^* - m_e \le \Delta m_e / 2$	0
$m_e^* - m_e < -\Delta m_e / 2$	-1

B. DTC with Modified DTC Technique (m_DTC)

In order to overcome the mentioned drawbacks, there are different solutions. One of the possible methods to improve the DTC performance is sector modification. Similar to the classical DTC six sectors is used including change in their orientation. Hence, instead of a first sector from 60º up to 120º, it will be from 30º up to 90º. The new sector division is shown in Fig. 2.

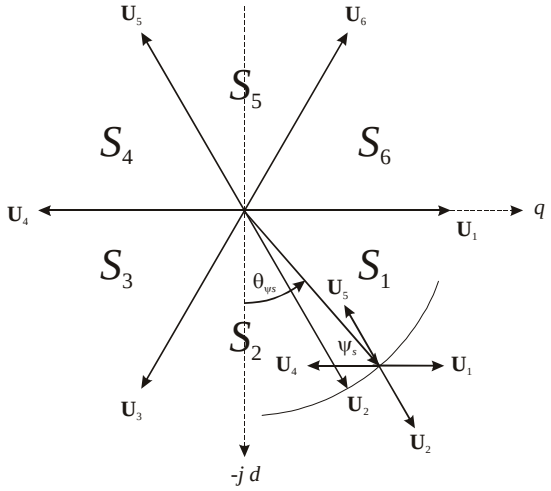


Fig. 2. Modified DTC and its sectors

Control of the flux and torque can be done by the similar procedure as for the classical DTC method. It can be observed that the states $\mathbf{U}_k$ and $\mathbf{U}_{k+3}$ are not used in the classical DTC (c_DTC) because they can increase or decrease the torque at the same sector depending on if the position is in its first 30 degrees or in its second ones. In the modified DTC (m_DTC), $\mathbf{U}_{k+2}$ and $\mathbf{U}_{k+5}$ are the states not used. However, now the reason is the ambiguity in flux instead of torque, as it was in the c_DTC. This is considered to be an advantage in favour of the m_DTC as long as the main point is to control the torque. Therefore, it is better to loose the usage of two states for flux ambiguity that for torque one [7-8].

C. DTC with Twelve sector Switching technique (12_DTC)

In classical DTC there are two states per sector that present a torque ambiguity. Therefore, they are never used. In a similar way, in the modified DTC there are two states per sector that introduce flux ambiguity, so they are never used either. It seems a good idea that if the stator flux locus is divided into twelve sectors instead of just six, all six active states will be used per sector. Consequently, it is arisen the idea of the twelve sector modified DTC (12_DTC). This novel stator flux locus is introduced in Fig. 3. Notice how all six voltage vectors can be used in all twelve sectors. However, it has to be introduced the idea of small torque increase instead of torque increase, mainly due to the fact that the tangential voltage vector component is very small and consequently its torque variation will be small as well.

As it has been mentioned, it is necessary to define small and large torque variations (S_m=1 - torque small increase, S_m=2 - torque large increase, S_m=-1 - torque small decrease, S_m=-2 - torque large decrease). Therefore, the torque hysteresis block should have four hysteresis levels and eight levels of flux and torque variation, Table III. It is obvious that $\mathbf{U}_2$ will produce a large increase in flux and a small decrease in torque in sector 2. On the contrary, $\mathbf{U}_3$ will decrease the torque in large proportion and the flux in a small one.

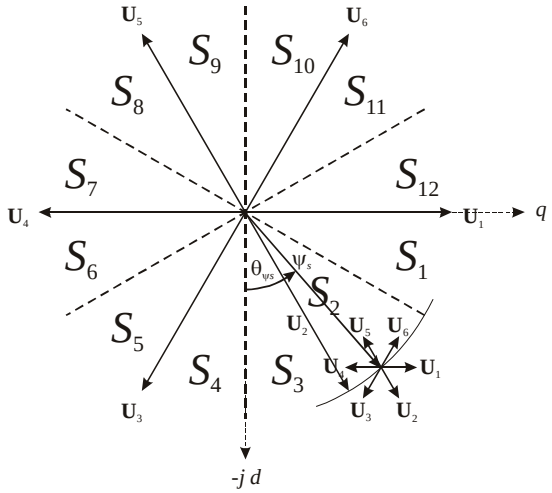


Fig. 3. Twelve sectors DTC

TABLE III LOGIC FOR TORQUE ERROR (12_DTC)

State	S_m
$m_e^* - m_e > \Delta m_e / 2$	2
$\Delta m_e / 2 \ge m_e^* - m_e \ge 0$	1
$0 > m_e^* - m_e \ge -\Delta m_e / 2$	-1
$m_e^* - m_e < -\Delta m_e / 2$	-2

D. DTC with PI Controllers (PI_DTC)

In subsections *A* to *C* the classical DTC method and its modifications are described, where regulation is in discrete values of output voltage inverter, with 8 discrete state (6 non-zero and 2 zero state).

The application of space vector modulation, SVPWM, enables to select the inverter output voltage of any phase position and amplitude in the domain of possible values. This approach allows the development of new DTC algorithm to improve the performance of existing ones. Fig. 4 shows the stator flux vector $\boldsymbol{\psi}_s$, which, in relation to the *–jd* stationary reference system has a phase position $\theta_{\psi s}$. If we adopt the *q*-axis of synchronous reference system, q^e, coincides with stator flux vector, it is clear that the *q* component of inverter output voltage in synchronous reference system $\mathbf{U}_{qs}^{\ e}$, affects only to the amplitude of stator flux vector. Also, the *d* component of inverter output voltage in synchronous reference system, $\mathbf{U}_{ds}^{\ e}$, affects only to the phase position of the stator flux vector and consequently to the torque response.

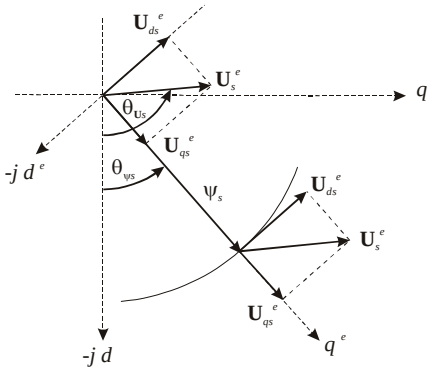


Fig. 4. Stator flux vector in synchronous reference frame

The above statements represent the basis for a modified DTC method with PI controllers (PI_DTC method). PI_DTC method uses PI controllers for calculation *qd* components of inverter output voltage in synchronous reference system. Trigonometric functions, necessary for the transformation from synchronous to stationary reference system, can be avoided by using the appropriate equations [8]. The need for adjusting the parameters of PI controller makes the method more complex than the traditional DTC methods described in subsections *A* to *C*. However, with well-designed PI controllers, we should expect much better performances in stationary conditions and slightly slower in the transient responses.

III. Experimental Results

In order to verify described DTC algorithms an experimental model of induction motor is formed. Its flow chart is presented on Fig. 5.

Control algorithm model of different DTC methods is formed in MATLAB/Simulink software. Measurement of motor currents is performed using LEM current probe and acquired with two analog inputs with sampling frequency 10 kHz. Speed measurement is realised using incremental encoder interfaced to dSPACE quadrature decoder with sampling frequency 1 kHz. In DTC methods described in subsections *A* to *C* (c_DTC, m_DTC and 12_DTC), inverter switching elements controled by three digital outputs with maximum switching frequency equal to 10 kHz. In DTC method described in subsection *D* (PI_DTC), control of inverter switching elements is performed by SVPWM with 5 kHz switching frequency. Discretization time in all experiments is 100 μs. Control of experiments, visualization, parameters variation and data acquisition are realised by dSPACE software *ControlDesk Developer*.

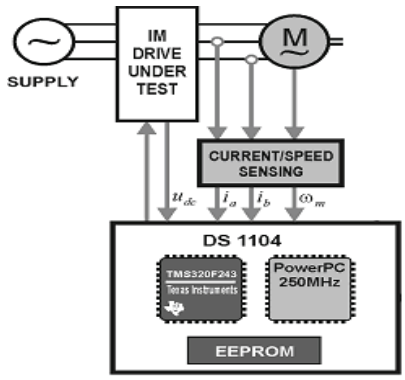


Fig. 5. Block diagram of experimental model

For DTC methods from subsections *A* to *C*, parameters: stator flux hysteresis band, $\Delta\psi_s$ and torque hysteresis width are adjusted on values 2% and 20%, respectively, in regard corresponding rated values. In this section experimental results of induction motor operation with DTC in torque control mode are presented. Reference stator flux is set to rated value, $\psi_s^*=\psi_{sn}$. Reference torque has square waveform, $m_e^*=\pm 3$ Nm, with the period 0.25 s. Motor is unloaded.

On Figs. 6 to 9 stator flux and estimated torque during the operation of described drive for previously explained DTC methods are presented. Also, torque is zoomed at the moment of its passing through zero for better insight. On the basis of the figures it can be concluded that the waveforms of stator flux and torque correspond to the conclusions given in section II. As was expected, PI_DTC method yields incomparable lower ripple waveform. Considering this criteria, presented control algorithm is equally good as vector control. Torque response during the change of reference is very fast and for DTC methods described in subsections *A* to *C* is about 0.0003 s. As expected, the PI_DTC method has a slower torque response and in the given case is 0.0012 s.

IV. Conclusion

In this paper, the theoretical principles of conventional DTC method, its modification in order to reduce the torque and stator flux pulsations, and constant switching frequency DTC method with PI controllers (PI-DTC), is presented. Each of the considered methods was implemented and experimentally verified in the dSPACE development system in the laboratory. Also, their advantages and limitations have been examined.

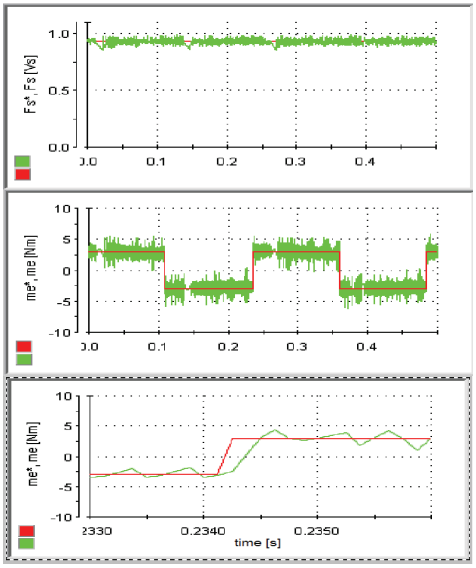


Fig. 6. Classical DTC method

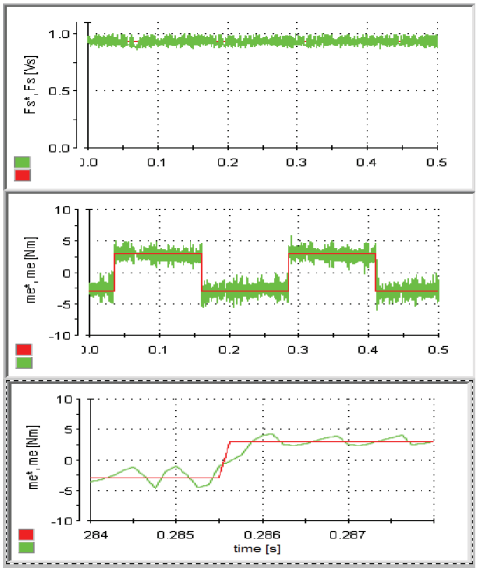


Fig. 7. Modified DTC method

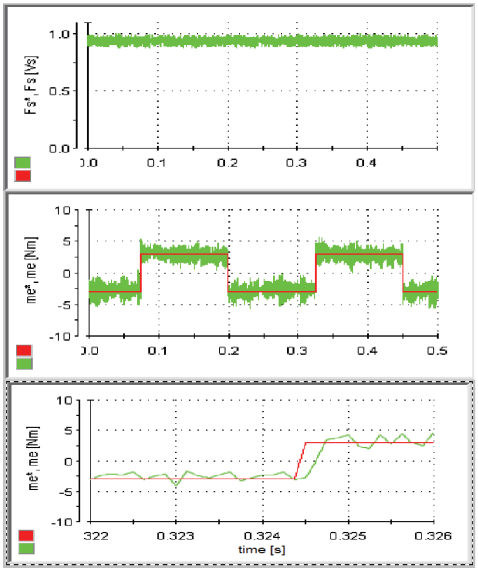


Fig. 8. DTC with twelve sector switching technique

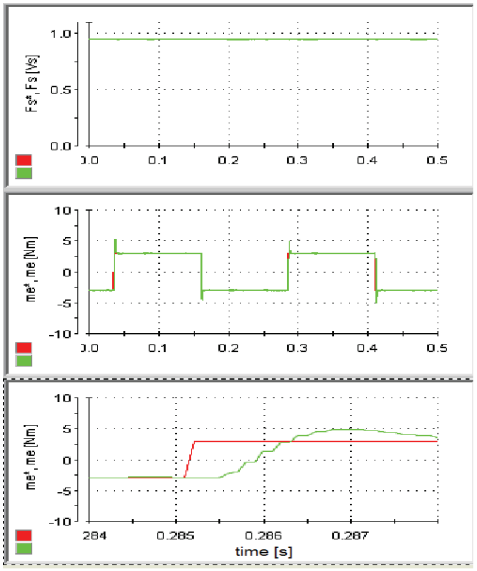


Fig. 9. DTC with PI controllers

ACKNOWLEDGEMENT

This paper is supported by Project Grant III44004 (2011-2014) financed by Ministry of Education and Science, Republic of Serbia.

REFERENCES

[1] I.Takahashi, T.Noguchi,"A new quick-response and high-efficiency control strategy af an induction motor," IEEE Trans. on Ind. Appl., Vol.22, No.5, pp.820-827, 1986.

[2] M. Depenbrock, "Direct Self Control (DSC) of Inverter Fed Induction Machine", IEEE Trans. on Power Electronics, Vol. 3, No.4, pp.420-429, 1988.

[3] Nebojsa Mitrovic, Vojkan Kostic, Milutin Petronijevic and Borislav Jeftenic, "Simulation of Direct Torque Control Schemes for Electric Drive Application, Part I", Scientific Bulletin of "Politehnica" University of Timisoara, Transactions on Automatic Control and Computer Science, Vol. 49(63), No. 1, 2004, ISSN 1224-600X, pp. 83-86.

[4] Y.S. Lai, and J.H. Chen, "A New Approach to Direct Torque Control of Induction Motor Drives for Constant Inverter Switching Frequency and Torque Ripple Reduction", IEEE Transaction on Energy Conversion, Vol. 16, No. 3, September 2001, pp. 220-227.

[5] D. Casadei, G. Grandi, G.Serra, A. Tani," Switching Strategies in direct Torque Control of Induction machine", International Conf. on Electrical Machines, Paris, France, 5-8 Sept 1994.

[6] T.G. Habatler, F.Profumo, M. Pastorelli, L. Tolbert,"Direct Torque Control of Induction Machines using Space Vector Modulation", IEEE Trans. on Ind. Appl., Vol.28, No.5, pp.1045-1053, Sept/Oct 1992.

[7] N.R.N. Idris and A.H.M. Yatim, "Reduced Torque Ripple And Constant Torque Switching Frequency Strategy For Direct Torque Control Of Induction Machine", In Conf Rec. IEEE-APEC, pp. 154-161, vol .1, 2000.

[8] Vojkan Kostić, Milutin Petronijević, Nebojša Mitrović, Bojan Banković, "Experimental Verification of Direct Torque Control Methods for Electric Drive Application", FACTA UNIVERSITATIS, Series: Automatic Control and Robotics, Vol. 8, No. 1, 2009, ISSN 1820-6417, UDC 621.313.33 004.424.451.2, pp. 111-126.

Session PO9:

PO9 - EDUCATION QUALITY

One Approach for Defining Students' Motivation in E-Learning

Donika Valcheva[1] and Margarita Todorova[2]

Abstract – **The success or failure of any e-learning initiative can be closely correlated to learner motivation. This paper presents a method for defining the Students' Motivation in E-learning, which uses the main concepts of the Keller's ARCS Model and the Gagne's events. An experiment is provided and some conclusions are made.**

Keywords – **Students' Motivation, ARCS Model, E-learning**

I. INTRODUCTION

The success or failure of any e-learning initiative can be closely correlated to learner motivation. Even the most elegantly designed training courses will fail if the students are not motivated to learn. Many students are motivated only to "pass the test.". The developers of e-learning course must strive to provoke a deeper motivation in learners to learn new skills and transfer those skills back into the work environment.

Some reasons for decrease of the students' motivation

- Learners can feel isolated.
- Difficult navigation within course.
- Confusing instructions for tasks.
- Irrelevant material for learners' needs and learning style
- Technical breakdowns.

As a first step, the e-learning course developers should ask the prospective learners questions such as:

- What would the value be to you from this type of course?
- What do you hope to get out of this course?
- What are your interests in this topic?
- What are your most pressing problems?
- What is your learning style?

The answers to these types of questions are likely to provide insight into learner motivation, as well as desirable behavioral outcomes.

MOTIVATION MODELS

According to [1,2,3] the most popular motivation models are:

- **The Time Continuum Model**

The model is presented in the form of a handbook for developing instruction and draws on approaches from linguistics, cognitive psychology, and motivation research. The model is not based on any one scientific theory or philosophy. Wlodkowski's Time Continuum Model of Motivation identifies three critical periods in the learning process where motivation is most important. Those periods are the beginning of the learning process, during the learning process, and at the end of the learning process. Each of those three periods has two distinct factors associated with it, yielding six basic questions to aid motivational planning. The factors to be considered at the beginning of the learning process are attitudes and needs. When planning the beginning of a learning experience, the designer should consider how the instruction will best meet the needs of the learners, and how a positive learner attitude can be developed. It is suggested that when possible, the instruction should focus on the physiological needs of the learners and experiences familiar or relevant to the learners. The instruction should allow for choice and self-direction in assignments. A needs assessment should be performed prior to developing the instruction to aid in appropriate planning. Stimulation and affect are to be considered during the learning experience. To maintain a stimulating learning environment, learner participation via questions, humor, varying presentation style using body language and voice inflection, and the use of different modes of instruction from lecture to group work to class discussion are strategies suggested. Wlodkowski's primary strategy is to make the learning experience as personalized and relevant to the learner as possible. Finally, competence and reinforcement are to be considered at the end of the learning experience. Frequent feedback and communicating learner progress are the author's main methods for developing confidence in the learners.

- **Keller's ARCS Model for Motivation and Gagne's events of instruction**

John Keller synthesized existing research on psychological motivation and created the ARCS model. ARCS stand for Attention, Relevance, Confidence, and Satisfaction.

Attention

The first and single most important aspect of the ARCS model is gaining and keeping the learner's attention, which coincides with the first step in Gagne's model. Keller's strategies for attention include sensory stimuli, inquiry arousal (thought provoking questions), and variability (variance in exercises and use of media).

Relevance

Attention and motivation will not be maintained, however, unless the learner believes the training is relevant. Put simply, the training program should answer the critical question, "What's in it for me?" Benefits should be clearly stated.

Confidence

The confidence aspect of the ARCS model is required so that students feel that they should put a good faith effort into the program. If they think they are incapable of achieving the

[1] Assistant Prof. Donika Valcheva, "St. Cyril and St. Metodius", University of Veliko Tarnovo, Veliko Tarnovo, Bulgaria, e-mail: donika_valcheva@abv.bg.

[2] Associate Prof. Margarita Todorova, "St. Cyril and St. Metodius", University of Veliko Tarnovo, Veliko Tarnovo, Bulgaria, e-mail: marga_get@abv.bg

objectives or that it will take too much time or effort, their motivation will decrease.

Satisfaction

Finally, learners must obtain some type of satisfaction or reward from the learning experience. This can be in the form of entertainment or a sense of achievement. A self-assessment game, for example, might end with an animation sequence acknowledging the player's high score. A passing grade on a post-test might be rewarded with a completion certificate.

This model is not intended to stand apart as a separate system for instructional design, but can be incorporated within Gagne's events of instruction.

Gagne's nine learning events are the most popular and effective model for creating e-learning contents. Gagne proposed that the content should have nine distinct instructional events to be effective. They are:

1. Gaining attention (reception)
2. Informing learners of the objective (expectancy)
3. Stimulating recall of prior learning (retrieval)
4. Presenting the stimulus (selective perception)
5. Providing learning guidance (semantic encoding)
6. Eliciting performance (responding)
7. Providing feedback (reinforcement)
8. Assessing performance (retrieval)
9. Enhancing retention and transfer (generalization).

II. METHOD FOR DEFINING THE STUDENTS' MOTIVATION IN E-LEARNING, WHICH USES THE MAIN CONCEPTS OF THE KELLER'S ARCS MODEL AND THE GAGNE'S EVENTS

For defining the students' motivation in e-learning, we use as a base the ARCS model and the Gagne events. The reason for this choice is that these models can be easier implemented and applied according to the specific nature of the e-learning process.

After finishing given e-learning course the students could be kindly asked to fulfill a questionnaire, based on the concepts of the Keller's ARCS Model and the Gagne's events, in order their motivation to be defined. The results of this investigation will be very useful for the course developers (teachers, trainers or software developers), because they will obtain important feedback information about the students' motivation and satisfaction after finishing the course. Thus the quality of the e-learning courses can be measured and if necessary the learning content can be modified. The questionnaire will consist of the following questions, divided into 4 sections, according to the Keller's ARCS Model and the Gagne's events: The scale that will be used will consist of the following possible answers:"Absolutely yes", "Yes, but not so much" and "Absolutely no".

Attention section

1. The course offered me appropriate for my learning style e-materials.
2. The interface design and navigation were easy to work with.
3. The visual aspect of the content (i.e. rite size and color of fonts, proper line spacing,, relevant diagrams, positioned at right places) has a positive impact on the accessibility of the content.
4. The objectives of the course are clearly stated.

Relevance section

1. The new content was based on my previous knowledge and skills in this field.
2. The received new information will be very important for my future work and study.
3. The course offer me links to additional information in the field.

Confidence section

1. During the course I felt myself sure I can manage with the problems.
2. During the learning process I received feedback and support from my teachers.
3. My success in this course is a direct result of the amount of effort I have put forth.

Satisfaction section

1. I am satisfied with the results of my study, after finishing the course.
2. I am feeling rewarded.
3. I will use the newly received knowledge and skills in my work.

III. SUMMARIZED RESULTS FROM THE PROVIDED INVESTIGATION

According to the suggested in this paper method, an experiment was provided. In this experiment were invited 20 bachelor students from Computer science specialty to participate. They assessed their motivation when they finished the e-learning course, titled Software technologies. They were asked to answer the presented in the paper questionnaire.

After processing the students' answers, the following results are obtained:

About the Attention section:

- Most of the students (13/20) think that the course offered them appropriate for their learning style e-materials;
- The half of the students assessed the interface design and navigation as easy to work with.
- Most of the students believe that the visual aspect of the content has a positive impact on the accessibility of the content and that its objectives are clearly stated.

About the Relevance section:

- Most of the interviewed students think that the new content was not so much based on their previous knowledge and skills in this field and it was a little bit difficult for them to start learning.
- The half of the students does not know for now where they can apply the new knowledge.
- 14 of the students think that the course offer them enough links to additional information in the field.

About the Confidence section

- During the course 11 of the interviewed student felt themselves sure they can manage with the problems
- During the learning process all of the students received feedback and support from their teachers
- Half of the students think that their success in this course is a
- direct result of the amount of effort they have put forth. The other 10 are not sure that this was the main factor for success.

About the Satisfaction section

- All of the students participating in this experiment are satisfied with the results of their study, after finishing the course.
- 12 of them are feeling rewarded.
- 15 students are not exactly sure where they will use the newly received knowledge and skills in their work.

Summery of the results

According to the results from the experimental inquiry it could be concluded that the e-learning course "Software technologies" motivates in grate degree the students. The strong sides of it are the high level of personalization, the feedback with the teachers, the good navigation and interface and the opportunity for additional information. The weak aspect of this course according to the students' opinion is that there is no enough information about the application in their real work. It is good if in the course description the developers add practical examples about where in the students' future work this knowledge will be useful. Also some information for the needed basic knowledge and skills before entering the course is necessary to be presented

On fig.1-8 are visualized the summarized results.

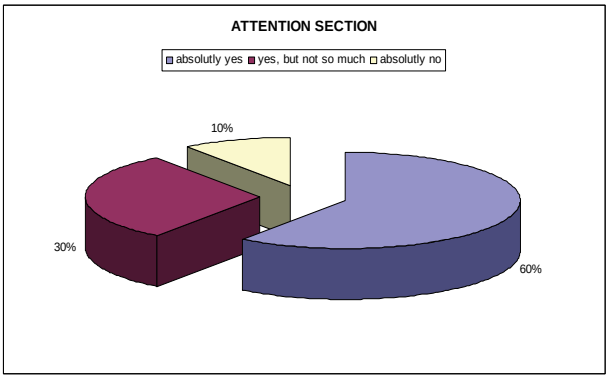


Fig. 1 Results from the students' opinion concerning the Attention in%

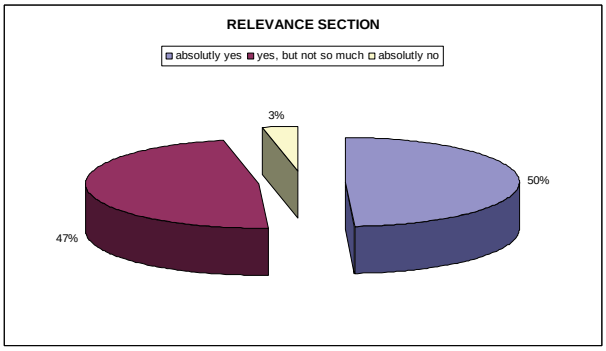


Fig. 2 Results from the students' opinion concerning the Relevance in %

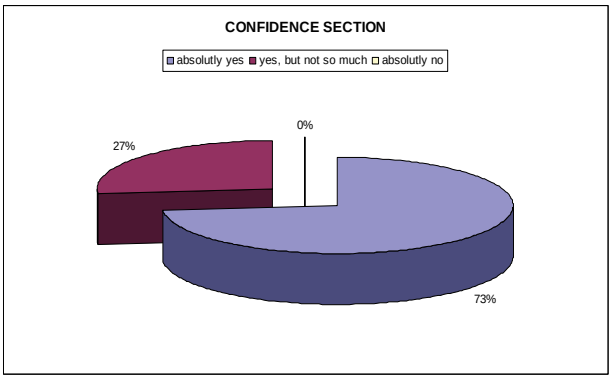


Fig. 3 Results from the students' opinion concerning the Confidence in %

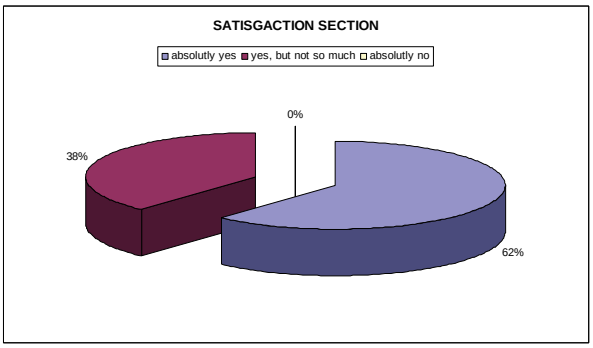


Fig.4 Results from the students' opinion concerning the Satisfaction in %

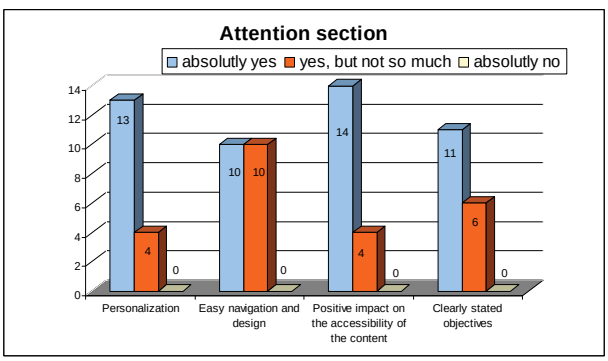


Fig.5 Results of the criteria in the Attention section

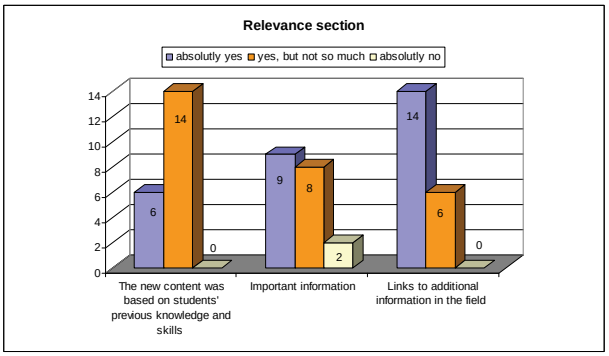


Fig.6 Results of the criteria in the Relevance section

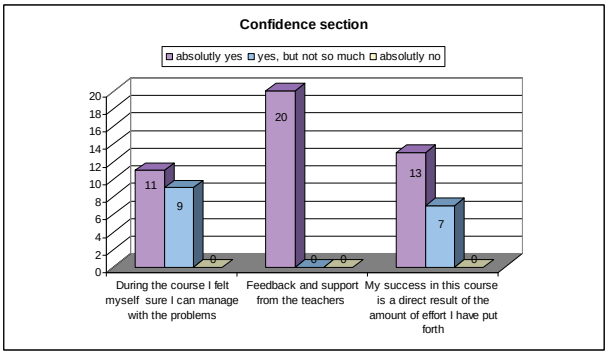


Fig.7 Results of the criteria in the Confidence section

Fig.8 Results of the criteria in the Satisfaction section

IV. Conclusion

One of the most important themes in psychology of learning is motivation. In order to include motivational factors in online learning, factors known as depending on the learner, assessment of the learner's motivation is required and this is the problem addressed by this research.

As a result from the presented in this paper research some important concepts for keeping the learners motivated could be summarized in the following list:

- Defining the target audience and their learning preferences;
- Course designers must realize that learning styles are different: visual learners, kinesthetic learners, auditory learners. E-learning courses must cater for all otherwise learners will lose interest;
- Defining clear learning objectives of the course;
- Use of interactivity/Games/Simulations - using interactivity in e-learning contents has many benefits. It keeps the learners involved, breaks the monotony of a single way communication, enhances the learning experience by participation and facilitates active experimentation;
- Use of real life scenarios - Cognitive Theories say that any new information is compared to existing cognitive structures called 'schema'. Meaningful information is easier to learn and remember. It is very important for the students to know where they can apply the newly received knowledge.
- Assessment of the students' motivation.

Acknowledgement

This paper is financed by the project: Creative Development Support of Doctoral Students, Post-Doctoral and Young Researches in the Field of Computer Science, BG 051PO001-3.3.04/13, EUROPEAN SOCIAL FUND 2007–2013. OPERATIONAL PROGRAMME "HUMAN RESOURCES DEVELOPMENT"

References

[1] Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), Instructional-design theories and models: An overview of their current status. Hillsdale, NJ: Lawrence Erlbaum Associates

[2] Keller, J. M. (1984). The use of the ARCS model of motivation in teacher training. In K. Shaw & A. J. Trott (Eds.), Aspects of Educational Technology Volume XVII: staff Development and Career Updating. London: Kogan Page

[3] Keller, J. M. (1987). Development And Use Of The ARCS Model Of Motivational Design. Journal Of Instructional Development, 10(3), 2-10. John Keller's Official ARCS Model Website

An Approach to Teaching "Software Design Patterns"

Violeta Bozhikova[1], Mariana Stoeva[2] and Veneta Aleksieva[3]

***Abstract* – This paper is about some problems of teaching "Software Design Patterns" in the Master's degree education and our approach to teaching this discipline. The paper underlines traditional and specific requirements respected by the approach developed and presents its main characteristics, based on the method of electronic textbook. Then, the approach for self-testing and examination of knowledge is discussed and some ideas for further development of the testing part are marked.**

***Keywords* – Electronic Textbook, Computer-based Learning, Teaching Strategy, Knowledge Testing and Evaluation.**

I. INTRODUCTION

Design patterns are attracting more attention now because they encapsulate valuable knowledge to solve recurring design problems and improve the quality of programming work. These architectural constructions for reuse become popular after 1994 with the book of Erich Gamma and al. [2]. A design pattern provides a general solution to a common problem in software design and is a language-independent description (or template) of the problem in general, which can be directly transformed into a code.

This paper is about our approach to teaching „Software Design Patterns. The course „Software Design Patterns" is included as an elective, in the second semester of the master's degree program for specialty Computer Systems and Technologies (CST), in the Technical University of Varna. The course includes lectures and practical exercises. The main topics of the course are related to the study of the three main groups of patterns (building, structural and behavioural) and the ways of their multiple use, their combination, their documentation and testing. Unified modelling language (UML) is mainly used for the pattern presentation.

It is seen that the course is based on extensive field of knowledge. Students should know the principles of object-oriented analysis and design, the unified modelling language (UML) and at least one object-oriented language. There are a lot of problems in teaching such a discipline: how to motivate the students to learn this complex technology, how to overcome the problem associated with the different background of the students, how to solve the problem of their employment (the majority of students enrolled in Masters Education are employed). A flexible teaching approach must be developed, an approach that encourages the students, regardless of their different background, an approach that holds the students attention and complies with their employment, providing them and individual workspace.

Section 2 presents the main characteristics of the proposed approach which is based on the method of electronic textbook: traditional and the specific requirements, respected by the teaching approach; structure and implementation of developed electronic textbook and specially – our approach to check the learning.

The last section presents our conclusions and future work.

[1]Violeta Bozhikova is with the Faculty of Computing and Automation, TU-Varna, Studentska 1, 9010 Varna, Bulgaria, E-mail: vbojikova2000@yahoo.com.

[2]Mariana Stoeva is with the Faculty of Computing and Automation, TU-Varna, Studentska 1, 9010 Varna, Bulgaria, E-mail: mariana_stoeva@abv.bg

[3]Veneta Aleksieva is only presenter of this paper and is with the Faculty of Computing and Automation, TU-Varna, Studentska 1.

II. OUR APPROACH TO TEACHING „SOFTWARE DESIGN PATTERNS"

A. Traditional and specific requirements to electronic teaching materials

We offer an approach for teaching software design patterns based on the electronic textbook. E-books (electronic textbooks) are computer-based systems [3, 4], mainly oriented towards training students, also examination and knowledge evaluation of the students. To electronic teaching materials have both traditional and specific requirements. As "traditional" we could define the following properties:

- Adequacy (in terms of curriculum) and completeness of the statement;
- Logic and coherence of the presentation.
- Accessibility statement of material
- Scientific character of the material.

To achieve the objectives of the course „Software Design Patterns" and to overcome the above problems, the electronic textbook, proposed by us has a structure shown in Figure 1 in order to respect the following specific requirements:

- To focus on the practical aspects of applying patterns in software development instead of theory: the aim is to motivate the students to use this technology and maintaining enough self-discipline to understand its general ideas. During laboratory exercises, students learn patterns through small examples check their knowledge (through tests) and finally - develop specific software tasks with higher complexity. The aim is, after graduation, students are able to independently design, develop, and document complete software solutions using a combination of patterns.
- To be based on examples written in C #.Net, a language that is studied in the previous semester in the Master Course.
- To focus mainly on individual work rather than teamwork.
- To be bilingual (to support Bulgarian and English language) in order to serve as a textbook on the subject "Software Design Patterns" for Bulgarian-speaking and English-speaking students.

- enable self-testing and self-assessment of the student's knowledge at different levels; for each topic, for each module and for the whole material;
- be convenient for maintenance and future evolution; Since software design patterns are numerous, and the developed electronic textbook focuses on the universal patterns only [5], the architecture of the electronic textbook is simplified to allow easy further development without impairing the quality in. Each topic presents a pattern and is stored in separate .rtf file when the tests are stored in .txt files (shown in the left part of Figure 2). Thus the change and complete replacement of the content of a pattern and the adding of new patterns is easy, with minimal technical effort, without program intervention in the system itself.

B. Structure and implementation of developed electronic textbook

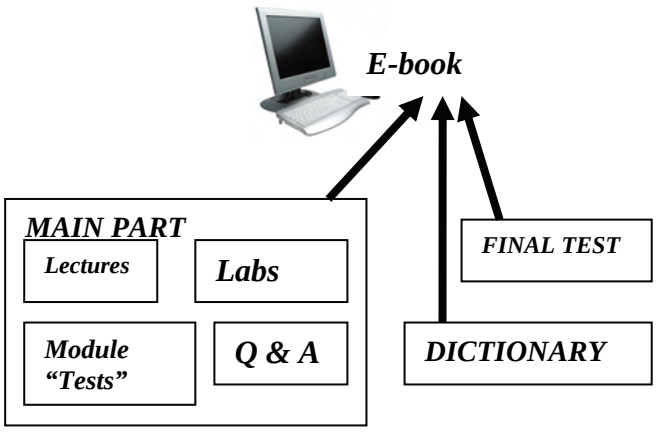


Fig. 1. The generalized structure of the created electronic textbook

The main modules (see figure 1) of the created electronic textbook on the subject " Software Design Patterns" are:

- MAIN MODULE: the most important part of the book, which exposes the contents of the course on modules and themes, each theme includes descriptions of lectures and laboratory exercises. This part is divided into three main modules, according to the three main groups of the patterns (building, structural and behavioral). Each lecture and each laboratory exercise are accompanied by graphic illustrations (UML diagram, etc.). Each laboratory exercise includes a description of assignments of tasks to be performed in order to assimilate the material in practice. This part includes also Q & A MODULE with frequently asked questions and answers.
- TEST MODULE: includes questions about self-assessment at different levels: at topic level, module level and final test level;
- MODULE DICTIONARY: includes a glossary of terms;

Figure 2 shows the main window of the electronic textbook, realised as a desktop C#.Net application.

C. Our approach to self-testing and examination of knowledge

Module "Tests" is used by the students to self-checking of learning. There are different ways of organizing test questions. Our module randomly generates a sequence of questions ("close" or "open"), which answers can be given in one of the following ways:

a) For "close" question: By selecting an option (or options) from a list of answers (menu), in which each question provides a list of correct and incorrect answers (figure 3). The student selects one or several answers. The system displays the correct answer but does not provide an assessment of the learner. In the wrong answer to this type of question, the system provides the both the correct answer and the wrong answer (figure 5) - displayed in red.

b) For "open" question: Response in the form of text (figure 4). The system displays the correct answer in the form of "reference text" but does not provide an assessment of the learner (figure 6). The ability to see the correct answers gives the learner a real idea of the extent of assimilation of the material.

Students can go back and change answers to previous question before the test is completed.

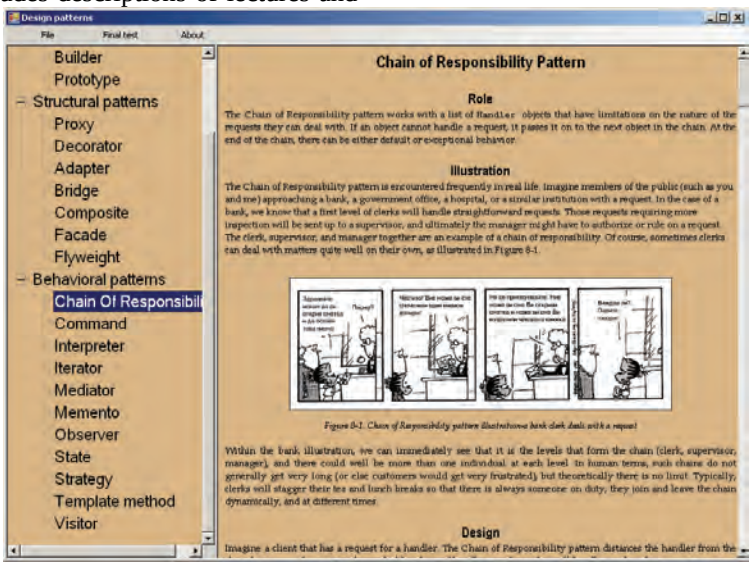


Fig. 2.The main window of the electronic textbook

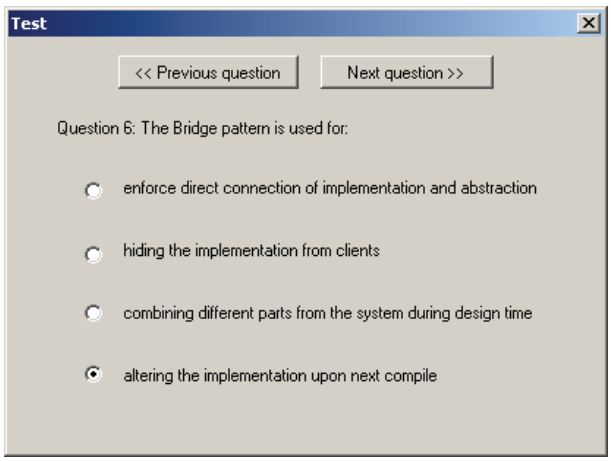


Fig.3. A “close” test question

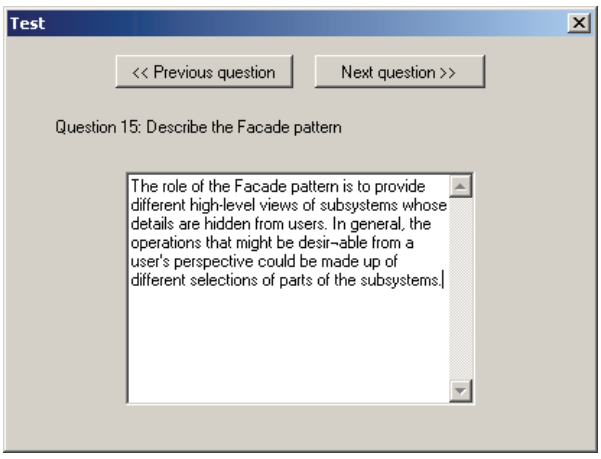


Fig.4. An “open” test question

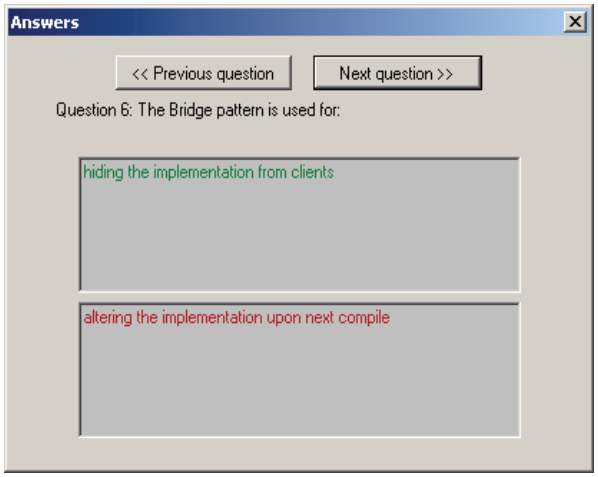


Fig. 5. A “close” question answer

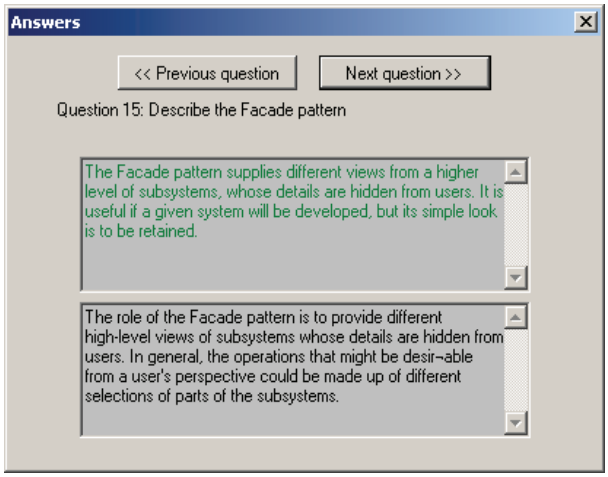


Fig. 6. An “open” question answer

III. CONCLUSION

In this paper, we discuss some problems of teaching “Software Design Patterns” in the master's degree education of specialty Computer Systems and Technologies in the Technical University of Varna. Based on our experience and observations in teaching similar courses we could claim that the discipline "Software Design Patterns" is undoubtedly useful for students in master's degree. It is because design patterns are elegant solutions to typical problems in the software design, with the possibility of reuse. They are a way to increase efficiency and quality of programming work in developing advanced software applications.

Next, the paper presents our approach to teaching „Software Design Patterns" - an approach based on the method of an electronic textbook. The paper highlights the traditional and the specific features of the developed electronic textbook and comments its structure, its implementation and mainly - the approach used for self-testing and student’s knowledge examination. We could also argue that the proposed teaching approach for the course "Software Design Patterns" is innovative and effective. This is because it is computer based, i.e. it is based on the latest teaching methodology both to the learning process, and also to students who are trained. The created textbook on the subject "Software Design Patterns" contains material for various levels of complexity, it is much more compact than traditional printed textbooks (it is collected in one CD) and it is accessible from any workstation (it is installed on each computer). It focuses on practical examples instead of theory. Based on our more than twenty years practice in Computer Sciences and Engineering Department of the Technical University in Varna we are persuaded that accentuating on practical examples instead of theory is the best way of motivating the students to use some technology. It provides personalized information space for each student and much higher visibility than the traditional printed textbook. It provides a variety of tests with varying degrees of complexity in interactive teaching mode; there is a feedback - in an incorrect response from the student, the system offers the right answer.

There are many directions for development and improvement of the textbook developed, although it complies with both the traditional and the specific requirements. Let mention only some of them: adding audio, video fragments and animations in MAIN PART; adding new software patterns; adding of recommendations and examples to assist

students when implementing the stated laboratory tasks; transition from Desktop implementation to WWW-based application. There are also many opportunities to improve the testing module: firstly, the e-book has to enable the teacher to control the absorption of knowledge, next - implementation of a system for assessing the acquired knowledge is needed. The evaluation of the students' knowledge for each “close” question could be realized using for example the formula, given in [4]: $y = r / (N + f)$, where r is the number of correctly selected items from the list of responses, N - number of correct answers in the list, f - number of incorrect items from the list of answers. Criterion for the correctness of the answer for an open question might be the presence of a number of key words in the reference text

All these ideas for further development aimed at achieving even greater visibility of the presented teaching material, even greater completeness of the contents of the course and accordingly: higher efficiency of the learning process.

ACKNOWLEDGEMENT

This work is supported by a scientific research project in Technical University of Varna.

REFERENCES

[1] Shirley Tessler, Avron Barr, Nagy Hanna, National Software Industry Development: Considerations for Government Planners, http://www.aldo.com/Publications/Papers/National_SWI_Development_050303.

[2] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software, http://www.uml.org.cn/c++/pdf/DesignPatterns.pdf

[3] E-book http://en.wikipedia.org/wiki/E-book

[4] Норенков И.П., Информационные технологии в образовании, http://bigor.bmstu.ru/?cnt/?doc=Default/050_iteduc.cou

[5] C# 3.0 *Design Patterns*, Judith Bishop, O’Reilly Media, 2008

Quality Monitoring in Higher Education: Elements of a Software Support System

Danijel Mijic[1] and Dragan Jankovic[2]

***Abstract* – This paper describes the basic elements of the software support system for quality monitoring in the area of higher education. We described the current situation at University of East Sarajevo, problems and shortcomings of existing solutions and the motivation for the development of new solutions. We proposed several suggestions for improvements. Some of the segments of the system for monitoring the quality of the teaching staff are also described.**

***Keywords* – higher education, quality, software support.**

I. INTRODUCTION

Quality assurance, control and monitoring are the processes that are continuously taking place in all areas of human work in order to improve quality of products and services. Minimum quality requirements are defined by the applicable standards in the appropriate workfield. Quality standards in higher education usually define the minimum requirements that higher education institutions must meet in order to satisfy the basic criteria of the quality of their work. In addition to requirements for the quality of institutions themselves, requirements for the quality of the study programmes, that are realized at higher education institutions, are also defined. In order to monitor the quality it is necessary to identify relevant quality indicators and monitor their values over time. Based on the results obtained by monitoring the values of quality indicators it is necessary to take appropriate actions that should improve quality. The effects of the activities undertaken are assessed based on comparison of these parameters before and after their implementation, afterwards appropriate decisions on the next steps are made.

Quality of products or services in any industry is affected by many factors. The quality of a product depends on the quality of raw materials and the quality of the entire production process. A similar analogy is valid and can be applied in higher education, where the quality of graduates and their output level of qualifications and skills, is affected by factors such as level of knowledge at the time of enrollment, quality of study programmes, physical and human resources involved in the teaching process, student motivation to learn the course material, and so on. For monitoring quality in higher education there is a need to constantly collect and analyze vast amounts of data about the teaching process, and to get precise information from this data that is necessary for making the right decisions to improve quality. Data to be collected and analyzed usually include data about students, teachig staff, material resources which the institution has at its disposal, various elements of the teaching process, and data related to administrative and business activities of a higher education institution.

It is clear that the use of modern information and communication technologies (ICT) in this area is imperative. The usage of ICT enables more efficient and easier way to carry out activities of collecting, processing and analysis of relevant data needed to monitor the quality, as well as its adequate presentation. In the case of possibly large higher education institution, especially if it is heterogeneous and distributed and if there is a large number of external associates (as it is the case with the University of East Sarajevo), efficient data collection and its processing and analysis is possible only in the case of properly designed, implemented and used special-purpose software. In this paper, the situation at the University of East Sarajevo (UES) in terms of software support for quality monitoring activities is described. We presented some of the elements of a software system for monitoring quality, with special emphasis on monitoring the quality of teaching staff as one of the key factors for quality assurance which greatly affects the students (eg how the teacher motivates students to work in a given subject).

II. QUALITY MONITORING AT UNIVERSITY OF EAST SARAJEVO

UES consists of 17 organizational units that are distributed across the eastern part of Bosnia and Herzegovina (BiH). The University operates on the model of integrated university in the legal sense, but in geographical terms it is highly distributed since its 17 organizational units are located in 8 different cities. Similar to other higher education institutions in BiH, more intensive activities in the field of quality assurance at UES began only in recent years. The first self-evaluation report for UES and all of its organizational units was made in 2009. Starting with the year 2010, the report is prepared in accordance with the Criteria for accreditation of higher education institutions in Bosnia and Herzegovina [1], in the form defined by the State Agency for Higher Education Development and Quality Assurance in BiH (the Agency).

At the time of writing, there is no clearly defined set of quality indicators of higher education institutions in BiH or study programmes that are realized in them. Document containing the reference standards against which accreditation of higher education institutions will be made, was adopted during the year 2010 by the Agency, in the form of the framework of criteria for accreditation of higher education institutions without precisely defined quality indicators. Criteria for accreditation of study programmes are still in the

[1]Danijel Mijic is with the University of East Sarajevo, Faculty of Electrical Engineering, Vuka Karadzica 30, 71123 Istocno Sarajevo, BiH, E-mail: danijel.mijic@etf.unssa.rs.ba.

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

process of development. Therefore, higher education institutions a left with a lot of freedom in the interpretation of criteria for accreditation and selection of quality indicators. In this sense, the UES selected the quality indicators based on the criteria for accreditation of higher education institutions. Selected indicators of teaching process quality are attached to the self-evaluation reports in the form of additional statistical reports containing information on students, student success rates at exams, teaching and administrative staff, material resources and scientific research.

Part of the required information is available in the official records of the UES, in the current information system databases, while a significant portion of the data is not collected in a uniform and systematic manner. The UES information system (IS) is a software solution based on open-source technologies. It is used individually at all organizational units of the UES. Some of the shortcomings of the current version of the IS, in the context of organizational structure of the UES, are that there are no aggregate data on the university level, nor the system can support the creation of statistical reports in the form required to monitor quality. In order to provide the necessary data in an efficient way, the development of new and improvement of the existing modules of the IS is initiated, for the purpose of the software support to quality monitoring.

III. Software Support for Quality Monitoring

In order to solve the previously described problems, the development of software systems to support quality monitoring was initiated. One of the main functions of the system is the integration of the existing data, from the individual databases of organizational units, in a central repository of data that is used to monitor quality indicators and generate reports at the university level or the individual organizational unit level. Another important feature is the recording of data for which there is currently no electronic record.

A subsystem for integration of the existing data implements the functionality for extraction, transformation and loading a central data warehouse at the university level. Information stored in the data warehouse includes data about students, teaching staff, study programmes and course. The data on teaching staff also stores the results of student surveys for evaluation of the teaching quality. Software tools based on the application of business intelligence are used for analysis and presentation of these data. The data from the data warehouse was processed and stored in the form of OLAP cubes. Several OLAP cubes were designed and implemented for analysis of data on various aspects of the teaching process.

For analyzing data about students, we realized OLAP cubes that enable the following analysis:

- analysis of the quality of students enrolling the university, based on their succes in the high school,
- analysis of data about students at the level of university, organizational unit or lower levels,
- analysis of success of students at exams and assesments,
- analysis of the average total length of study and success of students in their studies,
- analysis of the rate of student passing from one year of the study to next higher year of the study.

Each of the mentioned OLAP cube provides an analysis of data about students by a number of criteria, such as year of entry to the study, study programme, gender, type of funding, year of the study, student success from high school, type of high school, place of birth, place of graduating from high school, etc.

An important segment of the system for monitoring the quality is a module for the analysis and monitoring of the quality of the teaching staff. One indicator of the quality of the teaching staff are teacher ratings in student surveys. Collecting student opinions about the quality of teachers is done electronically, using a special web application used for conducting the student surveys at the university level, which is used from the academic year 2009/10 [2]. Faculty of Electrical Engineering of the UES has been using this application for more than five years. The application enables tracking results and trends of teachers ratings in the student surveys. In addition to teachers, students are using this application to evaluate the quality of their study programmes, the quality and availability of other resources used to support the educational process. In the next chapter we will give more details about the system for monitoring the quality of the teaching staff.

For a more detailed analysis of data about the quality of the teaching staff, based on the results of student surveys, we designed and implemented the OLAP cube that enables the analysis of teacher ratings on an arbitrary level (university, organizational unit, study programme, year of study, course unit) and the arbitrarily chosen criteria (type of involvement of teachers , teacher gender, student gender and other criteria) [3].

Another important indicator of the quality of the teaching staff are publications, participation in projects and other results of scientific research. In addition to individual results related to the teachers, it is required to monitor the aggregate performance of university and organizational units in the field of scientific research. Unfortunately, this segment is not covered by the current functionalities of the UES IS. Module for recording the results of scientific research of teachers is one of the additional modules of IS whose development is recently started, and whose implementation is expected by the end of 2012.

Other additional modules of the quality monitoring system, whose implementation is planned by the end of 2012, are a module for recording teaching activities of students and teachers by application of RFID technology, and a module for tracking employment of the UES graduates.

Module for recording teaching activities of students and teachers should enable automated recording of the student attendance to the classes, number of classes perfomed by teaching staff, and also recording the usage of the resources for teaching process like classrooms and laboratories. Implementation of this system should contribute to the quality of teaching process through the automated records of

important data about the educational process. Based on the analysis of these data it is possible to identify certain problems and weaknesses in the teaching proces and to influence on their elimination in order to improve the overall quality of the teaching process.

The quality of the study programmes can be monitored using several indicators. Some of them have already been mentioned and are related to the indicators of success of students, quality of teachers, quality and availability of material resources for the implementation of study programmes. Some additional indicators of quality of study programmes are related to employment of graduates. Based on data about employment of graduates, such as areas of work in which graduates are employed, the average employment rate, the average waiting time to employment, one can conclude about the quality of study programmes. These data can not be acquired easily, especially if not adequately using the possibilities offered by a modern information and communication technology. It is necessary to maintain the relationship with graduate students after graduation, and to collect and analyze data relevant to the analysis of the quality of study programmes. This need is the motivation for the development of additional modules of the system for quality monitoring, which will enable the collection, recording and analyzing data on employment of the UES graduates, development of their careers and obtaining feedback from relevant areas of the profession in order to improve the quality of the study programmes.

IV. Monitoring the Quality of Teaching Staff

An important segment of monitoring quality in higher education is monitoring the quality of the teaching staff. Evaluation of teachers is common practice in many universities around the world. The most commonly used instrument for collecting students' attitudes about the quality of the teaching staff are student surveys carried out in written or electronic form. In recent years, many universities use web based online surveys because of the many benefits that they provide compared to paper based surveys [4,5].

All organizational units of the UES have been conducting student evaluation of the teaching quality in the last few years. At the beginning, the surveys were performed in different ways, using the paper form or electronically, without the use of unique forms and unique procedure. Starting from the 2009/10 academic year, this process is uniformed so the entire university uses the same system for electronic survey of students, and surveys are conducted in the same way in all organizational units. The content of the questionnaires and the software system for online students' surveys are constantly improved in accordance with proposals and recommendations of quality assurance bodies at the UES. Besides the assessment of the quality of elements of the teaching process, the questionnaire contains a section with general information about the participants of the survey, such as gender, type of funding and the average grade on the exams, which can be used for additional analysis of survey results according to profiles of participants. It is important to note that students' anonymity is guaranteed during the process and there is no way to establish a connection between the ratings given by students and their identity. This is very important for students to objectively evaluate all elements of the teaching process, especially the teaching staff, without fear of beeing punished in any way if they give negative comments or ratings.

Web application for student survey was developed at UES Faculty of Electrical Engineering. It was first put into use in the summer semester of the academic year 2006/07. The architecture of the application is shown in Fig. 1.

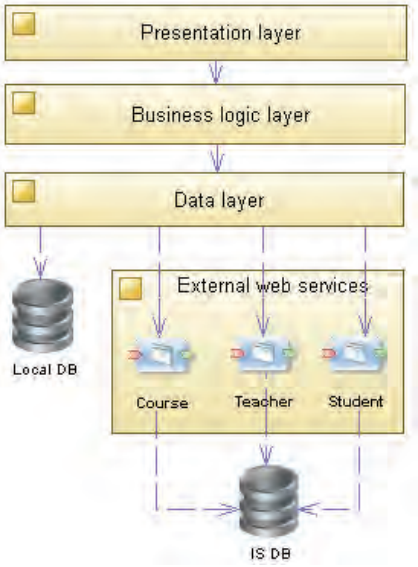


Fig. 1. Architecture of web application for students' survey

Analysis of teachers' ratings becomes more important if the results for one teacher are compared with results for other teachers who were evaluated by the same group of students, or if they are brought in connection with other factors that have a certain influence on teachers' ratings.

Comparison of teacher's ratings with an average ratings of teachers in the same study group, it could be seen that the ratings of teachers to some extent depend on the characteristics of the student population who participated in the survey, or to the performance of teachers in a given period. If the trend of ratings of individual teacher follows the trend of average ratings of all teachers at the same study group, it can be concluded that small variations in the teacher's ratings are probably not the result of degradation or improvement of the performance of teacher, but could be caused by different characteristics of the student population. An example of the individual teacher's ratings on one course unit, compared with an average grade of students on the same course unit and with the average ratings of all teachers evaluated by the same group of students is shown in Fig. 2. It could be noticed that the trends are similar which indicates the that teachers' ratings depend on certain characteristics of the student population. Smaller ratings in academic year 2010/11 are the result of change in scale for the evaluation. Since the 2010/11 academic year, five degrees scale from 1 to 5 is used, while in the previous period we used six-grade scale with ratings of 5 to 10.

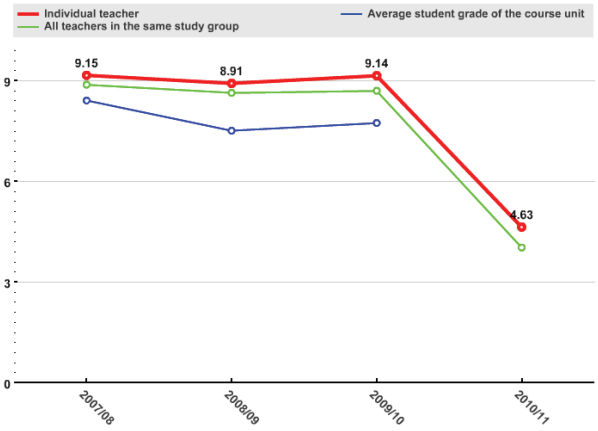


Fig. 2. Teacher's ratings for one course unit compared to student's grades on a given course unit and to overall teacher's ratings for all course units of the same study group

Comparing the trend of ratings of the individual teacher for all his course units, with the trend of overall teachers' ratings of the faculty is shown in Fig. 3. The discrepancy in the trends is understandable considering the fact that a group of students who rated the work of individual teacher differs significantly from the rest of the student population who participated in the survey and rated the work of other teachers.

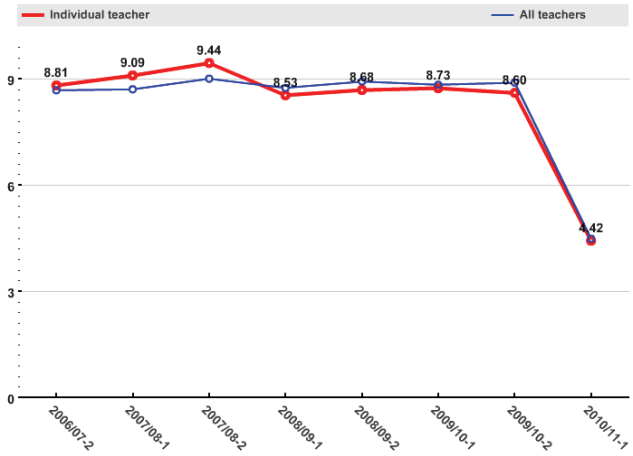


Fig. 3. Individual teacher's ratings compared to overall teachers' ratings

Comparing the trends of teacher's ratings with the trends of certain characteristics of the student population participating in the surveys may reveal additional relationships and dependencies and to allow better analysis of results. Some research shows that the average grade of students participating in the survey affects the ratings of teachers [6]. Also, the average rating of teachers is higher if the study group of students is less. Some of other factors that could influence the average ratings of teachers are teacher age, teacher gender, students gender and so on. In order to analyze the ratings of teachers from some of these aspects, it is necessary to have information about the profile of the student population participating in the survey. Some of the information can be obtained during the survery process without compromising the anonymity of students, as is the case in the questionnaire with general information about the respondents used at the UES. More reliable and accurate data about the student population could be obtained from official data about students from the UES information system. The data that can be obtained from IS refers to all students who are eligible to participate in the survey. Since the survey process ensures the anonymity of students, it is not possible to establish a direct connection between the characteristics of the student population and the results of the survey, but this information still can be useful for analysing the results.

V. Conclusion

This paper describes several software tools used for software support to the quality monitoring at the UES. The main advantage of these tools are capabilities of integration of existing data from various segments of the teaching process and insight into the key indicators of the quality of the teaching process at the university level. By analyzing trends of quality indicators it is possible to monitor the quality of certain elements of the teaching process and conclude about the effects of the measures introduced to improve quality.

In addition to software tools that have already been developed and used at the UES, we mentioned other tools whose development is in progress and whose role is to contribute to automating the process of collection, analysis and presentation of data from the teaching process. These new tools will provide means for monitoring additional quality indicators, which in turn will ensure that the process of monitoring and improving the quality of the UES to become more efficient and effective.

References

[1] *Decision on Criteria for Accreditation of Higher Education Institutions in Bosnia And Herzegovina*, available online at http://hea.gov.ba/Kvalitet/dokumenti/?id=1315, April 2011.

[2] D. Mijić, "Measuring Teaching Quality in Higher Education: Instrument For Collecting Student Feedback", ICT Innovations 2010 (Editor M. Gusev), Web proceedings, pp.117-128, Ohrid, 2010.

[3] D. Mijić, D. Janković, "Application of Business Intelligence in a System For Software Support To Monitoring Quality at Higher Education Institutions", 18th Telecommunication forum TELFOR, pp. 1297-1300, 2010.

[4] K. Hoffman, "Online Course Evaluation and Reporting in Higher Education", New Directions For Teaching And Learning, no. 96, 2003.

[5] H. Anderson, J. Cain, E. Bird, "Online Student Course Evaluations: Review of Literature and a Pilot Study", American Journal of Pharmaceutical Education, 2005.

[6] D. L. Sorenson, C. Reiner, "Charting the Uncharted Seas of Online Student Ratings of Instruction", New Directions For Teaching And Learning, no. 96, 2003.

Advantages, Structure and Capabilities of the Electronic Assessment System

Mariyana Nikolova[1] and Margarita Todorova[2]

Abstract - **The basic structure, the possibilities and advantages of e-assessment system are discussed in this paper. Architecture of the system for e-assessment is presented in two modules: Database examination system and Database management system. System requirements for e-Assessment are examined as well.**

Keywords - **Education, e-Assessment, the structure e-assessment system, e-test, type of question, advantages of the e-Assessment.**

I. INTRODUCTION

The information society is built on the base of million information systems that realize variety of functions in different fields. The construction, development and operation of these systems require specific approaches and skills to apply the modern information technologies in the particular area. This situation requires individuals to adapt their skills and competencies. Consequently, educational objectives and societal expectations have changed significantly in the recent years. Modern learning settings have to consider learning community aspects as well as learner-centered, knowledge-centered and assessment-centered aspects.

Educational community has also changed and updates the forms of training and testing. The development of a good information system for training and education significantly increases the opportunities for improving the quality of the teaching process. It could work effectively only if all participants in the educational process take part in it, and especially the leading personal.

The assessment is an important component of the teaching and learning processes. It is necessary organizational prerequisites to be created, consistent with the specific school, as well as support the from the management representatives.

Technology can support almost every aspect of assessment somehow - from the administration of individual tests and assignments to the management of assessment across a school, faculty or institution; from automatically marked on-screen tests to tools for students` support and feedback.

In order the technology-enhanced assessment to be effective, pedagogical applications need to be supported by robust and appropriate technology, within a supportive institutional or departmental context.

[1]Assistant Prof. Mariyana Nikolova, PhD, Department of Computer Systems and Technologies Department, St. Cyril and St. Methodius University of Veliko Turnovo, Bulgaria, E-mail: mnikolova_vt@abv.bg

[2]Assoc. Prof. Margarita Todorova, PhD, Head of Computer Systems and Technologies Department, St. Cyril and St. Methodius University of Veliko Turnovo, Bulgaria, E-mail: marga_get@abv.bg

This paper is financed by project: Creative Development Support of Doctoral Students, Post-Doctoral and Young Researches in the Field of Computer Science, BG 051PO001-3.3.04/13, European social fund 2007–2013г. operational programme "Human resources development"

II. STRUCTURE OF THE E-ASSESSMENT SYSTEM

The formative e-assessment is understood as the use of ICT (Information and Communication Technologies) to support the iterative process of gathering and analyzing information about process of learning by teachers as well as by learners for evaluating the results in relation to prior achievement and attainment of intended, as well as unintended learning outcomes.

The system for e-assessment comprises two components: an assessment engine and an item bank. An assessment engine consists of hardware and software required for creation delivery of test. Most e-testing engines run on standard hardware so the key characteristic is the software's functionality [7].

There is a wide range of software packages for e-testing. E-assessment system that provides analysis and statistics include many specific software modules. The software does not include pure questions; these are provided by an item bank. Once created, the engine uses item bank to generate a test. Traditional paper-and-pencil testing is similar, but the test is generated from the bank at only one time, when it is sent to publishing.

The creation of the item bank is more expensive and time consuming than the installation and configuration of the assessment engine. This is due to the fact that the assessment engines can be bought "off the shelf" whereas an item bank must be developed for each specific application.

An e-assessment system designed to focus on more sophisticated forms of knowledge requires some sort of interactive activity and a system for inviting students to reason or solve problems around that activity [7]. The structure of the system for e-assessment is presented in Figure 1.

The e-assessment system includes the following modules: Database management system (DMS), Databases examination system (DES) and user interface for teachers, student and system administrator. The basic task of this paper is organization of DES. We suggest it to contain two modules: Storage of questions and Repository of the assessment results.

Storage of questions is the important part of the system. It stores questions that will be use by the students. Construction of an unambiguous, productive and unrepresentative questionnaire is a difficult process. It is very important all possible options to be offered by the teacher [2], [4]. The questions can be used in various tests once created. This requires teachers to add information for each question. Let call this information code.

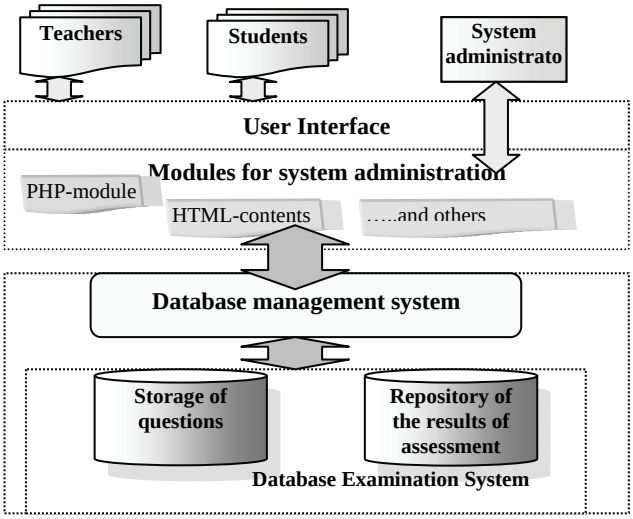


Figure 1. The structure of the e-assessment system.

For each question is entered code by teacher with the information for: **Discipline, Section, Class, Difficulty**.

The types of questions can be very different:

✓**Direct** - they offer the respondent to express his own position.

✓**Indirect** - to the respondent is given an opportunity to express agreement or disagreement with the position of other people.

✓**Questions filters** – They are called so because it is possible to select the respondents according to given indicator. For example - gender, age, profession, etc.

✓**True - false statement** - suggest two mutually exclusive response options (type 'yes-no ").

✓**Question-menu** - requires answers when the respondent can choose a combination of variants of answers.

✓**Questions - rocks** – those questions which answer is putting in order something in preliminary determined scale.

✓**The Table issues** - suggest as a response to fill table.

✓**Open questions** - do not contain any version of the response, a respondent answers his opinion in certain place in the questionnaire.

✓**Half-questions** – part of the variants of the answers are preliminary suggested, but the student may write something additional.

There are some special rules for formulation of questions and answers, compliance with which ensures maximum reliability of the answers of the respondents [1], [3].

The most important principles of questions are:

✓ Questions and answers must not contain suggestion that one way or another direct to some answers or make them more desirable or more prestigious.

✓ Creating artificial opportunity some of the answer to be given more frequently than others;

✓ When the question is "closed" (i.e. there are pre-defined options), all possible cases, have to be predicted;

✓ Question and all answers must be formulated in an equivalent manner so that various responses to have the same conceptual value;

✓ The answers should form a unified scale and relate to one and a same sign;

In education a well constructed survey can be a tool for feedback, which can significantly enhance quality of education. In this case there must be adequate and rapid response as a result of the aggregated answers of students in order to change their approach and training institutional.

III. Life Cycle of Creation of the E-tests

Life cycle of creation of the tests has three phases [5].

I. Preparatory phase:

- Target;
- Design of questions, answers and practical tasks;
- Determination of weights indicator - determining the difficulty of the problem and which part of the material relates;
- Design Test;
- Defining structural variations for different types of tests - linear (conventional), adaptive (in difficulty), interactive;
- Written instructions for students.

II. Operating phase

- Administration of various types of tests - optional version control;
- Providing a test dynamic or static, with the possibility to derive the correct answer, visualization of information relating to the material of the wrong question;
- Navigation.

III. Evaluation phase

- Scoring, evaluation and analysis of results;
- Assessing the quality of knowledge;
- Extraction the necessary data for statistics (for the management of educational process).

The repository of the results from assessment stores (Figure 2):

- e-tests;
- E-portfolio;
- Another type of theoretical and practical examinations

Just like traditional testing, e-Test is a method of assessing learners' ability to meet the required standards. The difference is that e-Tests are extracted from a computer, rather than pen and paper.

In the electronic portfolio (e-Portfolio) [8] the learners could upload and submit except all other information for themselves, also their work – term papers, solved problems and other tasks assigned by teachers. Unlike traditional paper based methods, e-Portfolios provide much richer and varied ways of recording and presenting proofs about their knowledge and skills. Learners can submit a range of file formats including word processed documents, spreadsheets, images, video and sound files. Then the content of an e-Portfolio can be shared with others.

Another kind of examinations that teacher provides could be practical and oral excises, projects, etc.

All that is needed to work with the system for electronic verification and assessment is PC or laptop with Internet access or LAN, where the server of the system is located. No specialist software or hardware is necessary because everything you need may be accessed via the network. Everything is access on the network via a user account, i.e. after registration and granted access to the system for electronic verification and assessment.

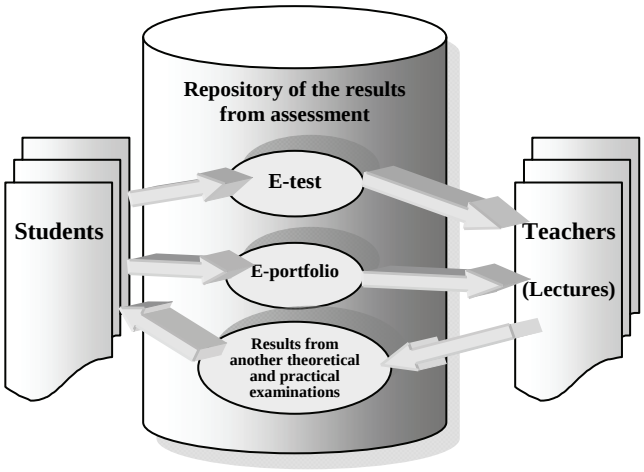


Figure 2. The repository of results from the assessment stores

Link between the teacher and student and Databases examination system

Architecture of the link between the user and database consists of three components: server, browser and client database. Handling and recording data is realized by using scripts that run the (web) server. User connects to the web server using a client browser. The server receives the request and processes it by script.

CGI (Common Gateway Interface) [9] is a specification for the interaction of Web-server with other applications. A typical CGI-program run of Web-server implementation of the task,returns the results to the server and ends his performance.

IV. The Advantages of E-assessment

E-assessment is becoming widely used. It has many advantages over traditional (paper-based) assessment. The advantages include [3], [6], [8]:

- **Richer assessment experience** – questions can be made clearer and more detailed through the use of text, sound and video which can aid motivation. For example, e-Portfolios allow the use of digital video, animations, presentations etc to be submitted electronically for assessment.
- **Increased flexibility** – assessment can be provided at a greater range of locations. This means assessment on-demand can become completely achievable.
- **Instant feedback** – results are often available within minutes of taking an e-Test, as well as diagnostic information on a learner's performance, highlighting areas that can be improved upon.
- **Reduce the administration burden** – fewer paper forms to complete, no posting of test papers.
- **Greater storage efficiency** - tens thousands of answer scripts can be stored on a server compared to the physical space required for paper scripts.
- **Enhanced question styles** which incorporate interactivity and multimedia.

There are also **disadvantages**. E-assessment systems are expensive to establish and not suitable for every type of assessment (such as extended response questions).

The best examples follow a Formative Assessment structure and are called "Online Formative Assessment". This involves making an initial formative assessment by sifting out the incorrect answers. The author/teacher will then explain what the student should have done with each question. It will then give the student at least one practice at each slight variation of sifted out questions. This is the formative learning stage. The next stage is to make a Summative Assessment by a new set of questions only covering the topics previously taught.

V. Conclusion and Future Work

The term e-assessment is becoming widely used as a generic term to describe the use of computers within the assessment process. Specific types of e-assessment include computerized adaptive testing and computerized classification testing.

E-assessment can be used to assess cognitive and practical abilities. Cognitive abilities are assessed using e-testing software; practical abilities are assessed using e-portfolios or simulation software

Assessment systems may support parts or the entire chain of the assessment lifecycle. This lifecycle includes authoring and management of assessment items, compilation of specific tests, performance of assessments, and compilation and management results.

The use of the different forms of control and assessment of knowledge and the e-technologies give the opportunity for:

- The results of the assessment and tasks performed by students to be saved in e-form;
- Free access to the tests and the results of the assessment process;
- Full and complex analysis of the results from the teaching;
- Oopportunity for feedback and adjusting the structure of curriculum and technology training.

In order to be developed a model of an optimal assessment system and the components, their characteristics and structure to be defined. Also an open assessment system may be developed. This system should consist of database with the results from the assessment with the classical and electronic methods, data base with the different types of questions (static and dynamic), mechanism for generation of different tests, according to the complexity, mechanism for access and of the results from the statistical analysis.

References

[1] Carles D., C., G. Joughin & M.M.C. Mok, Learning-oriented assessment: principles and practice, Assessment & Evaluation in Higher Education, 31(4), pp. 395-398, (2006).
[2] Carless D., G. Joughin, N.F. Liu, and Associates, How assessment supports learning: learning-oriented assessment in action. (Hong Kong University Press, 2006).
[3] G. Rodríguez etc., "Developing the *e*-Learning-oriented *e*-Assessment", Research, Reflections and Innovations in Integrating ICT in Education, pp. 515-519, (2009).
[4] Савчук, В.П., Журнал "&Стратегии"2004
[5] Nikolova M., Todorova M., Valcheva D., "The role of Computer Assisted Assessment System as an instrument for assessing and improving the students` knowledge and competence", Conference ICT, (CD), Vilach - Austria, 2009
[6] Nikolova M., Todorova M. „The essence and the structure of the online questionnaires in the education", International Conference on e-Learning and the Knowledge Society - e-Learning'10, (p.166-170), Riga, 08.2010
[7] http://en.wikipedia.org/wiki/E-assessment
[8] http://www.ocr.org.uk/eassessment/etesting/index.html
[9] http://www.wdvl.com/Authoring/CGI/

Indirect Identification of the Disturbances by Programmable Logic Controller Simatic S7-200

Vasil Dimitrov[1]

***Abstract* – The programmable logic controller (PLC) Simatic S7-200 offers maximum automation at minimum cost. It can be used for simple controls as well as for complex automation tasks. It finds application in many branches of industry, power engineering and transport. These advantages cause to teaching the students how to use and program this PLC.**

***Keywords* – indirect identification, - PLC, education quality.**

I. INTRODUCTION

The development of the energy and transport equipment entered to a new stage during the past years. Contemporary electrical transport vehicles are designed on the base of power drives controlled on high efficient devices and microprocessor safety and control systems. The new technologies used in electrical equipment make headway at accelerated rates. Regenerative converters (based on IGBT technology) and energy-saving motors are introduced into many modern trams, trolleybuses and locomotives.

On the other hand, in many branches of industry the positioning systems realize very often motions determining production quality and efficiency. The contemporary positioning systems require efficient solutions of the problem of the indirect identification of the disturbances in case of fast running processes.

Therefore, the efficiency of the traction drives and the accuracy of the positioning systems are directly connected to the possibility of control devices of recognition the peculiarities of the dominant disturbances.

II. APPLICABILITY OF THE PLC SIMATIC S7-200 TO THE MOTION DRIVES AUTOMATION AND TO THE POWER ENGINEERING

The SIMATIC Modular Controllers have been optimized for control tasks and specially designed for ruggedness and long-term availability [1]. They can be flexibly expanded at any time using plug-in I/O modules, function modules, and communications modules. The modular controllers can also be used as fault-tolerant or fail-safe systems.

The S7-200 series of programmable logic controllers can control a wide variety of devices to support automation needs. These PLCs are compact and highly powerful (e.g. in relation to its real-time response), they are fast, feature great communications capabilities and they are based on very user-friendly software and hardware. The SIMATIC S7-200 family is suitable for applications where programmable controllers would not have been economically viable in the past.

Siemens provides different S7-200 models with graduated range of CPUs with many basic PLC functions; with a diversity of features and capabilities that help to create effective solutions for varied applications [2].

S7-200 has been already introduced in the transport. It is used in the object level in SCADA system (Supervisory Control And Data Acquisition system), which controls the electrical equipment in the traction substations and Electricity transmission network. PLCs are used in many contemporary electrical transport vehicles (locomotives, under ground wagons etc).

S7-200 can be used in automation configurations based on Automation and Drives standard products for easy, fast and cost-saving implementation of automation tasks for small-scale automation. It is particularly suitable for industrial applications requiring the positioning of objects [5]. The compatible with S7-200 product combination in conjunction with the software library enables a cost-effective positioning solution in the following applications: cutters, for example, for pipes; conveyors; feeders; lifting stages; rotary tables; hoisting devices.

III. LABORATORY SIMULATOR FOR CONTROL ON ELECTRICAL DRIVES WITH A PLC S7-200

The fast industrial progress set up higher requirements of education quality. The power engineering and the transport are attractive areas for Bulgarian and foreign investment for progress, modernization and expert education. The training under the bachelor and master programmes Power Engineering and Electrical Equipment in The Todor Kableshkov University of Transport prepares highly qualified experts in the fields of the Electrical Equipment of the transport, industry and power engineering. The Todor Kableshkov University of Transport's lecturers realize the necessity of training of skilled workers for electrical transport and power engineering needs. There is a modern laboratory simulator built on contemporary devices (Fig.1). It includes an energy-saving induction motor (AD), controlled by a frequency converter Sinamics G120. A synchronous generator (SG) and resistors realize the load of the motor. There is a positioning system, too. A PLC Simatic S7-200 supplements the simulator and can be used for control and optimization both of the inverter drive and the positioning system. An HTL encoder reports the traveled distance of the positioning axis to

[1]Vasil Dimitrov is with Todor Kableshkov University of Transport, 158 Geo Milev Str., Sofia, Bulgaria

E-mail: vdimitroff@abv.bg

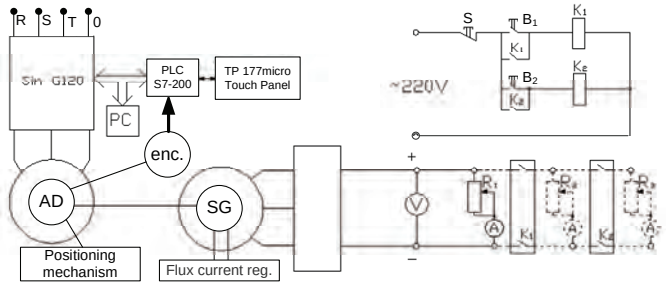


Fig. 1. Laboratory simulator

the S7-200 in the form of pulses. This encoder can also be used for closed loop motion control on the drive. TP 177micro Touch panel, connected to the PLC, gives the possibility of visualization and control of the drive system, as well as of entering the initialization and parameterization data and work conditions. A WINDOWS-based computer system is used for configuring and data archiving. The necessary configuration software is installed on this computer: Step 7 MicroWin for programming the PLC Simatic S7-200, WinCC Flexible Micro for configuring the Touch panel and STARTER for setting-up the frequency converter Sinamics G120.

The SINAMICS G120 frequency converters family is designed to provide precise and cost-effective speed and torque control of AC motors [3]. They are modular frequency inverters for standard drives. SINAMICS G120 is ideally suited as a universal drive in various industrial applications, e.g. in the automotive, textile, printing and chemical industries, for higher-level applications, in conveyor systems. Each SINAMICS G120 comprises two operative units – the Power Module and Control Unit. The Power Module is controlled by a microprocessor in the Control Unit. State-of-the-art IGBT technology with pulse-width modulation is used to achieve the highest degree of reliability and flexible motor operation. Comprehensive protection functions provide a high degree of protection for the Power Module and the motor. The Control Unit controls and monitors the Power Module and the connected motor using several different control types that can be selected. It supports communication with a local or central control and monitoring devices.

The simulator offers various possibilities of implementation into practice many laboratory exercises, for example:

- Examination of a synchronous generator at a constant speed and variable load;
- Examination of an open loop motion control in case of inconstant load torque. A sensorless Vector control (SLVC) without speed feedback or a V/f characteristic control can be examined. The V/f characteristic can be linear, flux current control (FCC), square-law characteristic (f^2 characteristic) or programmable characteristic, which takes into consideration the torque characteristic of the driven load [4];
- Examination of a closed loop speed control in case of inconstant load torque. There are several versions. In one case the closed-loop speed control with the evaluation of the encoder signal using a PID controller and V/f characteristic control can be examined. In other case the closed-loop vector control with speed encoder can be examined (this is the preferred solution in the most applications).
- Examination of a positioning system in case of inconstant load torque. The speed influence over the positioning accuracy can be examined;
- Optimization of the travel profile by the PLC Simatic S7-200. The encoder measures the traveled distance of the positioning axis and the motion speed can be calculated to achieve the wanted accuracy at highest efficiency.

The travel profile is essentially defined by the technological application and an energetic optimization. Depending on the requirements of the particular application various optimization methods can be used. Determining factors are the time within which the positioning operation has to be completed, the acceleration and deceleration whose upper limits are determined by the motor performance, the application's mechanical system and the actual process and the power demand of the application.

In all examinations the load torque can be changed by various techniques:

- It can be changed lightly altering the resistance of the load resistors or the generator flux current;
- It can be changed with a jerk switching the circuit closers K_1 and K_2 by pushing buttons B_1 and B_2.

IV. Indirect Identification of the Disturbances

The disturbances influence materially on the accuracy in the positioning systems. The dominant disturbance in the hoisting devices is the moment of inertia closely connected with the load mass. In the industry, the worse work conditions (like the moisture, dustiness, vibrations etc.) result in the load torque variation and decrease of the positioning accuracy.

In the electrical transport, the main disturbances are the supply pressure fluctuations, climatic conditions, load mass, road profile and track condition. They result in the vehicles speed variation.

The identification of the disturbances will optimize the efficiency and accuracy in the positioning systems as well as the passenger comfort and the energy consumption in the electrical transport. The direct identification will be very difficult depending on the various natures of the disturbances. The indirect identification takes into consideration the change of controlled parameters as a result of a disturbance. Sensors mounted on the system and connected to the PLC can easily measure these parameters. The S7-200 can be connected to the specific for positioning systems sensors (linear and rotary pulse encoders, voltage and current sensors), as well as with the speed, voltage and current sensors and energy consumption meters used in the power drive systems. After the calculating of the controlled parameters (for example, consumed active power, motor speed and current, power factor etc.) the disturbances influence can be eliminated and optimal drive behaviour can be achieved. Programmed PID controller can be used to obtaining a fast response of a disturbance appearance. In steady state operation, a PID controller regulates the value of the output so as to drive the error (e) to zero. A measure of the error is given by the difference between the setpoint (SP) (the desired operating point) and the process variable (PV) (the actual operating

point). The principle of PID control is based upon the following equation that expresses the output, M(t), as a function of a proportional term, an integral term, and a differential term:

$$M(t) = K_c.e + K_c.\int_0^t e.dt + M_{in.} + K_c.\frac{\partial e}{\partial t} \quad (1)$$

where:
M(t) is the loop output as a function of time;
K_C is the loop gain;
e is the loop error (the difference between setpoint and process variable);
$M_{in.}$ is the initial value of the loop output.

In order to implement this control function in a digital description, the continuous function must be quantized into periodic samples of the error value with subsequent calculation of the output. The corresponding equation that is the basis for the digital algorithm solution is:

$$M_n = K_c * e_n + K_I * \sum_{x=1}^{n} e_x + M_{in} + K_D * (e_n - e_{n-1}) \quad (2)$$

where:
M_n is the calculated value of the loop output at sample time n;
e_n is the value of the loop error at sample time n;
e_{n-1} is the previous value of the loop error (at sample time n-1)
e_x is the value of the loop error at sample time x;
K_I is the proportional constant of the integral term;
K_D is the proportional constant of the differential term.

The integral term is shown to be a function of all the error terms from the first sample to the current sample. The differential term is a function of the current sample and the previous sample, while the proportional term is only a function of the current sample. In a digital processor, it is not practical to store all samples of the error term, nor is it necessary. Since the processor must calculate the output value each time the error is sampled beginning with the first sample, it is only necessary to store the previous value of the error and the previous value of the integral term. As a result of the repetitive nature of the digital algorithm solution, a simplification in the Eq. (2), which must be solved at any sample time, can be made. The simplified equation is:

$$M_n = K_c * e_n + K_I * e_n + MX + K_D * (e_n - e_{n-1}) \quad (3)$$

where:
MX is the value of the integral term at sample time n-1 (also called the integral sum or the bias):

$$MX = K_I * \sum_{1}^{n-1} e_x + M_{in.} \quad (4)$$

The S7-200 uses a modified form of the Eq. (3) when calculates the loop output value. This modified equation is:

$$M_n = MP_n + MI_n + MD_n \quad (5)$$

where:
MP_n, MI_n *and* MD_n are the values of the proportional, integral and differential terms of the loop output at sample time n [2].

The proportional term MP is the product of the gain (K_C), which controls the sensitivity of the output calculation, and the error (e), which is the difference between the setpoint (SP) and the process variable (PV) at a given sample time. The equation for the proportional term as solved by the S7-200 is:

$$MP_n = K_c * (SP_n - PV_n) \quad (6)$$

where:
SP_n is the value of the setpoint at sample time n;
PV_n is the value of the process variable at sample time n.

The integral term MI is proportional to the sum of the error over time. The equation for the integral term as solved by the S7-200 is:

$$MI_n = K_c * \frac{T_s}{T_I} * (SP_n - PV_n) + MX \quad (7)$$

where:
T_S is the loop sample time;
T_I is the integration period of the loop (also called the integral time or reset).

The integral sum or bias (*MX*) is the running sum of all previous values of the integral term. After each calculation of MI_n, the bias is updated with the value of MI_n, which might be adjusted or clamped. The initial value of the bias is typically set to the output value ($M_{in.}$) just prior to the first loop output calculation. Several constants are also part of the integral term, the gain (K_C), the sample time (T_S), which is the cycle time at which the PID loop recalculates the output value, and the integral time or reset (T_I), which is a time used to control the influence of the integral term in the output calculation.

The differential term MD is proportional to the change in the error. The S7-200 uses the following equation for the differential term:

$$MD_n = K_c * \frac{T_D}{T_S}\left[(SP_n - PV_n) - (SP_{n-1} - PV_{n-1})\right] \quad (8)$$

where:
T_D is the differentiation period of the loop (also called the derivative time or rate);
SP_{n-1} is the value of the setpoint at sample time n-1;
PV_{n-1} is the value of the process variable at sample time n-1.

To avoid step changes or bumps in the output due to derivative action on setpoint changes, the Eq. (8) is modified to assume that the setpoint is a constant ($SP_n = SP_{n-1}$). This results in the calculation of the change in the process variable instead of the change in the error as shown:

$$MD_n = K_c * \frac{T_D}{T_S}(PV_{n-1} - PV_n) \quad (9)$$

The process variable rather than the error must be saved for use in the next calculation of the differential term. At the time of the first sample, the value of PV_{n-1} is initialized to be equal to PV_n.

In many control systems, it might be necessary to employ only one or two methods of loop control. Setting the value of the constant parameters makes the selection of the type of loop control desired. If the integral action is not required, then a value of infinity "INF" should be specified for the integral time T_I. Even with no integral action, the value of the integral term might not be zero, due to the initial value of the integral sum MX. If the derivative action is not required, then a value of 0.0 should be specified for the derivative time T_D. If the proportional action is not necessary (I or ID control is required), then a value of 0.0 should be specified for the gain K_C. Since the loop gain is a factor in the equations for calculating the integral and differential terms, setting a value of 0.0 for the loop gain will result in a value of 1.0 being used for the loop gain in the calculation of the integral and differential terms.

Eight PID instructions can be used in a program, which controls S7-200. Each loop has four constant parameters, two input variables (the setpoint and the process variable) and output value, which is generated by the PID calculation. The setpoint is generally a fixed value. The process variable is a value that is related to loop output and therefore measures the effect that the loop output has on the controlled system. A loop table stores nine parameters used for controlling and monitoring the loop operation and includes the current and previous value of the process variable PV_n and PV_{n-1}, the setpoint SP, output M_n, gain K_C, sample time T_S, integral time T_I, derivative time T_D, and the integral sum MX (bias). The output value field in the loop table is updated at the completion of each PID calculation.

Therefore, it is possible to realize a multidimensional system that includes up to eight closed-loop control processes. Such is the case, for example, that positioning and speed regulators are used and a current (toque) governor is in a state of subordination of the speed controller. In this case, the setpoint of the low range and the setpoint of the high range should correspond to the process variable low range and high range.

In the laboratory simulator, the load torque is depending on the output power and the speed of the synchronous generator. The output power is the product of the output current and voltage of the generator. Therefore, the load torque might be calculated and measuring the current, voltage and the speed of the generator, might identify disturbances appearance.

The S7-200 must calculate the load torque. The outputs of voltage and current sensors mounted on the output circuit of the synchronous generator have to be connected to the analog inputs of the S7-200 [6]. By multiplication of these two signals the output power of the synchronous generator must be calculated. The rotary encoder mounted in the induction motor might be used for speed measuring. The load torque can be calculated dividing the output power and the generator speed:

$$M_c = \frac{30 * K_U * U * K_I * I}{\pi * n} \tag{10}$$

where: M_c is the load torque;
U and I are measured voltage and current values;
K_U and K_I are the sensors constants;
n is the measured value of the speed.

The Eq. (10) might be used for a process variable in a PID controller. In this way, the drive system behaviour will be optimized.

When the positioning system is examined, the rotary encoder must be used for the measuring of the traveled distance. If a linear encoder will be mounted on the positioning mechanism, it could be used for precise position measuring. The friction error and the clearance in the gear will be eliminated. A high accuracy can be achieved, because the standard resolution of the traveled distance amounts to 0.01mm. Three output signals are available to be connected to the PLC's digital inputs: the channels A and B are 90° phase shifted and the index pulse is a periodic pulse, which is released each 5 mm's. A high-speed counter might be used for calculating the traveled distance and the absolute position of the mechanism. These values can be used for the travel profile optimization.

V. CONCLUSION

In this paper the examinations of the power drive and positioning systems are given. The laboratory simulator gives possibilities of student's practical training in many terms of reference, for example:

- Training the programming skills for setting-up the controllers, inverters and touch screens – a knowledge of the relevant software products is necessary;
- Synthesis of algorithms for optimal control on the positioning and drive systems, using PID controllers and selecting the values of their parameters (gain K_C, sample time T_S, integral time T_I, derivative time T_D);
- Using the vector control on the power drive systems - the Sinamics G120 has a current measurement function, which permits the output current to be precisely determined, referred to the motor voltage. This measurement guarantees the output current to be sub-divided into a load component and a flux component. Using this subdivision, the motor flux can be controlled and can be appropriately adapted and optimized inline with the prevailing conditions;
- Estimation of the positioning error at the random load torque disturbances as well as at the various speed;
- Evaluation of the PLC's possibilities for its using in the positioning and power drive systems for the efficiency optimization.

REFERENCES

[1] Products for Totally Integrated Automation and Micro Automation, Siemens, Catalog ST 70, 2009.
[2] SIMATIC S7-200 - Programmable Controller, System Manual, Siemens, 2008.
[3] SINAMICS G120 Standard Inverters, Siemens, Catalog D 11.1, 2009.
[4] SINAMICS G120 - Function Manual, Siemens, 2007.
[5] Closed-Loop Positioning Control with standard Drives, Siemens, 2008.
[6] Cherneva G. Computerized Laboratory in Science and Technology Teaching: Course Stress Resonance. Workshop on Multimedia in Physics Teaching and Learning, Prague, 2003

Teaching Interactive Cryptography: the Case for CrypTool

Saša Adamović[1], Irina Branović[2], Dejan Živković[3], Violeta Tomašević[4], Milan Milosavljević[5]

***Abstract* – The theory and applications of cryptography are complicated and hard to follow for undergraduate students with less mathematical background. For this reason, instead of plain theoretical teaching, we applied different, interactive approach. Open-source CrypTool software allowed us to practically demonstrate all current private and public-key algorithms and protocols. This paper describes our teaching model and experience. Positive feedback received from students confirms the advantages of adopted approach with respect to traditional teaching.**

***Keywords* – education, cryptography, interactive teaching, CrypTool**

I. Introduction

Cryptography course as part of computer science curriculum has become a necessity, considering that it nowadays serves as the basis for data communication security, information and network security. For today's computer professionals, secure data storage and communication have become vital competencies, unlike before when cryptography was considered to be a secret science. However, since it is directly based on diverse mathematical disciplines (number theory, abstract algebra, probability), students with less mathematical background are often intimidated and could benefit from teaching through practical examples.

Analyzing the results achieved and students' feedback, we noticed that the plain textbook-theoretical approach to teaching cryptography that we used to apply simply was not satisfactory, and subsequently decided to shift to interactive approach by introducing the open-source cryptography software CrypTool [4]. The course makeover required substantial instructor and teaching assistant efforts, especially when choosing the right examples to illustrate the most commonly used cryptography algorithms and protocols. In this paper we describe the teaching experiences and analyze students' results which confirm that teaching cryptography interactive, through practical demonstrations, is indeed advantageous.

[1]Saša Adamović is with the Singidunum University, Danijelova 32, 11000 Beograd, Serbia, E-mail: sadamovic@singidunum.ac.rs

[2]Irina Branović is with the Singidunum University, Danijelova 32, 11000 Beograd, Serbia, E-mail: ibranovic@singidunum.ac.rs

[3]Dejan Živković is with the Singidunum University, Danijelova 32, 11000 Beograd, Serbia, E-mail: dzivkovic@singidunum.ac.rs

[4]Violeta Tomašević is with the Singidunum University, Danijelova 32, 11000 Beograd, Serbia, E-mail: vitomasevic@singidunum.ac.rs

[5]Milan Milosavljević is with the Singidunum University, Danijelova 32, 11000 Beograd, Serbia, E-mail: mmilosavljevic@singidunum.ac.rs

II. Related Work

It is well known that cryptography is a hard-to-master discipline, which requires strong mathematical background because the security of a cryptosystem is often based on the inability to efficiently solve a problem in algebra, number theory, or combinatorics. Many instructors have made attempts to adapt their teaching methods to be flexible and to get students interested in the topic; one such approach is described in [1]. In [2], authors analyze 20 selected academic courses in cryptology with respect to their aims, scopes, contents, organization, and literature recommended to students, finally proposing the curricula tailored for different categories of students. In closely related [3], authors propose a "theory-algorithm-practice-application" teaching mode, which has proved to be efficient in achieving better teaching results, and helping students solve practical problems encountered in the engineering practice by using cryptography.

III. Interactive Cryptography

Our cryptography course curriculum mainly follows [5] and focuses on cryptographic principles, procedures, mechanisms, and techniques required for secure communications. Fig. 1 shows the building blocks of the course.

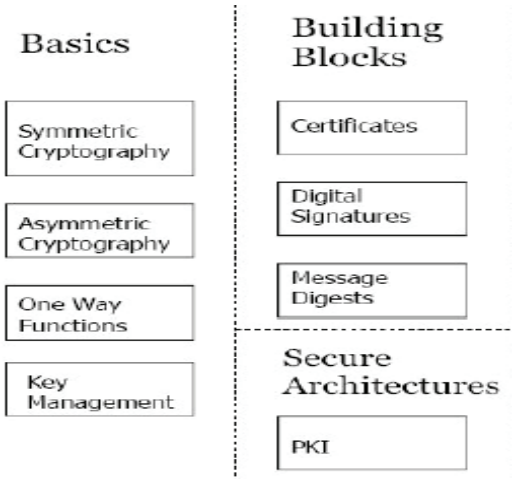


Fig. 1. Building blocks of the course

When implementing the course curriculum for the 2010 school year, our primary goal was to put emphasis on understanding the basics of information security, protection of cryptographic algorithms, and security services (authentication, authorization, confidentiality, non-repudiation and availability), as well as to provide practical examples

which integrate theory with practice. For this purpose we have chosen CrypTool, a free, open-source learning application, used worldwide in the implementation and analysis of cryptographic algorithms. From its graphical interface, CrypTool offers numerous interactive demonstrations and visualizations of classic and modern cryptographic algorithms, generation of the secure passwords, authentication, cryptanalysis, and encryption.

Fig. 2 shows an RSA key generation/encryption simulation based on the original algorithms approved by NIST. The simulation was developed by students during laboratory exercises.

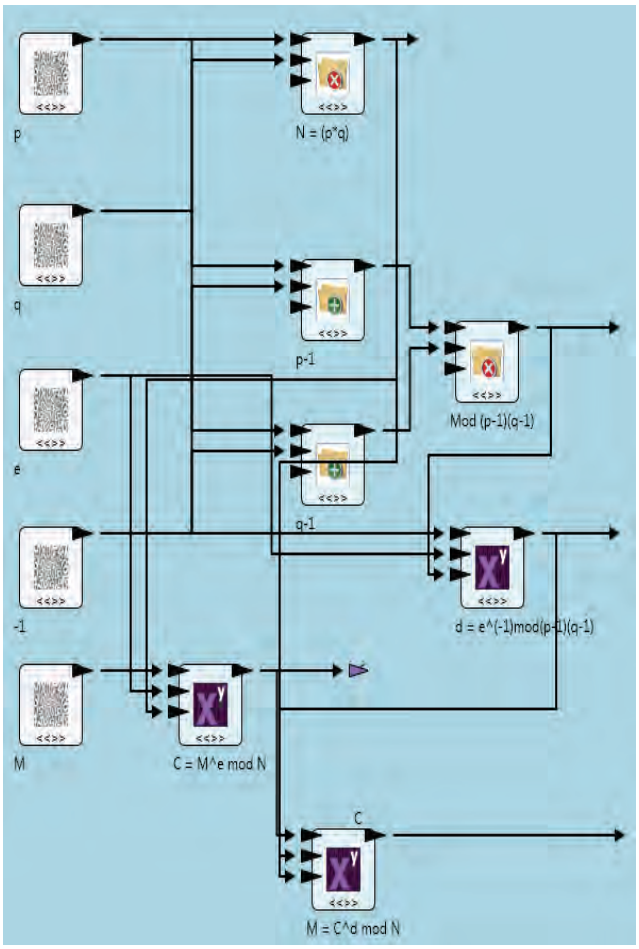


Fig. 2. An example of interactive CrypTool screen illustrating RSA key generation and encryption.

Before using interactive laboratory for complex mathematical operations, students with poor mathematical background were not able to view the simulation in a controlled interactive way. Namely, without CrypTool teacher could only show students the functionality of several commercial programs related to the basic cryptography principles that need to be respected in the construction of a cryptographic solution. This approach has a negative impact on the interest of students for the course and their final success.

Using CrypTool we have got the lab that works in real time with real parameters. Students have been now able to follow every cryptographic system step by step. More important, students can easily and quickly implement their ideas by dragging objects from the palette that contains algorithms and run the simulation in real time. This way students over time could get more experimental experience that not only creates more interest among them for the cryptography overall, but is also positively reflected on their final success.

IV. Analysis of Teaching Results

For the sake of comparison, Fig. 3 shows the student's attendance and grades for the 2008 and 2009 school year in the cryptography course without CrypTool, as well as for the 2010 school year when we started using interactive approach in the course.

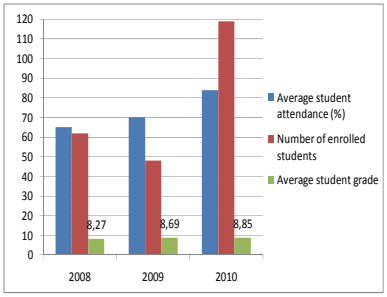


Fig. 3. Students' attendance and grades before and after using interactive approach in the course.

As the diagram clearly demonstrates, student attendance jumped well over 10%, and students grade also slightly improved. With some new ideas and building on previous experience, in the years that follow we hope to further increase students' involvement and learning outcome.

V. Conclusion

In order to combine theory and practice in teaching cryptography, this paper advocates using CrypTool as a powerful instructional tool. Interactive CrypTool software makes up for the students' lack of strong mathematical background and demonstrates inner workings of cryptography algorithms in a user-friendly way. The effectiveness of CrypTool is confirmed by a comparative analysis of students' attendance and grades, which clearly justifies its use in a cryptography course.

Acknowledgement

This work has been supported by the Serbian Ministry of Education and Science, projects no. 44006 and 32054.

REFERENCES

[1] X. Song, H. Deng, “Taking Flexible and Diverse Approaches to Get Undergraduate Students Interested in Cryptography Course”, Proceedings of the First International Workshop on Education Technology and Computer Science, 2009.

[2] Olejar, D., and Stanek, M., "Some Aspects of Cryptology Teaching”, Proceedings of the First World Conference on Information Security Education WISE1, 1999.

[3] Y. F. Zhong, C. Y. MengXiao, H. YiRan, “Teaching Cryptology Course Based on Theory-Algorithm-Practice-Application Mode”, Proceedings of the First International Workshop on Education Technology and Computer Science, 2009.

[4] CrypTool, available online at http://www.cryptool.com/

[5] W. Stallings, *Cryptography and Network Security*, Fourth Edition, Prentice-Hall.

The Appliance of OLAP and Microsoft SQL Server Analysis Services in the Analysis of User Behavior Patterns

Marija Blagojevic[1] and Sava Baric[2]

***Abstract* – This paper describes the use of OLAP and Microsoft SQL Server Analysis Services in the analysis of behaviour patterns. The research was conducted at the Technical Faculty in Cacak. The aim of research is the adaptation of traditional teaching and the concept of e-learning by the students patterns of behaviour. The future work relates to the obtainment of detail information about students' behaviour in the learning management system.**

***Keywords* – Olap, Microsoft SQL Server Analysis Services, e-courses, behaviour patterns**

I. INTRODUCTION

Analytical solutions are quickly becoming mission critical for many organizations. Recently in educational environments, e-learning solutions have been growing and developed rapidly. A critical problem facing e-learning is the lack of relevant and timely information. As information cost money, it must adopt innovative approaches to attain operational efficiently [1]. Nowadays, most universities, colleges and high school are vitalizing their teaching through e-learning platforms benefiting from their many advantages.

E-learning can be used on all levels of education and are used just as well in the combination with the traditional teaching as in the distance learning. That is why the knowledge about the users of electronic courses is essential for understanding their ways of learning and their learning approach. In order to obtain this, it is of essential importance to discover and analyze their patterns of behavior.

Analysis of users' behavior patterns can be used for new model designing that can be of high importance for understanding of users' behavior in virtual environment. E-learning systems accessible through the Internet are intranets that represent self contained versions of the data warehouses and human behavior found more broadly across the Internet [2]. In this study were used OLAP for extracting useful information and evaluating of user profiles. In addition here are used Microsoft SQL Server Analysis Services.

In [3] OLAP was applied to get information for learners observing interactions with the system. According to paper [4] data mining and warehousing are the two most important techniques for pattern discovery and centralized data management in today's technology. In [4] is defined e-learning model while stressing the importance of data mining in e-learning. In [5] is shown how to create OLAP cube and the operation with the cube. A concrete example of the cubes in this paper was applied to data relating to the online learner.

[1]Marija Blagojevic is with the Technical Faculty Cacak, Svetog Save 65, 32000 Cacak, Serbia

E-mail: marija_b@tfc.kg.ac.rs.

[2]Sava Baric is with the Technical Faculty Zrenjanin, Djure Djakovica, bb, 23 000 Zrenjanin, Serbia

E-mail: savaba@gmail.com

II. OLAP AND MICROSOFT SQL SERVER ANALYSIS SERVICES

Microsoft Analysis Services are a collection of Online Analytical Processing (OLAP) and data mining services supplied in Microsoft SQL server. Analysis services provide managers the possibility to explore a cache of collected and current data, define business trends and patterns and mine data to make discerning business decisions. Clients communicate with Analysis services using the public standard XML for analysis (XMLA), a SOAP-based protocol for issuing commands and receiving responses, exposed as a Web service. Data mining presents analysis of observational data sets with the purpose for detection of undetected links and data summing in a sophisticated manner, understandable and useful for data owner. The relations and summings that are obtained by the data mining process are defined as models or patterns. Microsoft Analysis Services is part of Microsoft SQL Server, a database management system. Microsoft has included a number of services in SQL Server related to business intelligence and data warehousing. These services include Integration Services and Analysis Services. Analysis Services includes a group of OLAP and data mining capabilities. Analysis Services provides managers the possibility to explore a cache of collected and current data, define business trends and patterns and mine data to make discerning business decisions. Microsoft SQL Analysis Services (MSAS) relies on Windows accounts for granting access to cube data as well as for administrative tasks such as processing cubes, altering server-wide configuration settings and modifying dimensional objects.

A dimension is the major analytical object. Dimensions have attributes, and they have relationships with facts. Their reason for being is to add qualitative information to the numeric information contained in the facts. Figure 1 shows Analysis services architecture.

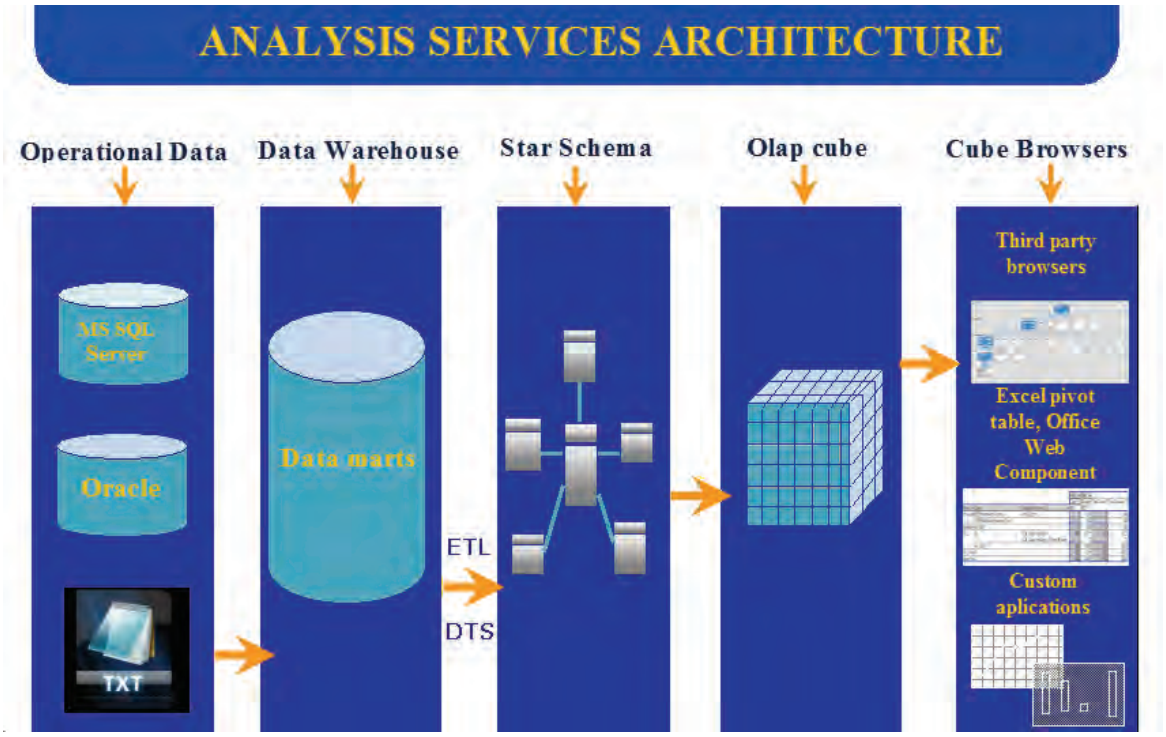


Fig. 1. Analysis services architecture [6]

III. Purpose of the Study

A learning management system (LMS) is a software application for the administration, documentation, tracking, and reporting of training programs, classroom and online events, e-learning programs, and training content [7]. However, LMS doesn't allow detail monitoring of the users' activities. In order to consider the complete teaching process that includes the usage of electronic courses within a specific LMS, a thorough analysis is a must. OLAP and Microsoft SQL Server Analysis Services are used in this study in the purposes of analysis of users' patterns of behaviors.

The objectives of the research:

- Data pre-processing: clean and prepare the Web server log file
- OLAP analysis: design a multidimensional structure in which the main factors under analysis:(year, month, day, time, minute, course, and modul activity) will be taken as dimensions and later build OLAP cube in order to analyze the recorded data.
- Pattern evaluation: determination of behaviour patterns based on obtained reports and their evaluation

The goal of the research:

- Professors will have an insight in students' patterns of behaviour and according to them thy will have a chance to organise their classes by using other concepts such as concept of e-learning.

The goal of research: Professors will have an insight in students' patterns of behavior and according to them they will have a chance to adapt their classes

Hypothesis: The activity of students in electronic courses on different modules varies during different days of the week

IV. Methods, Participants, Tools and Procedures

In order to detect and analyze behavior patterns, OLAP (On Line Analytical Processing) have been used in this study. In addition to OLAP were used Microsoft SQL Server Analysis Services.

The study was conducted at the Technical faculty in Cacak. In this case, traditional method of teaching is being combined with e-learning with the assistance of Moodle LMS (http://itlab.tfc.kg.ac.rs/moodle). This Moodle system and e-courses are designed to provide teaching material to students, and activities that provide collaborative learning.

System registered 1789 active users. One hundred of courses have been created within the system, and these courses are being used by students and/or teachers as well.

For purposes of this research are used Microsoft Visual Studio 2008 and Microsoft SQL Server 2008.

The procedure suppose that is used data pre-processing (clean and prepare the Web server log file), and OLAP analysis (design a multidimensional structure, and later build OLAP cubes in order to analyze the recorded data). After that, in

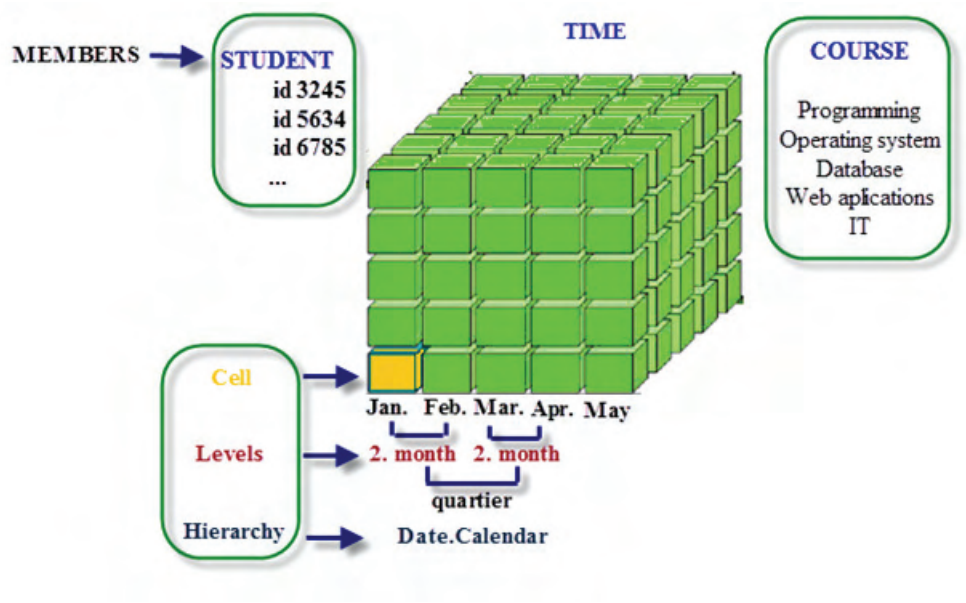


Fig. 2 Olap cube

Microsoft SQL Server Management Studio is prepared reports. To determine the activity of students during the different days of the week, all data were randomly selected for two weeks. After that are created the dimensions and cube.

Creating dimensions and cubes leads to the final phase of work in Microsoft Visual Studio 2008. Figure 2 shows the Olap cube. This phase involves deployment solutions. Deployment solution passes several phases, and the result is to be successfully completed deployment to continue the work within SQL Server Management Studio.

V. Results and Discussion

In this section are given the result of the research. Visualization of results is done in Microsoft Excel 2003.

Figure 2 shows the activities of the modules during the day of the week. Numbers from 1 to 7 indicated seven days a week and this Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday, respectively. On the x-axis is shown the actual number of activities or access logs of a given module. According to this figure, it can be concluded that students, for example, this week, had a different activity in various modules in the course.

Figure 3 shows the activities of the modules during the day of the second selected week. Numbers from 1 to 7 indicated seven days a week. On the x-axis is shown the actual number of activities or access logs of a given module. According to this figure, it can be concluded that students, for example, this week, had a different activity in various modules in the course.

Both of these figures confirm initial hypothesis. The activity of students in electronic courses on different modules varies during different days of the week. Results indicate the need for adaptation activities in the courses depending on the day of the week.

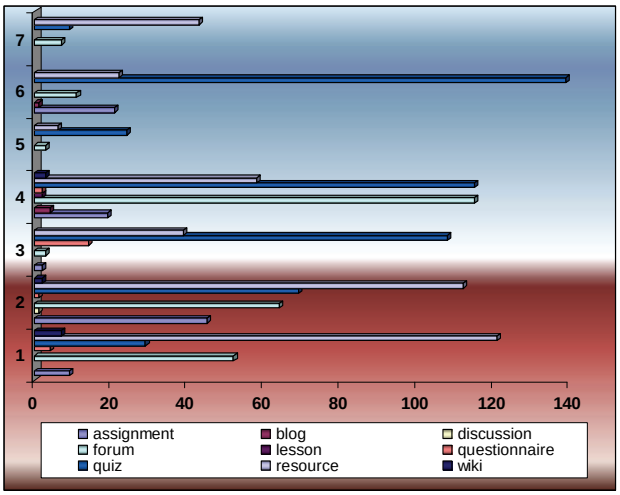


Fig. 2. Actions by modules in the day of the week

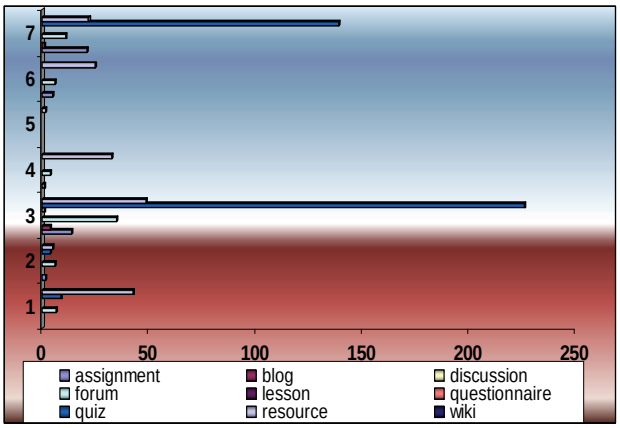


Fig. 3. Actions by modules in the day of the week

VI. Conclusion

There are number of techniques for evaluation of user behaviour patterns. Techniques used in this paper supported initial hypothesis. .

Similar in [8] here is made the learner distribution over time, but with one difference. In [8] is given distribution over time of day.

Benefits of using OLAP and Microsoft SQL server analysis services is the accurate and efficient processing of a large number of records in log files and highly accurate evaluation features user behaviour patterns.

The future work relates to the obtainment of detail information about students' behaviour in the learning management system.

References

[1] J.B. Donald, W.F. John, R.H. Alan, S. James, "Healthcare Data Warehousing and Quality Assurance", IEEE Computer, pp.56-65, 2001, December.

[2] J. Ai, J. Laffey,"Web Mining as a Tool for Understanding Online Learning", from: http://jolt.merlot.org/vol3no2/ai.htm, (2007). Retrieved November 23, 2010.

[3] C. Romero, S. Ventura, "Data mining in E-Learning" Volume 4 of Advances in Management Information. WIT Press. 2006.

[4] O. Zaïane, "Web Usage Mining for a Better Web-Based Learning Environment" Proc. of Conference on Advantage Technology for Education. Alberta, Canada. 2001.

[5] M. Zorrilla, "New Trends in Database Systems, Methods, Tools, Applications", retrieved from: http://personales.unican.es/zorrillm/BDAvanzadas/Teoria/Zorrilla07-DW-OLAP.pdf

[6] http://www.devx.com/dbzone/Article/21410/1763?supportItem=1

[7] R. Ellis, "A Field Guide to Learning Management Systems, ASTD Learning Circuits", 2009, from http://www.astd.org/NR/rdonlyres/12ECDB99-3B91-403E-9B15-7E597444645D/23395/LMS_fieldguide_20091.pdf, Retrieved on April 24, 2010

[8] Goyal, N, On-Line Analytical Processing, (2007) from http://csis.bitspilani.ac.in/faculty/goel/Data%20Warehousing/II%20sem%202006-2007/Lecture%20Slides/OLAP-2.ppt, Retrieved on April 24, 2010

AUTHOR INDEX