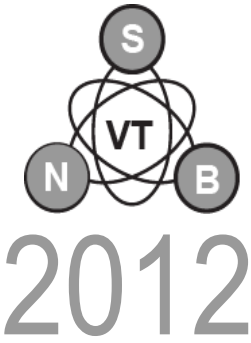
XLVII INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION, COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES



2012 ICEST

PROCEEDINGS OF PAPERS

VOLUME 2

SOFIA, 2012

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

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

ISBN: 978-619-167-003-1

T A B L E O F C O N T E N T S

VOLUME 1

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS – PART 1

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS – PART 2

TELECOMMUNICATION SYSTEMS AND TECHNOLOGY – PART 1

TELECOMMUNICATION SYSTEMS AND TECHNOLOGY – PART 2

SIGNAL PROCESSING

DIGITAL IMAGE PROCESSING

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES

INFORMATICS AND COMPUTER SCIENCE – PART 1

INFORMATICS AND COMPUTER SCIENCE – PART 2

ELECTRONICS

ENERGY SYSTEMS AND EFFICIENCY

CONTROL SYSTEMS

MEASUREMENT SCIENCE AND TECHNOLOGY

REMOTE ECOLOGICAL MONITORING

ENGINEERING EDUCATION

VOLUME 2

RADIO COMMUNICATIONS, MICROWAVES, ANTENNAS (POSTER)

TELECOMMUNICATION SYSTEMS AND TECHNOLOGY (POSTER)

SIGNAL PROCESSING (POSTER)

DIGITAL IMAGE PROCESSING (POSTER)

COMPUTER SYSTEMS AND INTERNET TECHNOLOGIES (POSTER)

INFORMATICS AND COMPUTER SCIENCE (POSTER)

ELECTRONICS (POSTER)

ENERGY SYSTEMS AND EFFICIENCY (POSTER)

CONTROL SYSTEMS (POSTER)

MEASUREMENT SCIENCE AND TECHNOLOGY (POSTER)

REMOTE ECOLOGICAL MONITORING (POSTER)

ENGINEERING EDUCATION (POSTER)

Sensor Network Topology as Low-interference Factor

Vasil Dimitrov[1], Rozalina Dimova[2] and Paskal Novakov[3]

***Abstract* – This paper presents investigation results of wireless sensor network topology connection with some main geometric structures. The aim is optimal communication organization between neighbour nodes. Sensor node range and distance between nodes as a factor for interference influence has calculated. MatLab simulation model has developed for interference investigation in sensor network design.**

***Keywords* – Wireless sensor, topology control, Interference, Relative Neighbourhood Graph, Gabriel graphs.**

I. INTRODUCTION

Wireless sensor network (WSN) consists many geographically dispersed units, that communicate wirelessly with each other [1]. Missing central infrastructure supposes that networks have no particular fixed topology. This way the control of the topology [2] appears to be necessary in planning the network structure in order to maintain connectivity, effective power and optimizing network performance. Many algorithms for topology control [3] were offered in the last decades, including the most famous Delaunay Graph (DG) [4], Relative Neighborhood Graph (RNG) [5], Gabriel Graph (GG) [6]. These algorithms are principally designed for energy efficiency and connectivity between neighboring nodes, but there is another important criterion for sensor networks such as productivity.

In this paper we consider geometric structures by which to analyze the wireless sensor nodes and their geographical distribution in the environment. There shall be accordance between the network topology and a geometric structure. When algorithms for topology control are used, based on the geometric structure, it's necessary to take into account the communication range of sensor nodes. The control of topology is used in sensor networks mostly to optimize the initial topology to save energy, reduce interference and prolong the life of the network.

The main aim is to reduce the number of active nodes, according to save resources necessary for future communications. The installation of wireless sensor networks in the area appear many problems mainly related to transmitter power, conservation of energy supply of batteries and maintaining communication connectivity. In case to save the packet, it's very important to choose the right node during the organization of the communication. A wrong choice may cause blocking of the communication.

The transmitting nodes affect the ability of the other nodes to receive data. The sensor node is not able to receive data from its neighbor, if in the same time another neighbor sends data. This mutually disturbance of the communication is named interference. Reducing the interference [7] in the network leads to much fewer conflicts during the sending of packets, which actually reduces the usage of energy and extends the life of the network. Therefore, reducing the interference via using the correct graph at the disposal is an important aim for the control of the topology.

II. GEOMETRIC STRUCTURES

The geometric structures graphics are graphics where vertices are points in a plane connected by straight segments between the individual points. This kind of graphics are known as proximity graphics [8].

A. Voronoi Diagram and Delaunay triangulation

In the diagram of Voronoi (dotted lines on fig. 1) set of sensor nodes placed in a plane form discrete units, which form the convex polygons (area Voronoi). All sides inside the polygon are most closely only to one node. Through this type of structure effectively create polygons with vertices located equidistant from neighboring node.

Delaunay triangulation signed as DT (the continuous line on fig. 1) is a double graph of Voronoi diagram. This is a unique set of points, where the peaks of all triangles lie on circles. And in these circles there are no other points. The special feature of this geometric structure is that it is well balances- the triangles do not differ significantly.

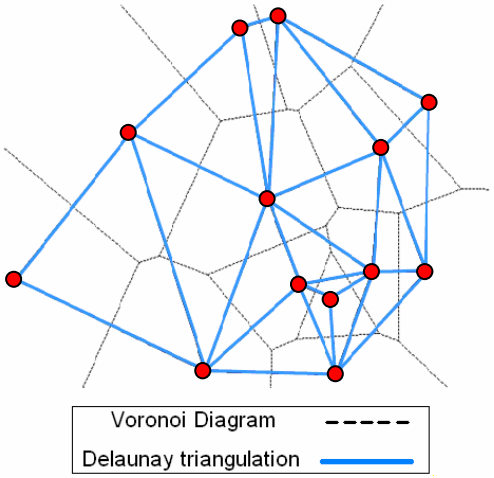


Fig. 1. Voronoi Diagram and Delaunary Triangulation

[1]V.Dimitrov is with the Technical University of Varna, ul. Studentska 1, 9010 Varna, Bulgaria (phone: +359-52-383350; e-mail: v_1986@abv.bg).

[2]R.Dimova is with the Technical University of Varna, ul. Studentska 1, 9010 Varna, Bulgaria; (phone: +359-52-383350, e-mail: rdim@abv.bg).

[3]P.Novakov is with the Technical University of Varna, ul. Studentska 1, 9010 Varna, Bulgaria (phone: +359-52-383350; e-mail: render3d@abv.bg).

B. Relative Neighborhood Graphs

Relative Neighborhood Graph (RNG) connects two points, *u* and *v* in the area *S*, then and only then when:

dist(u,v)≤max{dist(u,q), dist(v,u)}, for each point $q \in S$.

With *B(x,r)* assign the opened circle with radius *r*, and centre *x*, where *B(x,r)={y|dist(x,y)<r}*.

Figure 2. shows how to set up communication links with RNG.

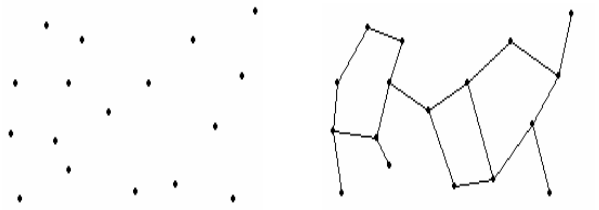


Fig. 2. Communication links in Relative Neighborhood Graph

D. Gabriel graph

Gabriel graph (GG) is under the graphic of Delaunay triangulation (DT) because of the ability called „empty circle" where in a circle with a diameter of a line joining the two nodes must be adjacent nodes (fig.3).

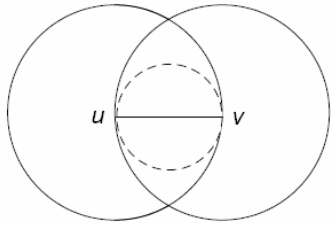


Fig. 3. Empty circle in Gabriel Graph

GG connects two points, *u* and *v* in the *S* area, then and only then when:

$$dist(u,v) \le \sqrt{dist^2(u,q) + dist^2(q,v)}\,, \qquad (1)$$

for each point $v \in S$.

Figure 4. shows how the communication connections in GG are formed.

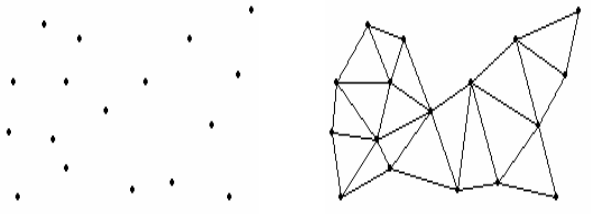


Fig. 4. Communication links in Gabriel Graph

III. FLUCTUATIONS OF ENERGY DEPENDING ON THE GEOMETRIC STRUCTURE

Saving energy is very important for the sensor networks, because the nodes are charged only vie batteries. To send the signal from one sensor node to another, the power that is used from these two nodes contains the following three parts [9]. First, the sensor node, which sends, needs some energy to prepare the signal. Second, in most cases, the power required for the realization of the communication link *uv* is $\|uv\|^{\beta}$, where $\|uv\|$ is Euclidean distance between *u* and *v* , and β is the real constant between 2 and 5 in transmission depending on the environment. Finally, when the node receives the signal needs some energy for the reception, storage and processing of this signal..

Let's take a look at each set $\pi(u,v)$ in the geometric structure from the node $u \in V$ to another node $v \in V$.

$\pi(u,v) = v_0 v_1 \ldots v_{h-1} v_h$, *where* $u=v_0$, $v=v_h$

Here the *h* appears to be the count of hops on the trail π. The general power for the emission $p(\pi)$, consumed by this path π is defined as [10] :

$$p(\pi) = \sum_{i=1}^{h} \| v_{i-1} v_i \|^{\beta} \qquad (2)$$

The trail connecting *u* and *v* with less usage of energy is known as least-energy path. Let *p(u,v)* be the less used energy of all paths, that connect *u* and *v*, and *H* is under graph of this geometric structure. Then the usage of energy in this graph may be defined as:

$$p(H) = \max_{(u,v \in V)} \frac{pH(u,v)}{p(u,v)} \qquad (3)$$

When we change some of the parameters, that are used for sending packets in the sensor network may be reached different results for the same topology.

This topology, where the interference is with very low percent appears to be optimized for the cases when the main criterion is the energy.

IV. NETWORK MODEL

In our analysis we consider the network topology from sensor nodes randomly distributed in a plane *G=(V,E)*, where $V=\{v_1,v_2,\ldots,v_n\}$ and *E* is the aggregation of connections between neighbouring sensor nodes in the network. We accept, that all the nodes transmit at maximum power. The range of a radio R_{max} is calculated as follows [11]:

$$R_{\max} = \frac{\lambda}{4\pi} \sqrt{\frac{P_t G_t G_r (1 - |G_r|^2)}{P_r}}\,, \qquad (4)$$

where:

P_t - power that the sensor sends;

G_t - gain of transmitting antenna;

G_r - gain of receiving antenna;

P_r - sensitivity of the receiver;

$|G_r|^2$ - ratio of reflected power from the receiving antenna;

Algorithms for topology control , which aim is to decrease the interference are besed on different parameters based on the topology of the network itself. All the nodes are potentially disturbing for a node. Before we determine we will introduce the following definition:

▪ The interference area: that part of the range of the node v_1, which overlaps with the ranging area of the node v_2 and form a negative impact on the communication of these two nodes with their other neighbors.

▪ Potentially disturbing node: a node v_1, is potentially disturbing on node v_2, if its range could cause danger on the communication between v_2 and the rest of the nodes.

From the definition RNG [5] we suppose that the interference area has to be:

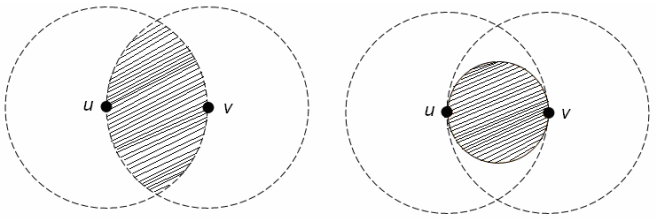
$$I_{u,v}=B(u,dist(u,v))\cap B(v,dist(u,v)), \quad (5)$$

Which is marked with (*a*) and it is shown on fig. 3. In another words, the two points *p* and *q* form RNG then and only then when $I_{u,v}$ does not contains any other points in the *S* area.

The interference area for GG is:

$$I_{u,v}=B((u+v)/2),dist(u,v)/2, \quad (6)$$

marked with (*b*) and shown on fig. 3. The interesting part here is the circled shape of the interference.



(a) $I_{u,v}$ for RNG (b) $I_{u,v}$ for GG

Fig. 5. Interference field in RNG and GG

V. SIMULATION STUDIES OF SENSORY UNITS IN INTERFERENCE

A. Model of interference jamming receivers of sensor nodes

The interference of node *v*, assigned as $I_{(v)}$, may be defined from the count of the other nodes, which areas of transmission are able to cover *v*.

$$I_{(v)} = |\{u \mid u \in V \setminus \{v\}, |uv| \le r_u\}| \quad (7)$$

The middle interference in the geometric structure G, assigned as $I_{avg}(G)$, may be defined as:

$$I_{avg}(G) = \frac{\sum_{v \in V} I_{(v)}}{V} \quad (8)$$

For a node with radius of transmission r_v, may be defined a interference made by it on the following way:

$$I_{(v,r_v)} = |\{u \mid u \in V \setminus \{v\}, |uv| \le r_v\}| \quad (9)$$

So we have:

$$I_{avg}(G) = \frac{\sum_{v \in V} I_{(v)}}{V} = \frac{\sum_{v \in V} I_{(v,r_v)}}{V} \quad (10)$$

Б. Simulation results

To estimate the effectiveness of the use of energy resources and the impact of interference in the transmission of data is necessary to analyze the relative position of sensor nodes and the relevant processes in communication between the sensor modules.

If all devices on the network do not change their position and are homogeneous in terms of hardware components, there is a fixed set of positions in which each sensor node will have the same radio range and will spend the same amount of energy for communication.

As we mentioned earlier the sensor nodes cannot receive and process several signals at the same time, and because of that for clustering of multiple nodes the probability for disturbing of the communication between them is very large.

Using the topology based on RNG we define the parameter θ (fig.6.), which is the angle, that composes from the sensor node of the two abutments of the circles, describing their own range and the range of the neighboring node.

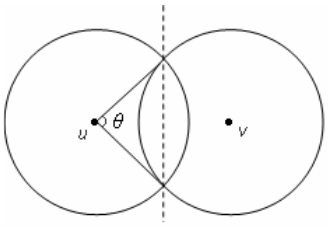


Fig.6 Graphical representation of the angle θ.

Fig. 7 shows the influence of the interference when θ changes, as to more accurately simulate real communication between neighboring nodes, we add an additional disturbing influence (Gaussian noise), which may be amended percent depending on weather, topography and more.

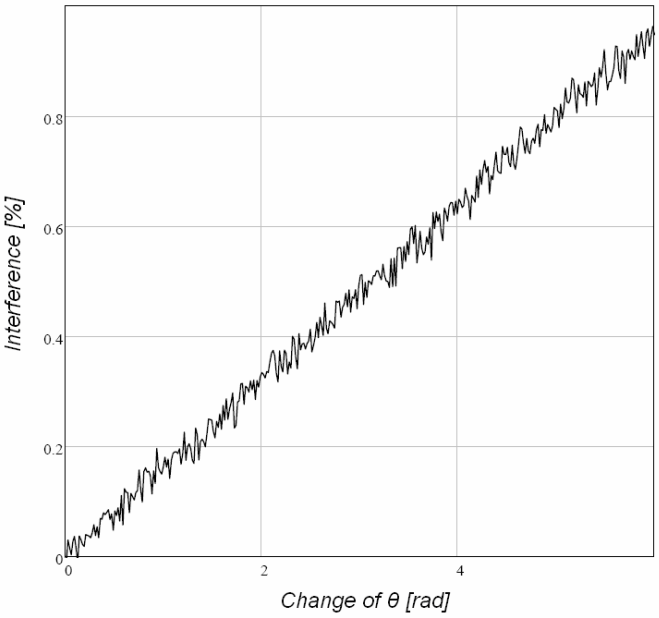


Fig. 7. The change of the interference depending on the parameter of θ

The distance between the mobile sensor nodes causes influence over the interference. When neighboring nodes are located very close, θ has large values, therefore, a high interference. It is the use of geometric structure to limit the communication links between nodes.

The subordination of θ from the distance is shown on fig. 8.

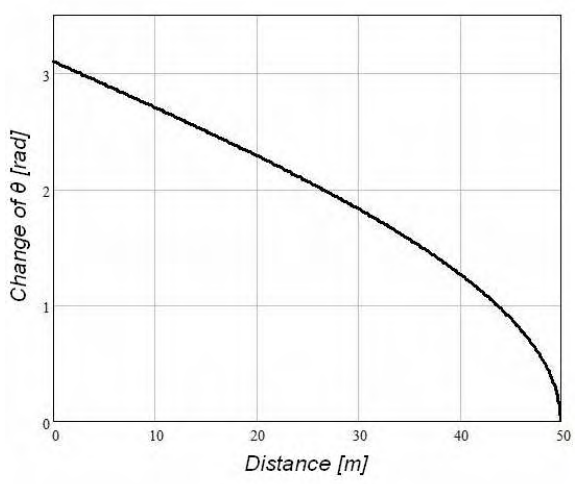


Fig.8. The change of the parameter θ when the distance changes

To define the amount of the interference at GG, shown on fig. 9., we use the recovering that the "Empty Circle" ability makes, and the distance between the neighboring nodes.

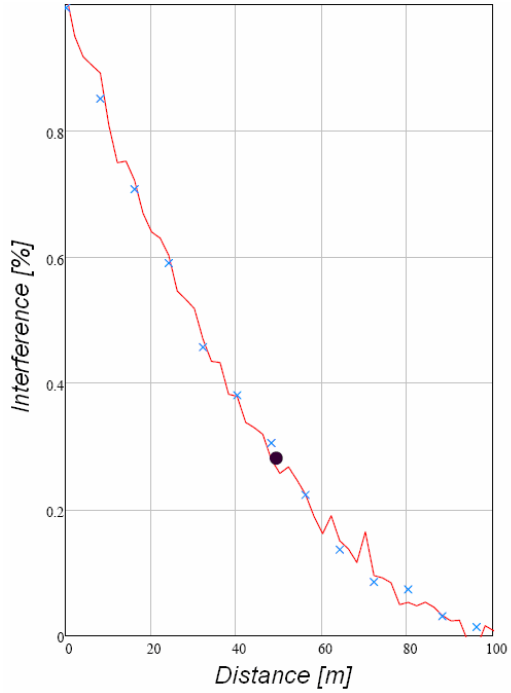


Fig. 9. Relation between the distance and interference when using topology with GG

The black point represents the certain event at a set distance between adjacent sensor nodes and the percent hedging, from which the influences of interference in this topology. It gets clear from the graphic that when the distance between the neighboring nodes is increased the interference will be significantly reduced.

VI. Conclusion

Although the algorithm for calculating RNG and GG is pretty easy to do, from the graphics we see either excessive accumulation of sensor nodes, or not enough count neighbors for a normal communication, which significant impacts on interference.

We study the influence of the interference at the topology, build on RNG and GG, and the results of the simulation are the following:

- The interference may be damped, if we localize the amount of communication connections;
- If we have n number nodes distributed in certain geographic area it is necessary to build particular communication structure;
- We must use a tree structure if we want to have optimal topology related with less interference;
- The interference will be less if the sensor nodes are placed in a distance that they communicate in a range close enough to r_{max}.

The results, based on the analysis provide guidelines for design optimization of interference in sensor networks.

Our studies in future are directed to improve the method that allows exchange of the sensor nodes between the different work groups. Also our simulation method of the interference between the sensors may be used for an online education. We also analyze other techniques to improve the interference effect in the future.

Acknowledgement

This paper was developed in the frames of the research project № DVU01/0109 "Research on Cross Layer Optimization of Telecommunication Resource Allocation" supported by FNI, Ministry of Education, Bulgaria.

References

[1] G. Ferrari, "Sensor Networks Where Theory Meets Practice", 2010.
[2] Miguel A. Labrador, Pedro M. Wightman, "Topology Control in Wireless Sensor Networks" 2009.
[3] Ivan Stojmenovic, "Handbook of Sensor Networks Algorithms and Architectures" 2005.
[4] X. Li, I. Stojmenovic, and Yu Wang. Partial Delaunay triangulation and degree limited localized Bluetooth scatter net formation. IEEE Transactions on Parallel and Distributed Systems, 15(4):350–361, April 2004
[5] G. Toussaint. The relative neighborhood graph of finite planar set. Pattern Recognition 12(4):261–268, 1980
[6] K. R. Gabriel and R. R. Sokal. A new statistical approach to geographic variation analysis. Systemic Zoology, 18:259–278, 1969
[7] K. Moaveni-Nejad and X. Li. Low-interference topology control for wireless ad hoc networks. Ad Hoc & Sensor Wireless Networks: an International Journal, 2004.
[8] M. J. Laszlo, Computational Geometry and Computer Graphics in C++, Prentice Hall, 1996.
[9] X.-Y. Li, P.-J. Wan, W. Yu, O. Frieder. .Sparse power efficient topology for wireless networks., IEEE Hawaiian Int. Conf. on System Sciences (HICSS), 2002
[10] Wang, W., Li, X., Moaveninejad, K., Wang, Y., Song, W., The spanning ratio of β-skeletons. Proc. CCCG, (2003) 35-38
[11] P.V. Nikitin and K.V.S. Rao, "Theory and Measurement of Backscattering From RFID Tags," IEEE Antennas and Propagation, vol. 48, No. 6, December 2006, pp. 212-216.

A Proposal for Harmonic Rejection Mixer Avoiding Irrational Weighting Ratios

Ludwig Lubich[1]

Abstract – **A harmonic rejection mixer (HRM) suitable for SDR/CR applications is proposed. Two main difficulties in HRM realization are addressed: The necessity to use irrational weighting ratios and the need of a clock frequency, which is a multiple of the desired LO frequency. In the proposed HRM the first not rejected harmonic is 7-th harmonic. Instead of being 8, as in a comparable traditional HRM, the clock division ratio here is 3 and no weighting is used.**

Keywords – **harmonic rejection mixer, receiver front-end, software receiver.**

I. INTRODUCTION

Due to the nonlinearity of the mixer and the presence of higher harmonics in the waveform of the local oscillator (LO), the heterodyne receivers are susceptible to interference from signals present at $f = nf_{LO} \pm f_i$, where f_i is the intermediate frequency (IF) used. In traditional receivers, high-order preselect filters are used to solve this problem. However, such filters are difficult or impossible to integrate on-chip and are space and sometimes power consuming. Therefore, it is desirable to relax the RF filtering requirements as much as possible.

A key requirement for software radio receivers, especially for cognitive radio applications is the ability to cover a wide frequency range. In such cases using a square wave LO is preferable to sinusoidal, since the wide range of LO frequencies needed, can be easily produced in an IC by means of a PLL synthesizer with a relative narrow frequency range which is combined with digital frequency dividers. However, the square LO waveform contains strong harmonics and this exacerbates the requirements to RF filtering.

Another popular solution when designing receiver ICs, especially zero-IF receivers, is to use passive switching mixers because of their better linearity, nearly zero power consumption, and lower 1/f noise [1]. But, the use of such mixers also adds up to the RF pre-filtering requirements.

The RF filtering requirements can be significantly relaxed when a harmonic rejection mixer (HRM) is used. In a HRM some unwanted mixing products can be cancelled [2]. A HRM (Fig. 1) consists of several switching mixers driven by separate LO square waves with different phases or duty cycles. The input signals for the individual mixers are delivered by separate RF preamplifiers with weighted gains. The individual mixer outputs are connected in parallel. Such a HRM is equivalent to a single mixer, driven by an effective

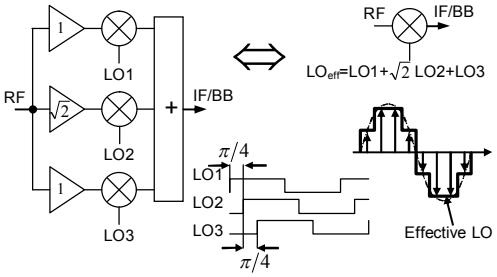


Fig. 1. A typical harmonic rejection mixer

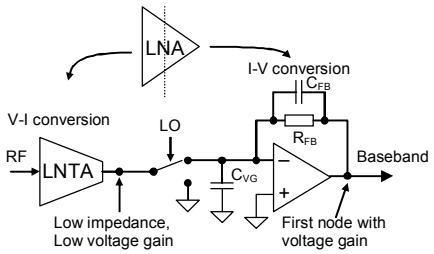


Fig. 2. Conceptual diagram of low-pass blocker filtering [3]

LO waveform in which some harmonics are removed. The LO waveform is in fact a weighted sum of the individual LO waveforms. In HRM, which have been proposed up to now, the effective LO waveform is based on a uniformly sampled sinusoid.

Relaxed RF filtering imposes increased requirements on the LNA and the mixer linearity in the presence of strong interferers. This is a very serious concern especially in receiver IC design, due to the current trend of decreasing supply voltages and the corresponding reduction of voltage headroom.

A successful solution of this problem is presented in [3], in which an integrated CMOS zero IF receiver front-end, using a HRM in conjunction with a well-devised architecture, is proposed. A conceptual block diagram of this receiver is presented in Fig. 2. Voltage amplification in a LNA is usually realized using voltage-to-current conversion via a transconductance followed by current-to-voltage conversion via some impedance or transimpedance. In the proposed receiver, the passive mixer is inserted between a low-noise transconductance amplifier (LNTA) and the virtual ground node of a transimpedance amplifier (TIA). Due to the RC network in its feedback loop, it also acts as a LPF. The mixer load impedance is maintained low in a wide frequency range, which leads to a low LNTA load impedance, too. Thus, large voltage swings at the LNTA output and in the mixer are avoided. The first voltage gain occurs at baseband after the low-pass filtering, where the out-of-band interferers are significantly attenuated.

[1]Ludwig Lubich is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: lvl@tu-sofia.bg.

A serious difficulty in HRM realization is the need of irrational weighting ratios. The required ratios have to be approximated by rational numbers, which leads to some errors, resulting in worsening the harmonic rejection (HR).

Another difficulty is the need of a multiphase LO. This can be realized by shift registers (SR) or using a delay locked loop (DLL) with a tapped delay line. The DLL implementation is disadvantageous due to the phase noise amplification in the DLLs [4] and the phase noise accumulation from one delay cell to the next one along the delay line [5]. A serious disadvantage of the SR implementation is its frequency dividing nature. If the first harmonic that is not rejected is *n*-th, a LO with (*n*+*1*) phases is required. In this case a clock frequency $f_{CLK} = (n+1) f_{LO}$ is needed.

It is clear by intuition that the values of a uniformly sampled sinusoid are related by irrational ratios. However, it is interesting to see whether there are special cases where the ratios in question are rational. We examined this problem and proved (see the Appendix) that the irrational ratios can be avoided only when the number of samples per period is not greater than 6. Therefore, another approach to harmonic rejection is desirable.

The purpose of this work is to develop a HRM in which the above mentioned problems are mitigated. This can be done at the cost of some HR degradation at the highest frequencies received. However, harmonic mixing is not such a serious concern at those frequencies.

The rest of the paper is organized as follows. Section II describes the proposed approach to harmonic rejection and the considerations related to the practical implementation of HRM based on this approach. Section III gives the simulation results.

II. HRM APPROACH PROPOSAL

Generally, a periodic pulse train with a frequency $f = 1/T$ contains components at all possible multiples of f including DC. In the case of rectangular pulses, the components with frequencies of n/T_P will be zero. If a copy of this pulse train is shifted by π and subtracted from original pulse train, all even harmonics and the DC component will be canceled. If a copy of the resulting waveform is shifted by $\pi/3$ and added to it, all harmonics of orders $3(2n+1), n = 0,\ 1,\ 2,\ 3,...$ will be canceled (see Fig. 3) and so on. Mathematically this can be described by following recursive formula:

$$p_0(t) = \mathrm{rect}(t/T_P) \otimes [\delta(t) - \delta(t - T_{LO}/2)] \otimes \sum_{n=-\infty}^{\infty} \delta(t - nT_{LO})$$

$$p_{i+1}(t) = p_i(t) + p_i(t) \otimes \delta[t - T_{LO}/(2h_{i+1})],$$

where h_{i+1} is the order of harmonic canceled by the current addition. Fourier series coefficients will be:

$$C_{n0} = \frac{T_P}{T_{LO}} \mathrm{sinc}\left(\pi n \frac{T_P}{T_{LO}}\right) \cdot \left[1 - (-1)^n\right] \text{ and}$$

$$C_{k(i+1)} = C_k \left(1 + e^{-j\pi n/h_{i+1}}\right).$$

No weighting is needed. A drawback of this approach is that the complexity doubles for each additional rejected group of harmonics. It seems that this approach can be applied for the cancellation of harmonics up to 7th order.

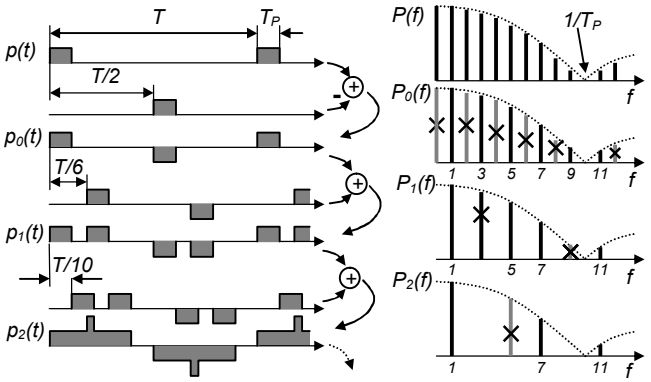


Fig. 3. Illustration of proposed HR approach

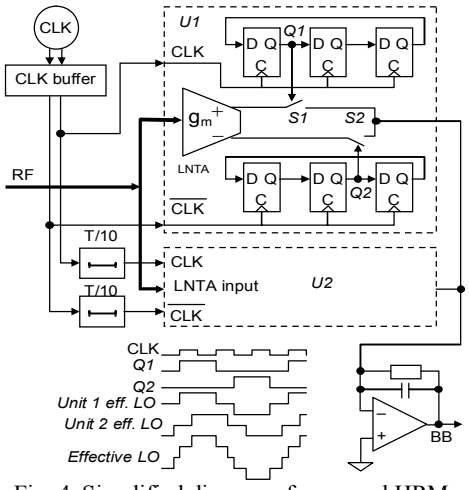


Fig. 4. Simplified diagram of proposed HRM

All the necessary time intervals can be obtained using some form of frequency dividers (e. g. shift registers). A drawback of this technique is the need of a clock frequency, which is a multiple of the LO frequency. Alternatively, delay cells can be used. However, delay cells increase the jitter variance roughly proportionally to the implemented delay [5]. Therefore, it is desirable that the largest delays are to be implemented without using delay cells.

Fig. 4 presents a simplified block diagram of HRM based on above considerations. The first harmonic that is not canceled is the 7th harmonic. The mixer is involved in a receiver architecture similar to the one described in [3]. The mixer contains two identical units - U1 and U2 – each one having two SRs with "100" bit patterns circulating. As a result, the pulse width at the flip-flop outputs is $T_{LO}/3$. Hence, the harmonics of orders $3n$ are suppressed. The switches S1 and S2, connected to the direct and inverse outputs of the LNTA respectively, are controlled by pulses spaced at $T_{LO}/2$. As a result, all even harmonics are cancelled. By adding the U1 and U2 output currents at the virtual ground node of the TIA, the harmonics of orders $5(2n+1)$ are canceled. Additional switches can be controlled by the unused shift register outputs and used to obtain three effective LO spaced at $2\pi/3$. We designate them as R, S and T.

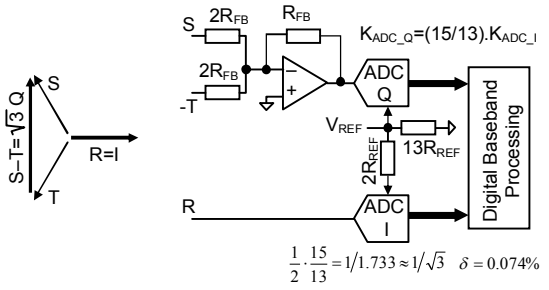


Fig. 5. Obtaining quadrature BB signals

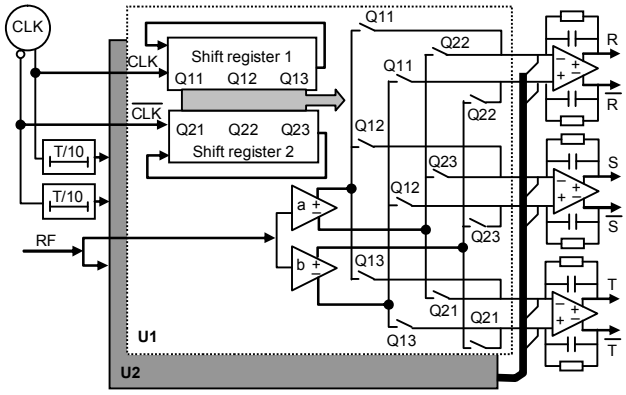


Fig. 6. Proposed HRM

Such a HRM appears to be unable to provide quadrature baseband output signals. A possible solution of this problem is suggested in [6]. We propose a simpler solution illustrated by Fig. 6.

A possible practical implementation of this HRM is presented in Fig. 6. The mixer consists of two identical units again. Each unit has two identical LNTAs. Special attention has been paid to the LNTA output assignment in order to minimize harmonic rejection worsening due to amplitude mismatches. It can be seen that when one input of some of the TIAs is connected to the direct output of a certain LNTA, the opposite input of the same TIA is always connected to the inverse output of the other LNTA. This reduces the influence of LNTA transconductance mismatches on the half-wave symmetry of the effective LO waveforms. Thus, the sensitivity of the even-order HR to LNTA mismatches is reduced.

Quadrature baseband signals can be obtained according to the way shown on Fig. 5. However, it is possible to perform the necessary subtraction directly at the virtual ground node of one of the TIA. In this way the third TIA and the summing amplifier can be spared. This is worth doing because the transimpedance amplifier occupies a large area on the IC die, as can be seen on the micrograph in [3]. In this case, special care should be taken to avoid the formation of low-ohmic paths in the input of the TIA if on-time overlaps occur during switching (Fig. 7) - an effect leading to a significant increase of the TIA noise [7]. When the zero-crossings of the Q LO waveform occur, the I LO waveform is constant, so two of the LNTA outputs (from a total of 4) are used and are unavailable. Therefore in Q zero-crossing moments when a TIA input is disconnected from a LNTA output, the opposite TIA input will be connected to the same LNTA output. If

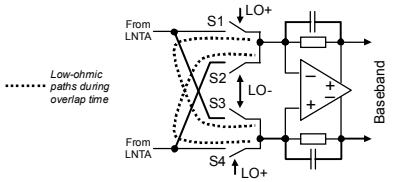


Fig. 7. Low-ohmic paths at the TIA input during on time overlaps

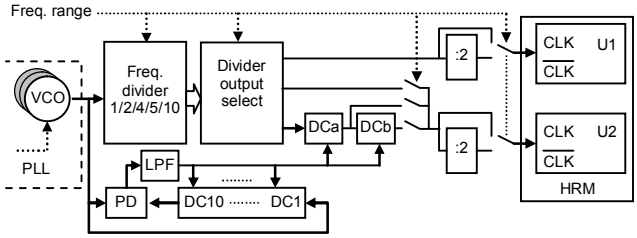


Fig. 8. Clock generation for proposed HRM

there is an on-time overlap, a low-ohmic path between two TIA inputs will arise. Therefore, no LNTA output assignment exists to prevent low-ohmic paths between TIA inputs.

The wide LO frequency range needed is achieved by a PLL synthesizer covering an octave using three VCOs with tuning ranges of about ±12% and frequency dividers. The considerations, leading to this solution, are presented in [8] and [9].

A conceptual block diagram of the CLK generation circuitry is given in Fig. 8. The logical signals are differential, but are presented as single ended for simplicity.

Even division ratios are used only to obtain pulses with a true 50% duty cycle. The latter is important for obtaining high even harmonic rejection ratios. When $f_{LO} > f_{LO\max}/2$ no frequency division is used and we rely solely on the symmetry of the VCOs, therefore even harmonic rejection is poor. However, at the highest receiving frequencies the harmonic mixing presents less concern.

The delay needed for the 5th order HR is implemented by controlled delay cells and an appropriate design of the frequency divider. It can be found out that only two delay values are required - $T_{VCO}/10$ and $2T_{VCO}/10$ - independent of the LO frequency. The corresponding delay cells are connected to appropriate outputs of the divider depending on the VCO division ratio used. When $f_{LO} \le f_{LO\max}/10$, the needed delay value is obtained exactly without using delay cells. The control voltage for the delay cells is produced in a DLL with a delay line consisting of 10 delay cells of the same type as DCa and DCb.

III. RESULTS

Monte Carlo simulations were performed (10 000 runs) to estimate HR worsening due amplitude and phase mismatches. CLK duty cycle error, timing errors due to shift registers and delay cell, as well as amplitude mismatches are assumed to be uncorrelated, normally distributed random variables with standard deviations σ_{DtC}, σ_{ShR}, σ_{DU} and σ_A respectively.

We assumed that $\sigma_{ShR} = 0.08$ ps (as in [3]), $\sigma_{DtC} = 10^{-2} T_{VCO}$

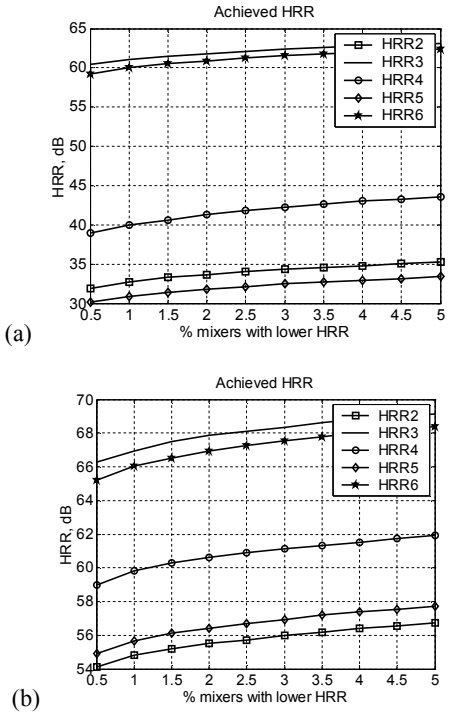


Fig. 9. Achieved HRR, (a) without VCO divider ($f_{LO} > f_{LO\max}/2$), and (b) VCO division ratio 2 , i.e. $f_{LO\max}/4 < f_{LO} < f_{LO\max}/2$.

when $f_{LO} > f_{LO\max}/2$ and $\sigma_{DtC} = 0.08\,\text{ps}$ in the other cases, $\sigma_{DU} = 10^{-2}\tau_{nom}$, and $\sigma_A = 1.10^{-3}$. VCO frequency was 6 GHz.

Fig. 10 shows the HRR versus the percentage of mixers with inferior HRR.

As it was expected, the HRRs worsen significantly for the highest LO frequencies when no frequency divider is used.

For LO frequencies below $f_{LO\max}/2$, the clock signal has a true 50% duty cycle. The latter is essential for the even order HR and 5-th order HR in presented HRM. In addition, relative time errors decrease when the LO period rises. As a result, the HRR's are significantly increased.

IV. Conclusion

A HRM was presented in which no irrational weighting ratios are needed. The clock frequency for up to the 6th order HR is $3f_{LO}$ instead $8f_{LO}$ as in comparable traditional HRM. Additionally, no weighting is used. The penalty is some HR degradation at highest LO frequencies. On the other hand HRR can be significantly improved if a digital cancellation technique is applied, similar to those in [3].

Appendix

We assume that there exist $P \in \mathbf{N}$, $P > 6$ and some phase displacement $\alpha \in (0, 2\pi/P]$, such that $\sin(\alpha + n \cdot 2\pi/P)/\sin(\alpha) \in \mathbf{Q}$ is fulfilled for each $n \in \mathbf{N}$.

If the above statement is true, then it is fulfilled for $n = 1$ and $n = P - 1$ in particular. Therefore, $\sin(\alpha + 2\pi/P)/\sin(\alpha) = \cos(2\pi/P) + \cot(\alpha)\sin(2\pi/P) = k/l$ and $\sin[\alpha + (P-1)\cdot 2\pi/P]/\sin(\alpha) = \sin(\alpha - 2\pi/P)/\sin(\alpha) = \cos(2\pi/P) - \cot(\alpha)\sin(2\pi/P) = p/q$, where $k, l, p, q \in \mathbf{N}$. After summing up the two equations above, we obtain $2 \cdot \cos(2\pi/P) = \frac{kq + pl}{lq}$. Therefore $\cos(2\pi/P) \in \mathbf{Q}$.

According to Niven's theorem [10] if x/π and $\sin(x)$ are both rational, then the sine takes values 0, $\pm 1/2$ and ± 1. We found, that $\cos(2\pi/P) \in \mathbf{Q}$. But $\cos(2\pi/P) = \sin(\pi/2 - 2\pi/P)$. Obviously $(\pi/2 - 2\pi/P)/\pi = (1/2 - 2/P) \in \mathbf{Q}$. Therefore $\cos(2\pi/P)$ can take values 0, $\pm 1/2$ and ± 1. As a result, the minimal non-zero value of $2\pi/P$ is $\pi/3$. Hence $P \le 6$ must be fulfilled contradicting our requirements. Therefore no $P \in \mathbf{N}$, $P > 6$ and $\alpha \in (0, 2\pi/P]$ exist such that $\sin(\alpha + n \cdot 2\pi/P)/\sin(\alpha) \in \mathbf{Q}$ for each $n \in \mathbf{N}$.

References

[1] Yeo, K. S., M. A. Do, C. C. Boon, *Design of CMOS RF Integrated Circuits and Systems*, World Scientific, 2010

[2] Weldon, J.A., Narayanaswami, R.S., Rudell, J.C., Li Lin, Otsuka, M., Dedieu, S., Luns Tee, King-Chun Tsai, Cheol-Woong Lee, Gray, P.R., "A 1.75-GHz highly integrated narrow-band CMOS transmitter with harmonic-rejection mixers", Solid-State Circuits, IEEE Journal of , Dec 2001, Volume: 36 Issue:12, pp. 2003 – 2015

[3] Ru, Zhiyu and Moseley, Niels A. and Klumperink, Eric A.M. and Nauta, Bram, (2009) "Digitally-Enhanced Software-Defined Radio Receiver Robust to Out-of-Band Interference", IEEE Journal of Solid-State Circuits, vol. 44 (12), 2009, pp. 3359-3375.

[4] Lee, M.-J.E. , Dally, W.J., Greer, T., Hiok-Tiaq Ng, Farjad-Rad, R., Poulton, J., Senthinathan, R., "Jitter transfer characteristics of delay-locked loops - theories and design techniques", Solid-State Circuits, IEEE Journal of, Apr 2003, Volume: 38, Issue: 4, pp. 614-621

[5] Xiang Gao, Nauta, B., Klumperink, E.A.M., "Advantages of Shift Registers Over DLLs for Flexible Low Jitter Multiphase Clock Generation", Circuits and Systems II: Express Briefs, IEEE Transactions on, March 2008, Volume: 55, Issue: 3, pp. 244-248

[6] Van Sinderen, J. et al., "Conversion System", US Patent Application Publication, US2012/0008717 A1, Jan. 12, 2012

[7] Redman-White, W. and Leenaerts, D. W. M., "1/f Noise in Passive CMOS Mixers for Low and Zero IF Integrated Receivers". Proceedings of the European Solid State Circuits Conference, September 2001, Villach, Austria.

[8] Razavi, B., "Challenges in the design of cognitive radios", IEEE Custom Integrated Circuits Conference, September 2009. pp. 391-398

[9] Abidi, A.A., "The Path to the Software-Defined Radio Receiver", Solid-State Circuits, IEEE Journal of, May 2007, Volume: 42, Issue: 5 pp. 954-966

[10] Weisstein, Eric W. "Niven's Theorem." From MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/NivensTheorem.html.

Efficient Estimation of the Antenna Noise Level Using Neural Networks

Ivan Milovanovic[1], Zoran Stankovic[2] and Marija Milijic[3]

Abstract – **This paper presents how through the use of artificial neural networks we can accelerate the prediction procedure of the external noise level at the receiving point of wireless communication systems. Were taken into account only the effects of natural noise sources, which are surrounded by the antenna system and considerably more stable than artificial. The case of microwave wireless transmission, where dominated influence of noise generated by emissions of gases from the atmosphere (primarily oxygen and water vapour), is considered. Accordingly, we developed a neural network model for antenna noise temperature prediction of the RF receiver based on Multilayer Perceptron (MLP) network. The architecture of this model, the results of its training and testing and simulation results are presented in this paper in the appropriate sections.**

Keywords – **Neural network, Antenna Noise, Brightness temperature.**

I. INTRODUCTION

The explosive growth of wireless systems poses increasing number of technical challenge and performance demanding necessary to support many wireless application. The wireless system design goal is to achieve the largest possible coverage area in which the received power is sufficiently strong compared to background noise. Consequently, one of fundamental parameters in wireless communication is signal-to-noise power ratio that indicates the reliability of the link between the transmitter and receiver. Therefore, it certainly helps to have a reliable tool to estimate noise power during the process of wireless systems designing.

Today, the most often used is recommendation ITU-R P.372-10 to estimate extern noise of RF transmitter [1]. Recommendation ITU-R P.372 provides data on radio noise external to the radio receiving system which derives from the following causes: radiation from lightning discharges (atmospheric noise due to lightning); aggregated unintended radiation from electrical machinery, electrical and electronic equipments, power transmission lines, or from internal combustion engine ignition (man-made noise); emissions from atmospheric gases and hydrometeors; the ground or other obstructions within the antenna beam; radiation from celestial radio sources. Many noise dependences [1] are represented by formula whose parameters should be determined from a lot of complex figures. Classical use of Recommendation ITU-R P.372 requires figures visual reading with applying challenging interpolation methods resulting in time-consuming, forceful process with non-satisfactory accuracy.

The use of Artificial Neural Network (ANN) is good tool to overcome all specified problems. ANN 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. ANN has the capability of a functional dependence's modeling exclusively on the basis of input data [2-5]. Neural network architecture which is consisted of connected small processing units (neurons). In this way, neural network can be used for modeling high-distributed and high-parallel problems [2-5]. The second is neural network ability to learn function dependence on the basis of solved examples rather then to learn to execute some well known function dependence. After successful learning process of neural network, it can be used not only for known examples but also for unknown examples (generalization).

Neural network has been used for estimation level of RF receiver external noise versus only frequency, without taking into account the parameters that describe the antenna environment [4]. In this paper, neural model for prediction temperature of noise source brightness is developed resulting in more effective estimation of receiver external noise dependence on frequency and antenna elevation in microwave range.

II. SPECIFICATION OF NOISE INTENSITY OF WIRELESS COMMUNICATION SYSTEM

The noise factor, *f*, for a receiving system is composed of a number of noise sources at the receiving terminal of the system [1]. Both internal and external noise must be considered. For receivers of the wireless communication system, the system noise factor is given by [1]:

$$f = f_a + (f_c - 1) + l_c(f_t - 1) + l_c l_t(f_r - 1) \quad (1)$$

where l_c is antenna circuit loss, l_t is transmission line loss and f_r is noise factor of ideal antenna and f_t is the noise factor associated with the transmission line losses. f_a is the external noise factor defined as:

$$f_a = \frac{p_n}{kt_0 b}, \qquad F_a = 10\log f_a \quad (2)$$

where p_n is available noise power from an equivalent lossless antenna, k is Boltzmann's constant = 1.38×10^{-23} J/K, t_0 is reference temperature taken as 290 K and b[Hz] is noise power bandwidth of the receiving system [1].

[1]Ivan Milovanovic is with the Singidunum University, DSL center Nis, E-mail: ivanshix@gmail.com

[2]Zoran Stankovic is with the Univeristy of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: zoran.stankovic@gmail.com

[2]Marija Milijic is with the Univeristy of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: marija.milijic@gmail.com

External noise factor can be presented using effective temperature of antenna noise t_a:

$$f_a = \frac{t_a}{t_0}, \qquad t_a = \frac{P_a}{kB} \tag{3}$$

where P_a is external noise power collected by antenna.

The available noise power is obtained by summing the contributions of each individual noise sources. To be able to perform the calculation it is necessary to introduce a parametar that determines the noise radiation sources. The parameter used in that sense commonly is brightness [6,7]. Taking into account the Planck law of black body radiation in the radio frequency spectrum and using the Raleigh-Johnson approximation, the brightness in the direction θ, φ from which noise of frequency f comes can be expressed as:

$$S(f,\theta,\varphi) = \frac{2kt_b(\theta,\varphi)}{\lambda^2} \tag{4}$$

where $t_b(\theta,\varphi)$ brightness temperature in the observed directionθ, φ, which originates from noise sources. Accordingly, effective temperature of brightness t_b from the body radiating noise is defined using power of noise radiation P_b [1,6]:

$$t_b = \frac{P_b}{kB} \tag{5}$$

Integrating noise power at all spatial angles and taking into account the characteristics of antenna F(θ,φ)antenna noise temperature can be expressed in a way [6,7]

$$t_a = \frac{\int_0^{2\pi}\int_0^{\pi} F(\theta,\varphi)\, t_b(\theta,\varphi)\sin\theta\, d\theta\, d\varphi}{\int_0^{2\pi}\int_0^{\pi} F(\theta,\varphi)\sin\theta\, d\theta\, d\varphi} \tag{6}$$

Natural source noise can be atmospheric noise, cosmic noise, noise from Earth and noise from different cosmic objects. Cosmic noise decreases approximately with the square of the frequency so that the above 1 GHz is very small and can be ignored by receiver operating in the microwave range. Noise from Earth, that correlates average noise temperature of 254 K, is important only for satellite antenna with the main radiation bean directed to Earth. There are number noises from many cosmic objects, but the only significant is the noise from the Sun. The Sun noise significantly affects on antenna noise only when large direction antenna with main radiation beam directed to the Sun. Atmospheric noise can derive from two sources. In The first is electrostatic discharge in atmosphere that overcomes for frequency range bellow 50 MHz. The last is emission in atmosphere due to water vapor and oxygen that is dominant in high frequency range Figure 1. shows temperature of atmosphere brightness versus antenna elevation and frequency when average concentrate of tropopause water vapor is 7.5 g/m^2. This figure is part of ITU-R P.372-10 Recommendation [1]. Considering only atmospheric influence and if space angle of antenna effective radiation Ω_a, is less then space angle of noise source radiation Ω_b, temperature of antenna noise can be equalized with temperature of noise source brightness

$$t_a \approx t_b \quad , \quad \Omega_a < \Omega_b \tag{7}$$

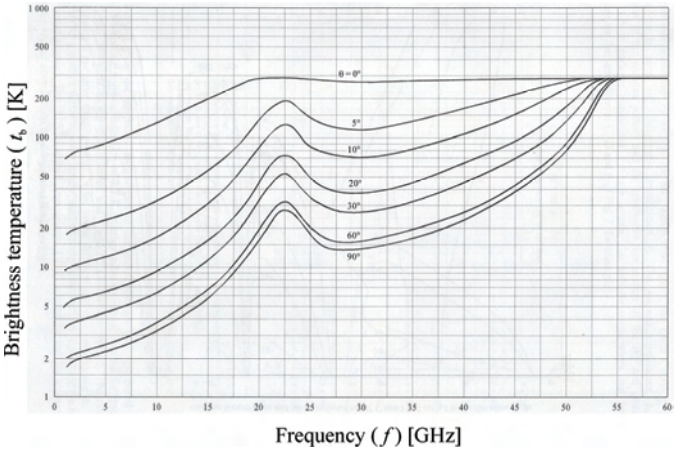


Figure 1. Temperature of atmosphere brightness versus antenna elevation and frequency when average concentrate of tropopause water vapor is 7.5 g/m^2 for calm and good in standard atmosphere weather

III. Neural Model of Microwave Antenna Noise Temperature

The model of noise temperature of receiver antenna in wireless communication system in microwave range considers only influences by atmosphere as dominant noise source while other noise sources are taken as inappreciable. For large space angle of antenna radiation, antenna noise temperature is approximately equal as temperature of antenna brightness from atmosphere that radiates noise. Also, it considers calm and good weather with constant atmosphere condition with average concentrate of water vapor is 7.5 g/m2. For given conditions, brightness temperature depends on antenna elevation angle and frequency. The problem should be modeled as the function

$$t_b = g(\theta, f) \tag{8}$$

The neural model given as function $\mathbf{y}=y(\mathbf{x},\mathbf{w})$, where y is neural network function and $\mathbf{w}$ is a connection weight matrice among neurons [2,3], has input vector $\mathbf{x}=[\theta, f]^T$ and output vector $\mathbf{y}=[t_b]$. The modeling brightness is done by using Multilayer Perceptron Network (MLP) with appropriate MLP neural model defined as:

$$t_b = y([\theta, f]^T, \mathbf{w}) = f_{MLP}(\theta, f, W) \tag{9}$$

where f_{MLP} is transfer function of MLP network used for realization neural model. If weight matrice $\mathbf{w}$ is presented as matrix structure, it can cause difficulties in implementation neural network and in its training algorithm. For this reason, neural network weight matrice $\mathbf{w}$ is replaced by set of neural network weights whose elements are weight matrices and vector of biases of neural network layers. During process of

training, values of weights W change to adjust function f_{MLP} to model function.

The figure 2. presents the architecture of MLP neural model of antenna brightness temperature versus atmosphere in microwave range while the atmosphere conditions are constant. The vector of l-th hidden layer outputs can be presented using vector $\mathbf{y}_l$ with dimension $N_l \times 1$ where N_l is number of neurons in l-th layer. i-th elements of vector $\mathbf{y}_l[i]$ is output of i-th neurons from s-th neural layer ($s=l+1$ considering input layer also) $v_i^{(s)}=v_i^{(l+1)}$, viz $\mathbf{y}_l = [v_1^{(l+1)}, v_2^{(l+1)}, \ldots, v_{N_l}^{(l+1)}]^{\mathrm{T}}$. Further

$$\mathbf{y}_l = F(\mathbf{w}_l \mathbf{y}_{l-1} + \mathbf{b}_l) \tag{10}$$

where $\mathbf{y}_{l-1}$ is a $N_{l-1} \times 1$ vector of (l-1)-th hidden layer outputs, $\mathbf{w}_l$ is a $N_l \times N_{l-1}$ connection weight matrix among (l-1)-th and l-th hidden layer neurons, and $\mathbf{b}_l$ is a vector containing biases of l-th hidden layer neurons. In the above notation $\mathbf{y}_0$ represents outputs of the buffered input layer $\mathbf{y}_0 = \mathbf{x}$. The element $\mathbf{w}_l[i,j]$ from weight matrix $\mathbf{w}_l$ represents connection weight between i-th neuron of (l-1) hidden layer and j-th neuron of l hidden layer, viz between i-th neuron network layer $s=l$ and j-th neuron in network layer $s=l+1$, while $b_i^{(l)}=\mathbf{b}[i]$ is bias value of i-th neuron in hidden layer l. F, the transfer function of hidden layer neurons, is hyperbolic tangent sigmoid

$$F(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} \tag{11}$$

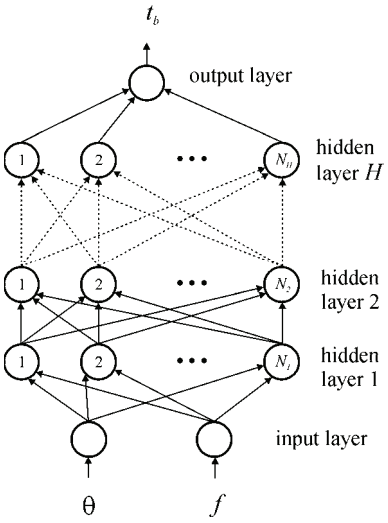


Figure 2. The architecture of MLP neural model of antenna brightness temperature versus atmosphere in microwave range while the atmosphere conditions are constant

All neurons from the last hidden layer H are connected with the neuron of the output layer. Since the transfer function of output layer is linear, the output of the network is:

$$t_b = \mathbf{w}_o \mathbf{y}_H \tag{12}$$

where $\mathbf{w}_o$ is a $1 \times N_H$ connection weight matrix among the H-th hidden layer neurons and output layer neurons (Figure 3). Thus, set of network weights is presented as

$$W = \{\mathbf{w}_1, \ldots, \mathbf{w}_H, \mathbf{w}_o, \mathbf{b}_1, \ldots, \mathbf{b}_H\} \tag{13}$$

The notation of MLP models MLPH-N_1-…-N_i-…-N_H where H represents hidden layers number and N_i is the numbers of neurons of i-th hidden layer.

IV. Modelling Results

MatLab 7.0 software development environment is used for realization and training MLP model. The training of neural model is done using 144 samples that are visual read from the graphics (Figure 1.) that is part of ITU-R P.372-10 Recommendation. The samples are read in frequency range 1.2 GHz ≤ f ≤ 57.5 GHz for antenna elevation θ = 0°, 5°, 10°, 30°, 60° i 90°. Levenberg-Marquartd method is used for training neural model with accuracy 10^{-5}. To achieve the best trained MLP model, many different MLPH-N_1-…-N_i-…-N_H models are trained where $1 \le H \le 2$ and $4 \le N_i \le 30$.

TABLE I. THE TESTING RESULTS FOR EIGHT MLP MODELS

MLP model	*WCE* [%]	*ACE* [%]	r^{PPM}
MLP2-9-5	2.31	0.55	0.9996
MLP2-10-4	6.02	0.50	0.9994
MLP2-8-8	5.49	0.58	0.9993
MLP2-9-8	7.32	0.53	0.9992
MLP2-10-9	4.49	0.85	0.9990
MLP2-20-15	4.56	1.05	0.9986
MLP2-18-16	6.89	0.94	0.9985
MLP2-12-11	9.67	0.93	0.9976

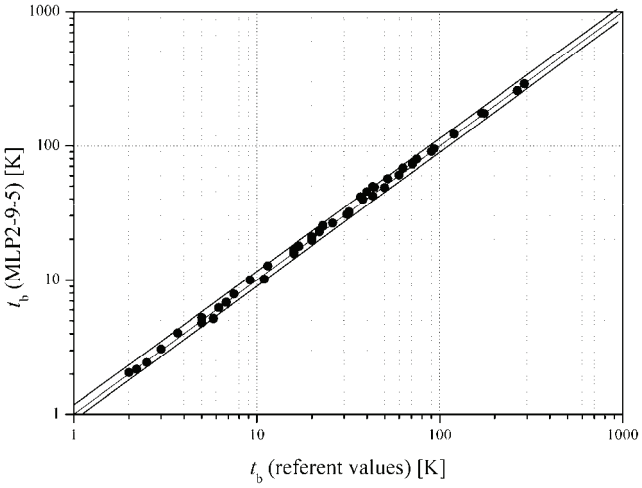


Figure 3. Scattering diagram for MLP2-9-5 model

The test of every trained MLP model is done with the set of 48 samples that are read in frequency 1.2 GHz ≤ f ≤ 57.5 GHz for antenna elevation θ = 20° i 60°. The samples that correspond to antenna elevation θ = 20° have not been used in training. The basic criterion for selection the best MLP network is the maximum value of Pearson Product-Moment correlation coefficient r^{PPM}. [2-5]. Test results of successfully trained MLP networks are presented in the Tables I together with the average test error (*ATE*) and the worst case error (*WCE*).

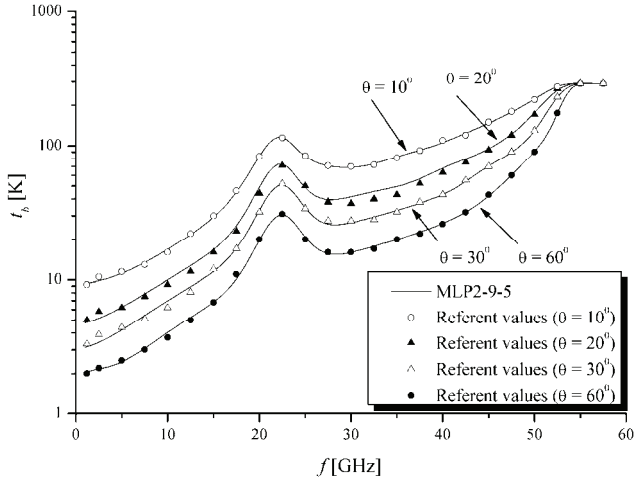


Figure 4. Antenna brightness temperature caused by atmospheric noise versus frequency obtainet by using MLP2-9-5 model for antenna elevation θ = 10°, 20°, 30° and 60° and comparison these values with referent values read from the graphics from ITU-R P.372-10 Recommendation

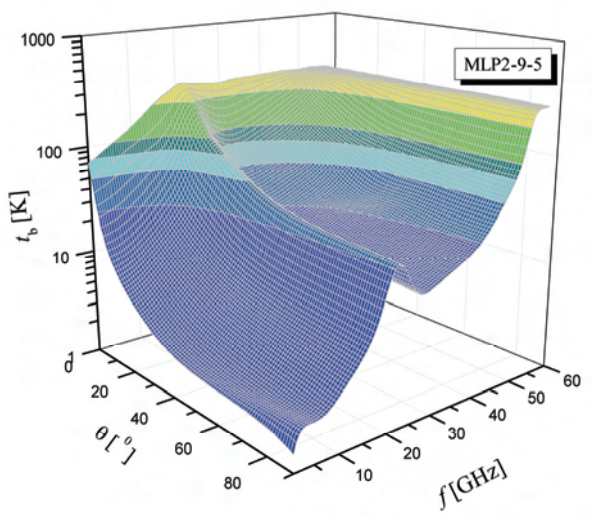


Figure 5. 3D presentation of antenna brightness temperature caused by atmosphere that radiates noise versus antenna elevation and frequency (results were obtained by using MLP2-9-5 model)

The model MLP2-9-5 is chosen as representative model of antenna brightness temperature caused by atmospheric noise. Figure 3. shows the scattering diagram that this model gives in testing process. It can be seen very satisfying agreement between neural model output and samples that are visual read from the graphics in figure 1.

The model MLP2-9-5 is used for simulation of antenna brightness temperature caused by atmosphere that radiates noise versus antenna elevation and frequency. Figure 4. show simulation results for antenna elevation θ = 10°, 20°, 30° i 60° and comparison these values with referent values read from the graphics from ITU-R P.372-10 Recommendation. It can be seen very satisfying agreement between these results and referent values proving the choice of this model. Figure 5 presents 3D dependence of antenna brightness temperature versus atmosphere that radiates noise versus antenna elevation and frequency using 10374 points (114 per frequency x 91 per azimuth). This dependence is got for less then 4secunds using Pentium IV 1.4 GHz and 1GB RAM proving great simulation speed of chosen neural model.

V. Conclusion

During the process of designing the modern wireless communication systems, procedures for estimation of external noise have a very important role due to external noise can significantly influences to services quality of wireless systems. Classic way of visual reading from different ITU recommendation can be time consuming and with great error possibility because of visual reading and applying interpolation formulas. The good alternative can be neural networks model of very complex graphs from ITU recommendation. Neural network model can avoid errors due to manual graphs reading enabling faster calculation of the level of external noise of receiver

Neural model also enables the automation of the process of predicting noise power of receiver making one suitable method for the efficient analysis of the entire coverage area of wireless communication system transmitters in a big number of points that is of vital importance for the design and analysis of all components of modern wireless communication systems.

Acknowledgement

The work was financially supported by the Serbian Ministry of Education and Science

References

[1] Recommendation ITU-R P.372-10, “Radio Noise”, published 10/2009.
[2] S. Haykin, Neural Networks, New York, IEEE, 1994.
[3] J. Hertz, A. Krogh, R. Palmer, “Introduction to the Theory of Neural Computation”, Addison-Wesley, 1991.
[4] Marija Milijić, Zoran Stanković, Ivan Milovanović, „Neuronski model za procenu snage eksternog šuma kod prijemnika bežičnih komunikacionih sistema“, 19th Telecommunications Forum (TELFOR), Belgrade, Serbia, November, 22-24, 2011, Telecommunications Society, pp.437-450, ISBN: 978-1-4577-1498-6.
[5] Zoran Stanković, Bratislav Milovanović, Marija Milijić, "Efficient Neural Model of Microwave Patch Antennas", ICEST 2006 Conference Proceedings, Sofija, Bugarska, 29. Jun- 1. Jul 2006, pp. 49-52
[6] M. Dragović, Antene i prostiranje radio talasa, Beopres, Beograd 1996.
[7] E. Zentner, Antene i radiosustavi, Graphis, Zagreb 1999.

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

Dragan Obradović[1], Igor Stančić[2], Aleksandar Kopta[3], Zoran Mićić[4] and Predrag Manojlović[5]

Abstract – **In this paper is described software and method for automated measuring antennas and antennas systems in wide frequency band. This software can be used with all antennas types and antennas with integrated microwave receivers with analog or digital output. It is also shown results of measuring the horn antennas system with the integrated microwave receivers in frequency range from 15 to 19 GHz.**

Keywords –**Antennas, Antenna measurement, Measurement software, Integrated microwave receivers**

I. INTRODUCTION

The measuring instruments for a long time have the capability of connection with PC computers. This capability helps to create very complex measuring methods which significantly accelerate and simplify measuring, especially if are used a lot of calibration data [1]. The PC computers are usually connected with the instruments over GPIB (IEEE488.2), RS232, USB or Ethernet interfaces.

It is usual to develop or adopt software for each measuring methods [2]. There are many commercial software applications that can be used in the measuring automation like LabView from National Instruments [3][4], Matlab from Mathworks [5] or OpenLab from Agilent [6]. Implementation of the measuring methods with commercial software applications should be developed in the development environment of that applications. If used instruments or equipments are not supported by the commercial applications, it should be developed. When commercial applications are used, it must be also paid for licensing. In some cases is more convenient to develop complete measuring software in the standard development environment, especially if there are many devices in the measuring methods which are custom designed and if they are not supported by commercial applications.

This paper will show the developed software and method for automated measuring antennas and antennas systems with integrated microwave receivers in wide frequency band.

[1]Dragan Obradović is with the IMTEL Komunikacije, Bul. Mihajla Pupina 165b, 11070 Belgrade, Serbia, E-mail: obrad@insimtel.com.

[2]Igor Stančić is with the IMTEL Komunikacije, Bul. Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, E-mail: stanco@insimtel.com.

[3]Aleksandar Kopta is with the IMTEL Komunikacije, Bul. Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, E-mail: kopta@insimtel.com

[4]Zoran Mićić is with the IMTEL Komunikacije, Bul. Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, E-mail: zoran@insimtel.com

[5]Predrag Manojlović is with the IMTEL Komunikacije, Bul. Mihajla Pupina 165b, 11070 Novi Beograd, Serbia, E-mail: pedja@insimtel.com

II. MEASURING METHOD

The measuring method which is shown on Fig. 1. can be used for measuring antenna patterns on a number of frequencies in the cases where received signal can be measured directly on the antennas or if antenna is integrated with receiver which has the analog output.

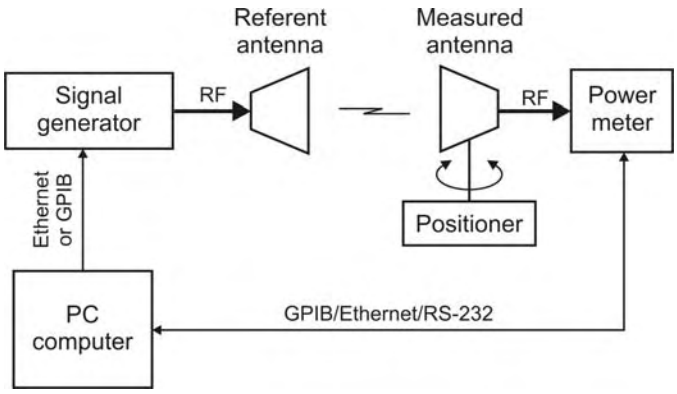


Fig. 1. Automated method for measuring antenna radiation pattern

The measured antenna is on the positioner which has capability of manual fine angle settings. The output of the antenna or integrated receiver is connected to the power meter which is connected to the computer with measuring software. Transmitting referent antenna is connected to the RF signal generator which is also connected to the computer with measuring software. For each angle of the measured antenna, the measuring software sends commands to the signal generator to set desired signal frequencies. For each set frequency, received level from measured antenna is read and stored to the database.

When the antenna radiation pattern with integrated receiver is measured at multiple frequencies in the case when the receiver output is in digital form, the measuring method which can be used is shown on the block diagram in Fig. 2.

This measuring method is similar like previously described, except that output of the integrated receiver is in digital format [8][9]. In some cases receiving frequency channel should be set by measuring software computer.

In both shown methods is used azimuth positioner with manual angle setting. The described methods can be used for

controlled positioner with some modification in the measuring software.

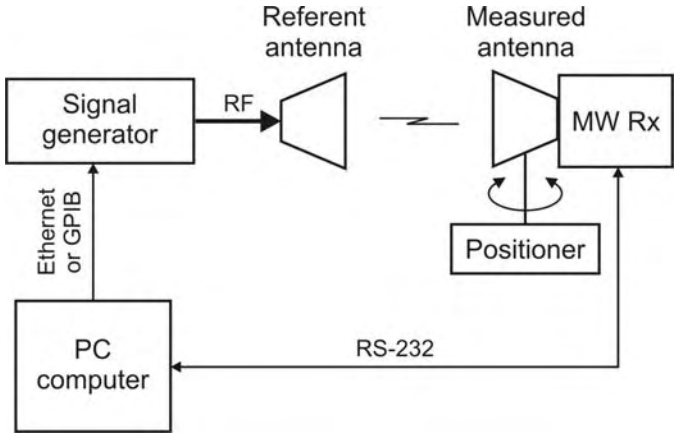


Fig. 2. Automated method for measuring radiation pattern of antenna with integrated receiver with output in digital format

III. Software Organization

On Fig. 3 are show components of the developed software for automated measuring antenna radiation pattern. The Components are same and for other methods.

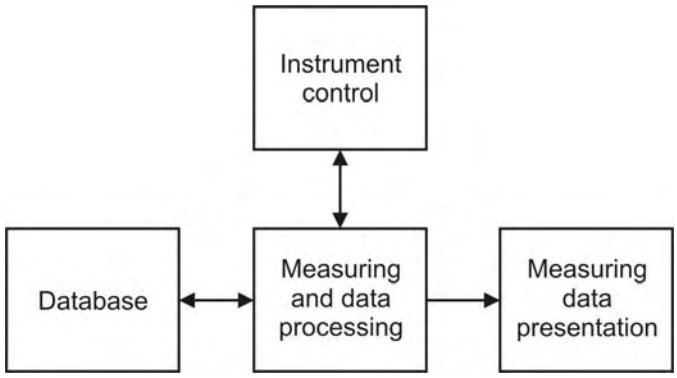


Fig. 3. The components of the software for automated measuring

The instruments control component is realized like class library where is for each type of instrument is created separately abstract class. The software for some measuring method can use any available instrument which is supported by library class and software developing can be focused on measuring method not on controlling and connecting with the instruments.

The results of measuring are stored in database for later processing and presentation. For the database can be used Access file format or database servers like MS SQL or MySQL. It is usually for each measuring method to use two database tables. The first is used for storage data about measuring conditions, and the second to storage measured data.

The presentation component is used to show raw or processed measuring data in desired form. The measured data from database can be shown in graphical form or exported to Excel or other file format for later analysis or using. In this stage of software development, it is used external software application Gnuplot [11] for drawing many types of graphics. Command and data files for Gnuplot are creating by the measuring software.

IV. Instrument Control Library

Source code for instrument control is separated from software for measuring methods into library, which can be used for different measuring applications [12]. The basic request in design of this library was that the programming classes for instruments be grouped by type and function of the instrument, not by type of connection. The library was developed in C# programming language for .NET Framework [13] like dynamic library DLL file.

In Fig. 4 is shown part of the class diagram for this library. All instrument classes are inherited from abstract class Instrument which has methods Open() and Close() for opening and closing connection to instruments. Those methods are implemented in inherited classes and depend from instrument type and connection type. All classes inherited from the class Instrument have attribute InstConnection which describe type and parameter for connection to the instrument.

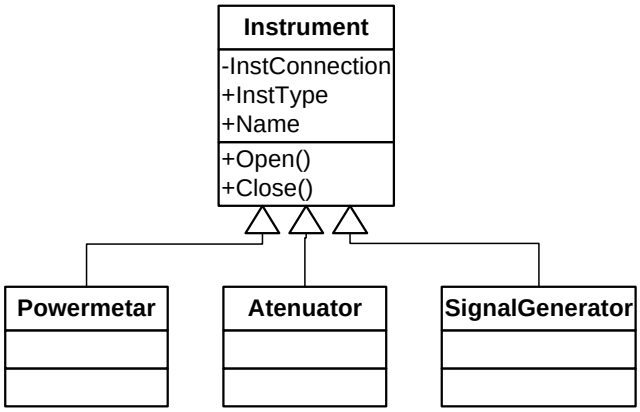


Fig. 4. The abstract class Instrument and inherited classes for instrument types

The classes for particular instrument type by function have methods which are specified for that instrument type. Implementations of those methods are in inherited classes for each instrument model. In Fig. 5. is shown the class diagram for microwave signal generators. For each models the structure SigGenSpec has characteristics of the particular instrument model (e.g. operating frequency and output levels ranges, etc.) and measuring software can read operating range for selected instrument model.

When measuring software is developed it will be made reference to base class for instrument type, not for particular instrument model. It is also possible to list available instrument models, and with static method of the base class it is possible to create object of the selected instrument model. In this example with signal generator, static method CreateSigGen() is defined for this purpose. In this way is possible to change DLL instrument library with new one which support new instrument model without changing and compiling existing measuring software.

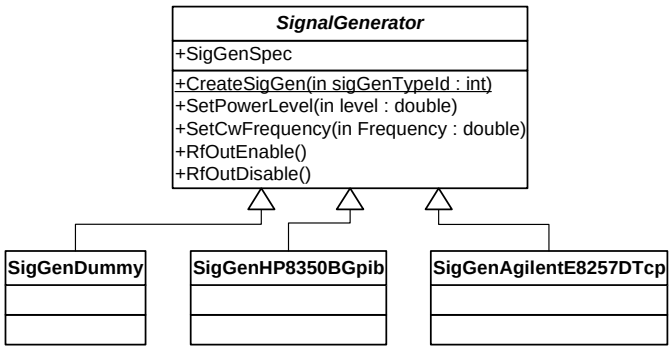


Fig. 5. The abstract class *SignalGenerator* and inherited classes for different instrument models

Not all instrument models should have connection to the real equipment. For example, the class SigGenDummy is virtual instrument and it is very useful during measuring software development.

V. Software Realization

Software for automated antenna radiation pattern measuring is realized like Windows form application which is executed under .NET Framework. Complete software was written in programming language C#. Measuring software application supports measuring for different methods. For each measuring methods were developed different customized windows forms. The software makes possible setup parameters for each measuring and choice of measuring instrument models. Configuration data are stored in XML files. Forms for application configuration are same for all measuring methods. Fig. 6 shows the windows form for selection and configuration instruments.

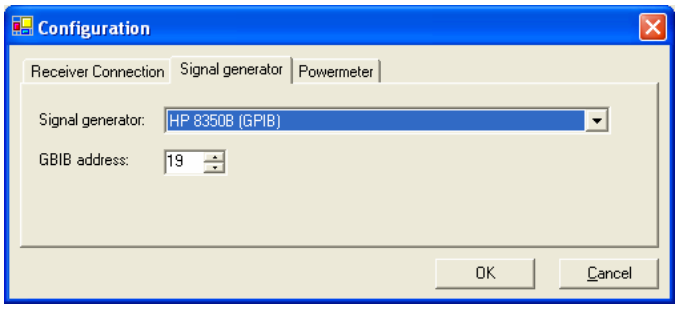


Fig. 6. Windows form for selection and confguration instruments

In the shown software, it is used local .mdb Access format file. Measured data in the antenna radiation pattern measuring method are stored into two database tables MeasAngle and MeasAngleResults. In MeasAngle table are saved general data about measuring like frequency range and step, time of start measuring, comment, etc. Unique ID of measuring will be got from field IdMeasAngle in this table after the record is saved. The second table MeasAngleResults is used for storage measured data with ID of measuring.

Software module for executing antenna radiation pattern measuring with integrated receivers with digital outputs is realized to use in advanced defined states. In Fig. 7. are shown diagram with transitions between states. For simplicity, in this diagram are not shown transitions to error states.

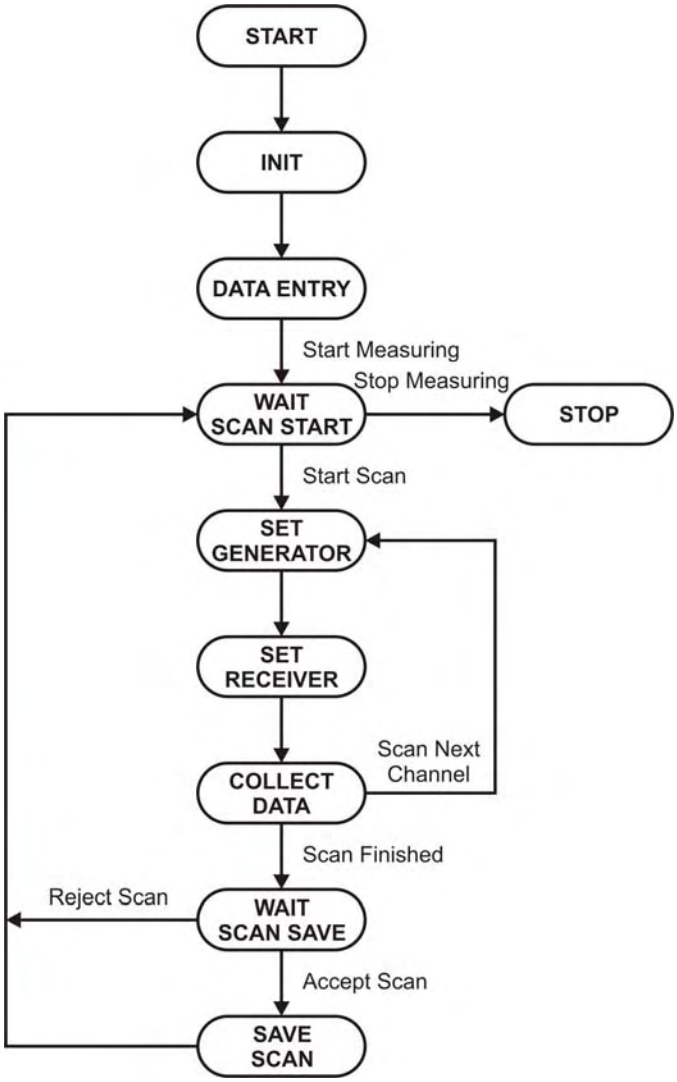


Fig. 7. Diagram states of the software modul for antenna radiation pattern measuring with integrated receivers wth digital outputs

When the software module is started initialization of the communications with instruments and receiver and opening database will be done. In the state DATA ENTRY, user can enter measuring parameters (frequency range and step, comment, etc.). When user press button to start with measuring, application is in the state WAIT SCAN START where application wait for user to setup antenna positioner on required angle. When user press button to continue measuring, the application in the state SET GENERATOR setup signal generator on needed frequency, and in the state SET RECEIVER setups receiver on needed frequency channel. In the state COLLECT DATA collects data from the receiver. Signal generator frequency and receiver channel are incremented while boundary frequency is not reached. Then the application goes to the state WAIT SCAN SAVE, where wait for user to accept this measuring which is shown in the form and save records to database. If measuring data are accepted, the next state is WAIT SCAN when user needs to change antenna angle.

On Fig. 8. are shown windows forms for antenna radiation pattern measuring.

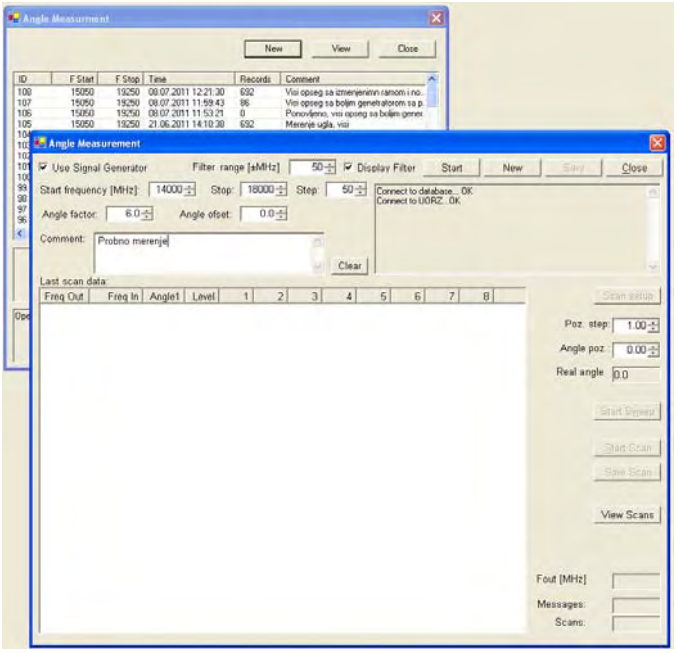


Fig. 8. Windows forms for antenna radiation pattern measuring

VI. Measured Data Presentation

Measured data from the database can be presented in tabular and graphical forms. Because there are large number of measured data, tabular view is not always suitable. The application supports export measured data to Excel and CSV file formats for analysis or representing in other programs.

Graphical views of the measured results are more convenient. Orthogonal, polar and 3D graphical views are supported for representing antenna radiation patterns. In Fig. 9. is shown 3D radiation pattern diagram for horn antenna which are measured on 46 different frequencies in ranges from 15 to 19.5 GHz with step from 100 MHz and with angle step from 22.5°. On this diagram is easy to observe some irregularities in antenna radiation patterns on some frequencies.

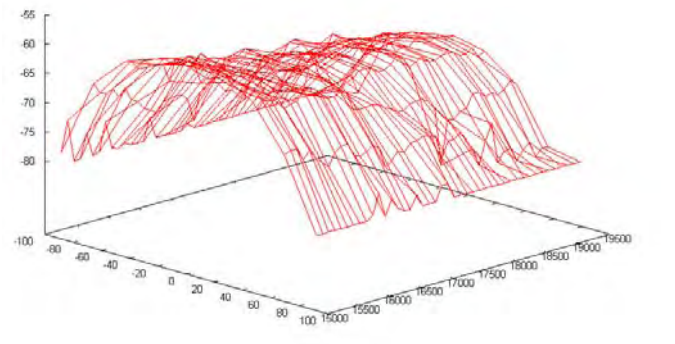


Fig. 9. Example of 3D antenna radiation pattern digram which depends from frequency

VII. Conclusion

Described method and software allow to make automated measuring of antenna radiation patterns on antenna with integrated receivers on desired number of frequencies. The software helps to make measurement for shorten time and to present on convenient way large amount of measured data. Different presentation types of he measured data can help to check for possible problems in realization of the tested antennas. The software also can be upgraded for supports for a new instruments types and models, for different kinds of integrated receivers and to support new measuring methods.

Acknowledgement

This work was partially supported by the Ministry of Education and Science of the Republic of Serbia under Grant TR-32024.

References

[1] E. Kruft, P. Packebush, “Calibration ensures accuracy,” IEEE Spectrum, vol. 37, no. 11, pp. 66-69, Nov. 2000.
[2] Joseph P. Keithley, “Customization is key to future instrumentation”, IEEE Spectrum, vol. 36, no. 1, pp. 94-95, Jan. 1999.
[3] P. Ponce-Cruz, F. D. Ramírez-Figueroa, *Intelligent Control Systems with LabVIEW*, London: Springer, 2010.
[4] www.ni.com
[5] www.mathworks.com
[6] www.agilent.com
[7] M. Pirola, V. Teppati, V. Camarchia, “Microwave measurements Part I: Linear Measurements,” IEEE Instrumentation and Measurement Magazine, vol. 10, Issue 2, pp. 14-19, Apr. 2007.
[8] S. Jovanović, P. Manojlović, D. Obradović, N. Mitrović: "Wideband Receiver for Signal Detection in Frequency Range from 15 to 19 GHz", Proceedings of Papers XLV ICEST, Vol. 1, Ohrid, FYR Macedonia, 2010, pp. 545-548.
[9] D. Obradović, S. Jovanović, N. Mitrović, Ž. Gajić: "Realization of IF signal processing for millimeter band signal detector system", 54. ETRAN, Conference Proceedings, pp. MT2.4-1-4, Donji Milanovac, Serbia, 2010.
[10] Jovanović Siniša, Manojlović Predrag, Tasić Siniša, Obradović Dragan, Mitrović Nemanja, "Wideband receiver for signal detection in frequency range 11-15 GHz", TELFOR 2010, Conference Proceedings, Belgrade, Serbia, 2010.
[11] T. Williams, C. Kelley, *Gnuplot 4.4 - An Interactive Plotting Program*, 2010., Available: www.gnuplot.info
[12] Dragan Obradović, Predrag Manojlović, Nemanja Mitrović, Siniša Jovanović, and Željko Bojović, "Software for automated measuring of the wideband microwave receivers", TELFOR 2010, Conference Proceedings, Belgrade, Serbia, 2010.
[13] J. Albahari, B. Albahari, *C# 3.0 in a Nutshell. 3rd ed.*, Sebastopol: O'Reilly Media, 2007.

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

Tatjana Cvetković[1], Vesna Milutinović[1], Nebojša Dončov[2], Jugoslav Joković[2] and Bratislav Milovanović[2]

***Abstract* – Numerical model of rectangular enclosure with a real receiving antenna is considered in the paper in order to investigate an impact of antenna presence on electric shielding effectiveness. Dipole antenna, often used in experimental setup to measure the level of electromagnetic field at some characteristic points in enclosure, is described by using a compact wire model implemented into transmission-line matrix (TLM) method. Using the proposed model, impact of receiving antenna on shielding effectiveness is illustrated on several examples of enclosure with apertures and compared with the corresponding numerical and circuital approaches in which the antenna presence is neglected.**

***Keywords* – Enclosure, shielding effectiveness, dipole antenna, wire TLM model.**

I. INTRODUCTION

The most frequently used types of electronic equipment for electromagnetic (EM) protection, are metallic enclosures. The shielding enclosure performances are quantified by shielding effectiveness (SE), defined as the ratio of field strength in the presence and absence of the enclosure. Integral parts of the shielding enclosure are apertures of various forms, intended for heat dissipation, control panels, outgoing or incoming cable penetration, airing or other purposes. The shielding enclosure with apertures should be designed based on the analysis of the EM coupling mechanism through apertures, in order to minimize the EM interference (EMI) and susceptibility risk due to inevitable discontinuities.

There are many techniques used to calculate SE, from analytical methods to numerical simulations. Analytical formulations [1] are a quick tool based on the 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]. Simple solution based on circuital approach has been proposed in [3] but with some limitations in terms of location of apertures, angle and polarization of incident plane wave and TE/TM modes that can take into account. Circuital approach has been modified in [4] to allow for considering oblique incidence and polarization and not to be limited by the location of aperture with respect of plane wave propagation direction. In that model, the SE of the enclosure with apertures on multiple sides can be simply calculated by superposition of one dimensional result, and the problem of SE considering the field with arbitrary incidence and polarization angle can be solved by vector decomposition.

Differential numerical techniques in the time domain, such as the Finite-Difference Time-Domain (FDTD) method [5] and Transmission Line Matrix (TLM) method [6], owing to their characteristics, have found their application in solving many electromagnetic compatibility (EMC) problems in a wide frequency range. In [7] TLM method has been successfully employed to calculate SE of shielding enclosure with apertures over a broad frequency band (up to 3 GHz). In parallel with this research, authors of this paper have conducted their own analysis of influence of various factors, such as aperture patterns, their dimensions, number and orientation with the respect of enclosure walls or plane wave propagation direction, on shielding efficiency of enclosure and the results have been presented in [8] and [9]. In addition, impact of plane wave excitation parameters of shielding properties of enclosure with multiple apertures has been considered by the authors in [10] and [11]. Again, TLM method was used in [8]-[11] to numerically study these various effects over a frequency range up to 2 GHz.

In practice, when EMC measurements are performed, to experimentally characterize the SE of enclosure, small dipole receiving antenna, is located inside the enclosure. Such antenna is used to measure the level of EM field, coming from external interference source through apertures, at the points in the enclosure in order to perform the SE calculation. Receiving antenna of finite dimensions could significantly affect the EM field distribution inside the enclosure [12] and thus affect the results for SE. Both either circuital or numerical approaches mentioned above did not take into account the presence of receiving antenna. Therefore, in this paper TLM method enhanced with compact wire model [13] to efficiently describe the dipole antenna, is applied in order to create a numerical model that can used to investigate the impact of receiving antenna on SE of enclosure. This model has been used here to calculate the SE of rectangular enclosure with one or two apertures of rectangular cross-section on the front wall and two adjacent walls, in the frequency range of up to 2GHz. Obtained numerical results illustrate the SE variation due to receiving antenna in comparison with the circuital or numerical case when its presence is neglected.

[1]Tatjana Cvetković is with the Republic Agency for Electronic Communications, Višnjiceva 8, 11000 Belgrade, Serbia, E-mail: tatjana.cvetkovic@ratel.rs.

[1]Vesna Milutinović is with the Republic Agency for Electronic Communications, Višnjiceva 8, 11000 Belgrade, Serbia, E-mail: vesna.milutinovic@ratel.rs.

[2]Nebojša Dončov is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: nebojsa.doncov@elfak.ni.ac.rs.

[2]Jugoslav Joković is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: nebojsa.doncov@elfak.ni.ac.rs.

[2]Bratislav Milovanović is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: bratislav.milovanovic@elfak.ni.ac.rs.

II. COMPACT WIRE TLM MODEL

The modelling of thin wires is a permanent difficulty for all differential numerical modelling methods. The simplest solution is to model wires by using short-circuit nodes or shorted link-lines adjacent to the wire surface [6]. This, however, is rarely a practical proposition, as computational resource limitation and geometrical disparity between the modelled space and fine features in the EMC problems result in a rather crude rectangular shape model of the wire. This can cause shift of the resonances by 5-10% to lower frequencies, a problem which is referred to as 'resonance error'.

Sophisticated solution in the form of compact wire model or wire node, which can allow for accurate modelling of wires with a considerably smaller diameter than the node size, has been introduced in TLM method [13]. It use special wire network embedded within TLM nodes (Fig.1) to model signal propagation along the wires, while allowing for interaction with the EM field. In order to achieve consistency with the rest of the TLM model, the each segment of wire network within one TLM node is formed by using additional link and stub lines (Fig.2) and it is coupled with the electric field component parallel to its direction. The parameters of link and stub lines are chosen to model the per-unit length wire capacitance and inductance, while at the same time maintaining synchronism with the rest of the transmission line network.

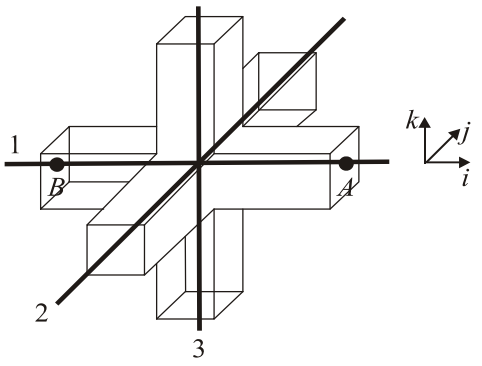


Fig.1. Wire segments embedded within the symmetrical condensed TLM node

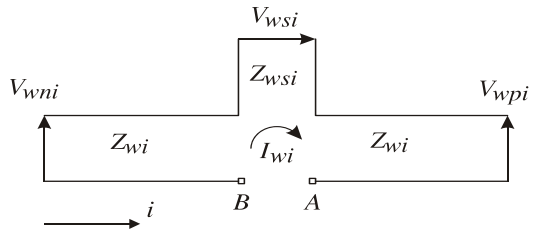


Fig.2. Link and stub lines network for straight wire segment running in *i* direction

The single column of TLM nodes, through which wire conductor segments pass, can be used to approximately form the fictitious cylinder which represents capacitance and inductance of wire per unit length. Its effective diameter, different for capacitance and inductance, can be expressed as a product of factors empirically obtained by using known characteristics of TLM network and the mean dimensions of the node cross-section in the direction of wire running [13]. For an example, for the node containing *i*-directed straight wire segment, as depicted in Fig.2, the effective diameters of fictitious cylinder for wire capacitance and inductance can be defined as:

$$d_{Ci} = 2k_{Ci}\Delta i_c \quad (1)$$

$$d_{Li} = 2k_{Li}\Delta i_c \quad (2)$$

respectively, where Δi_c represents mean cross-section dimensions of TLM node in *i* direction, $\Delta i_c=(\Delta j + \Delta k)/2$. Empirically found factors k_{Ci} and k_{Li} for the wire located in free space are:

$$k_{Ci} = 0.0511k_i^2 + 0.0194k_i + 0.617 \quad (3)$$

$$k_{Li} = 0.34 \quad (4)$$

and for the wire above the ground:

$$k_{Ci} = 0.0223k_i^2 + 0.024k_i + 0.606 \quad (5)$$

$$k_{Li} = 0.347 \quad (6)$$

Parameter k_i depends on time- and space-step discretization and EM properties of medium represented by TLM node and it can be calculated as:

$$k_i = 2\Delta t/(\sqrt{\varepsilon\mu}\Delta i_c) \quad (7)$$

Once the effective diameters are known, the per-unit length wire capacitance and inductance can be calculated as:

$$C'_{wi} = 2\pi\varepsilon/\ln(d_{Ci}/d_w) \quad (8)$$

$$L'_{wi} = \mu\ln(d_{Li}/d_w)/2 \quad (9)$$

where d_w is a real wire diameter. Wire per-unit length capacitance is then modelled by the link line of characteristic impedance Z_{wi}:

$$Z_{wi} = \frac{\Delta t}{\Delta i\, C'_{wi}} \quad (10)$$

while the wire per-unit length inductance is modelled by short-circuit stub of characteristic impedance Z_{wsi}:

$$Z_{wsi} = L'_{wi}\frac{\Delta i}{\Delta t} - Z_{wi} \quad (11)$$

III. RESULTS

For analysis purposes we used the rectangular enclosure with the following dimensions: l_x = 300 mm, l_y = 300 mm and l_z = 120 mm. The front wall of the enclosure was made of the 3 mm conductive material with rectangular apertures having different dimensions and numbers: one 50 mm x 10 mm aperture, one 50 mm x 30 mm aperture and two 50 mm x 10 mm apertures. We have calculated the enclosure SE for a plane wave of normal incidence to the front wall and vertical (z) electric polarization as excitation, as shown in Fig.3.

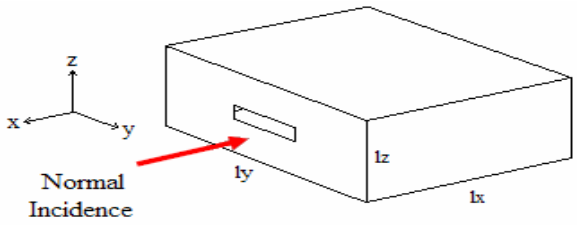


Fig.3. Rectangular enclosure with a rectangular aperture

In the first case, for calculating SE using the numerical model, we assumed that the presence of the receiving antenna could be neglected (empty enclosure). The SE was calculated at the point of the enclosure (145 mm, 150 mm, 60 mm) for all specified patterns of rectangular apertures, by using the conventional TLM method (Fig.4).

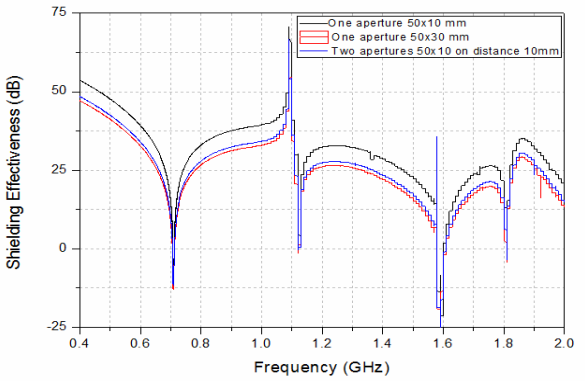


Fig.4. Numerical results for the SE of enclosure without antenna at the point 145 mm x 150 mm x 60 mm

As can be seen from Fig.4, the shape of the SE curves for all considered patterns of apertures remains the same, including the values of resonant frequencies. This indicates that the patterns and the number of apertures only affect the level of attenuation to which EM field propagating through apertures is exposed. As expected, the level of SE decreases with the increase of area covered by apertures. However, as it will be shown next, these results deviate to some extent from the case when the receiving antenna is included in the model.

For calculating SE when the receiving dipole antenna is included in the numerical model of the enclosure by using the compact wire model described in section II, the antenna is modelled as z-directed 80 mm long wire having the diameter of 1.6 mm. Its position within the enclosure is defined by points (145 mm, 150 mm, 20 mm) and (145 mm, 150 mm, 100 mm). The numerical results for the SE of enclosure, shown in Fig.5, are calculated at the centre point of the receiving dipole antenna (145 mm, 150 mm and 60 mm). In comparison with the case when enclosure is empty, it can be seen that the antenna presence significantly decreases the SE of enclosure in the whole frequency range.

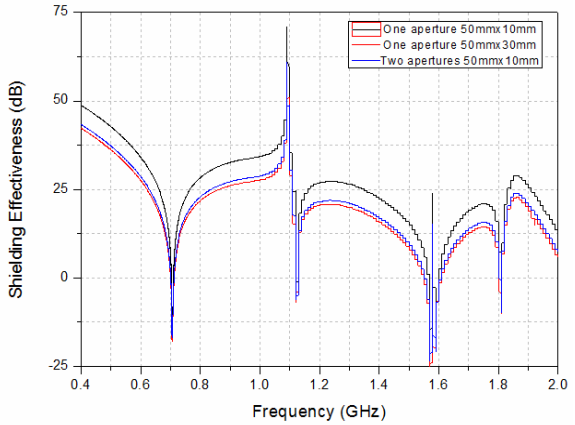


Fig.5. Numerical results for the SE of enclosure with antenna at the point 145 mm x 150 mm x 60 mm

Next, we presented numerical results obtained when the receiving antenna was not included in the model, when the receiving antenna was included in the model, and the modified circuital model results described in [4]. The calculated SE of the receiving dipole antenna at the centre point (145 mm, 150 mm and 60 mm) and of the empty enclosure at the same point, are shown in Figs.6, 7 and 8, for all considered aperture patterns. In the same figures the SE values obtained by circuital/analytical model are also shown.

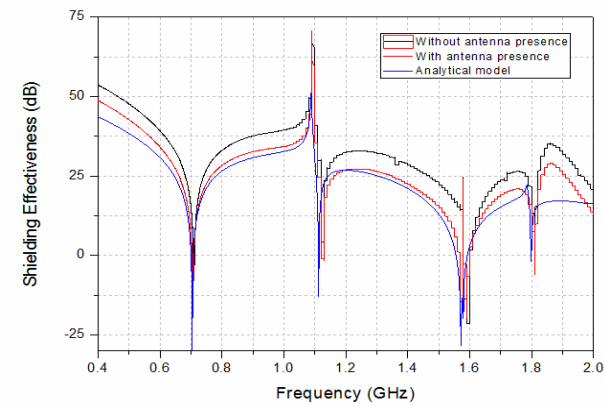


Fig.6. Results for SE of enclosure, with one aperture of dimension 50 mm x 10 mm, at the point 145 mm x 150 mm x 60 mm

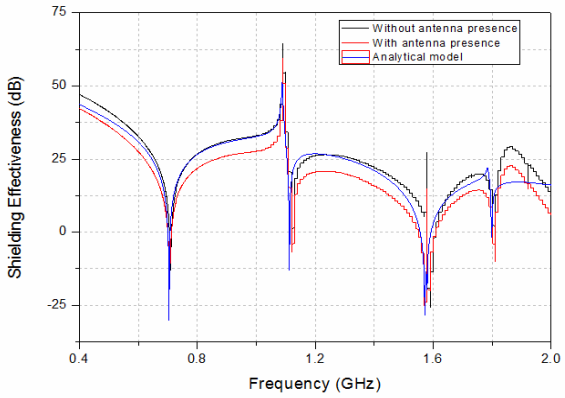


Fig.7. Results for SE of enclosure, with one aperture of dimension 50 mm x 30 mm, at the point 145 mm x 150 mm x 60 mm

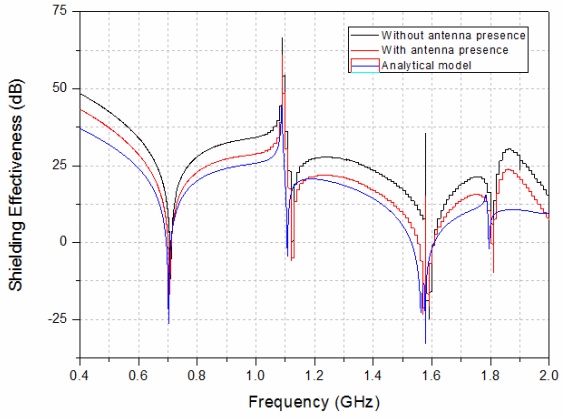


Fig.8. Results for SE of enclosure, with two apertures of dimension 50 mm x 10 mm, at the point 145 mm x 150 mm x 60 mm

It can be seen that the presence of the receiving antenna significantly reduces the shielding efficiency of the enclosure, as the SE level is always lower in comparison with the case when the enclosure is empty. Although the presence of antenna is neglected in the circuital approach, the SE values of enclosures with one and two 50 mm x 10 mm apertures are lower than the numerical results obtained. For the enclosure with one 50 mm x 30 mm aperture dimensions, the results obtained by circuital approach practically coincide with the numerical results obtained for an empty enclosure. Also, there is a tendency to slightly shift the resonant frequencies of enclosure.

Finally, the rectangular enclosure with apertures on multiple sides for oblique incident plane wave is considered. Numerical results for SE at the point 145 mm x 150 mm x 60 mm obtained by using TLM method with and without compact wire model and results obtained using circuital approach at the same point are shown in Fig.9. Two groups of two rectangular apertures of dimension 50 mm x 10 mm are placed on the adjacent enclosure walls while an obliquely incident wave with the azimuth angle 60°, elevation angle 90°, and polarization angle 30° is used as excitation. It can be seen that the presence of the receiving antenna significantly reduces the SE of the enclosure and that results obtained using circuital approach are lower than the numerical results.

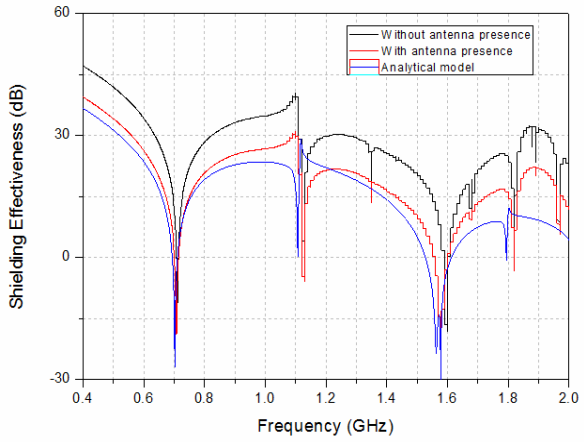


Fig.9. Results for SE of enclosure, with two groups of two rectangular apertures of dimension 50 mm x 10 mm placed on the adjacent enclosure walls, at the point 145 mm x 150 mm x 60 mm

IV. Conclusion

The TLM method, enhanced by the compact wire model, has been used here to generate a numerical model that can provide a tool for analysing the impact of the receiving dipole antenna on the enclosure electric shielding effectiveness. The given example confirm that the antenna presence affects the EM field distribution inside the enclosure and thus affects the SE level results, as well as the location of resonant frequencies. The results of circuital approach do not coincide in all cases with the numerical simulation results, but this fast analytical method can be used for approximate calculation of SE of enclosures. Future research will comprise numerical estimation of antenna impact on SE for given antenna dimensions and also modification of the circuital model in order to include the presence of the receiving antenna, since it is a real element in the measurement process.

Acknowledgement

This work has been partially supported by the Serbian Ministry for Education and Science within the project III 44009.

References

[1] H. H. Park and H. J. Eom, “Electromagnetic penetration into a rectangular cavity with multiple rectangular apertures in a conducting plane” IEEE Trans. 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, May 1978.

[3] 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 effectiveness of enclosures with apertures” IEEE Trans. Electromagn. Compat., vol. 40, no. 3, pp. 240–248, Aug. 1998.

[4] J. Shim, D.G. Kam, J.H. Kwon, J. Kim, ”Circuital Modeling and Measurement of Shielding Effectiveness against Oblique Incident Plane Wave on Apertures in Multiple Sides of Rectangular Enclosure” IEEE Trans. Electromagn. Compat., vol. 52, no. 3, pp. 566–577, Aug. 2010.

[5] K.S. Kunz, R. J. Luebbers, *The Finite Difference Time Domain Method for Electromagnetics*, CRC Press, Boca Raton, FL, 1993.

[6] C. Christopoulos, *The Transmission-Line Modelling (TLM) Method*, IEEE Press in association with Oxford University Press, Piscataway, NJ, 1995.

[7] B.L. Nie, P.A. Du, Y.T. Yu, Z. 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.

[8] B. Milovanović, N. Dončov, V. Milutinović, T. Cvetković, “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.

[9] V. Milutinović, T. Cvetković, N. Dončov, B. Milovanović, “Analysis of the shielding effectiveness of enclosure with multiple circular apertures on adjacent walls”, in Proc. Int. Conf. on Information, Communication and Energy Systems and Technologies – ICEST 2011, Niš, Serbia, vol. 3, pp.685-688, 2011.

[10] T. Cvetković, V. Milutinović, N. Dončov, B. Milovanović, “Analysis of the influence of polarization and direction of propagation of a incident plane wave on the effectiveness of rectangular enclosures with apertures”, in Proc. Int. Scientific-Professional Symp. INFOTEH, Jahorina, vol. 10, Ref.B-I-6, pp.90-94, 2011.

[11] V. Milutinovic, T. Cvetkovic, N. Doncov, B. Milovanovic, “Analysis of enclosure shielding properties dependence on aperture spacing and excitation parameters”, in Proc. IEEE Conf. TELSIKS , Niš, Serbia, vol.2, pp. 521-524.

[12] J. Joković, B. Milovanović, N. Dončov, “Numerical Model of Transmission Procedure in a Cylindrical Metallic Cavity Compared with Measured Results”, International Journal of RF and Microwave Computer-Aided Engineering, vol. 18, no. 4, pp. 295-302, July 2008.

[13] A.J. Wlodarczyk, V. Trenkic, R. Scaramuzza, C. Christopoulos, “A Fully Integrated Multiconductor Model For TLM”, IEEE Transactions on Microwave Theory and Techniques, vol. 46, no. 12, pp. 2431-2437, December 1998.

Reliability of Radio-Relay Systems

Nataša Bogdanović[1], Dejan Blagojević[2] and Dragiša Milovanović[3]

Abstract – **Reliability of radio-relay systems depends on the propagation conditions and the reliability of equipment. The aim of this paper is to analyze transmission parameters such as propagation loss, diffraction fading, attenuation due to atmospheric conditions etc., from the reliability of radio-relay systems point of view. Reliability of radio-relay links, depending on bandwidth channel, digital modulation types and bit rate, is considered using the simulation programme Pathloss. Annual rain and multipath unavailability are analyzed at the frequency of 18 GHz band on two links and four bandwidth channels. These results are particularly important in the design of transmission links over Wireless Internet.**

Keywords – Radio-relay, reliability, rain unavailability, multipath

I. INTRODUCTION

In order to properly plan terrestrial radio-relay line-of-sight systems it is necessary to have accurate propagation parameters or adequate prediction methods. These methods have been developed for the purpose of predicting the most important propagation parameters for radio-relay links [1]. In design of line-of-sights of radio-relay links several propagation parameters must be taken into consideration. These parameters are: diffraction fading due to an obstacle on terrain path, attenuation in the atmosphere, fading due to atmospheric multipath, fading due to multipath arising from the surface reflection, attenuation due to precipitation, variation of the angle-of-launch at the transmitter terminal and the angle-of-arrival at receiver terminal due to refraction, reduction in cross-polarization discrimination in multipath or precipitation conditions and signal distortion due to frequency selective fading and delay during multipath propagation.

The propagation losses on the terrestrial path in the free space consist of losses such as attenuation due to atmospheric gases, diffraction fading due to obstruction or partial obstruction of the path, fading due to multipath, beam spreading and scintillation, attenuation due to variation of the angle-of-arrival/launch, attenuation due to precipitation, dust and storms. The probability of appearance of these events is of great importance in consideration of reliability of line-of-sight radio-relay systems.

All these parameters are calculated in designing of terrestrial radio-relay links to points which are used to provide wireless Internet. From these points wireless Internet is distributed at standardized frequencies to end users.

[1]Nataša Bogdanović is with the School of Higher Technical Professional Education, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: natasa.bogdanovic@vtsnis.edu.rs

[2]Dejan Blagojević is with the School of Higher Technical Professional Education, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: dejan.blagojevic@vtsnis.edu.rs

[3]Dragiša Milovanović is with the Faculty of Electronic Engineering at University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: dragisa.milovanovic@elfak.ni.ac.rs

II. LINK RELIABILITY

Reliability represents systems ability to operate without outage during a certain period of time. Outage is the loss of ability of a device to perform a required function [2]. The reliability of a radio-relay system depends on equipment and connection availability.

An estimation of the equipment availability is used in order to statistically predict when the equipment or radio-relay system will be unavailable due to unintentional malfunctions on radio-relay devices.

Equipment unavailability Un_{opr} is determined with average time needed for repair (Mean Time To Repair) MTTR(h) and the mean time interval between the appearance of two successive errors (Mean Time Between Failures) - MTBF(h) :

$$Un_{opr} = \frac{\text{MTTR}}{\text{MTTR} + \text{MTBF}}. \quad (1)$$

Since, $\text{MTTR} << \text{MTBF}$, the previous expression transforms into:

$$Un_{opr} \cong \frac{\text{MTTR}}{\text{MTBF}}. \quad (2)$$

MTTR and MTBF are stated in the equipment datasheet and the designer of the radio-relay system has no influence on them. Let us consider the influence of connection unavailability on the reliability of the entire system.

The availability objectives, availability ratio (AR) and mean time between outage (Mo), and reciprocal outage intensity (OI) needed for design purposes are given in ref. [3]. The availability objectives applicable to fixed wireless link of length L*link*, can be derived from the values given in [3]. The variables in equations (3) and (4) are given in ref. [3];

$$AR = 1 - \left(B_j \frac{L_{link}}{L_R} + C_j \right) \quad (3)$$

$$Mo = 1/OI = \frac{1}{D_j \frac{L_{link}}{L_R} + E_j} \quad (4)$$

where the value of j is for section of national portion: j=5, for access network, j=6 for short haul, j=7 for long haul. LR: reference length LR = 2500 km. The lower limit of L*link* used to scale the objectives is L*min* = 50 km. National portion: length 30 km in access portion: The length is shorter than L*min* = 50 km, hence the value of L*link* = 50 km has been used. These values correspond to AR of 99.95% (unavailability of 263 min/year), number of events per year OI = 100 and the mean time between unavailability events Mo = 5257 min.

The length is in the range of 50-250 km. These values correspond to AR of 99.96% (unavailability of 210 min/year),

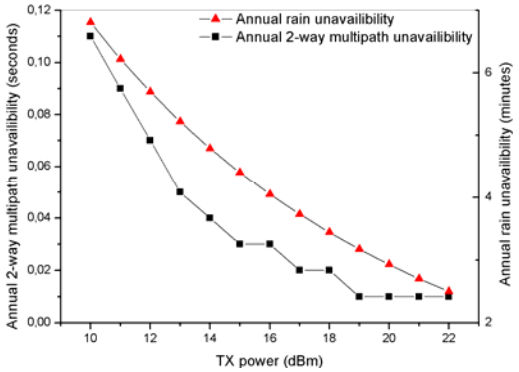


Fig. 1. Unavailability, 7 MHz bandwidth

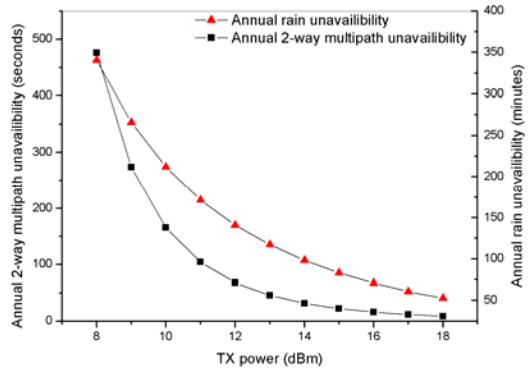


Fig. 2. Unavailability, 56 MHz bandwidth

number of events per year OI = 120 and the mean time between unavailability events Mo = 4381 min.

III. Analysis of the Unavailability

For the purpose of analysis of the influence of radiated power of transmitting antenna on the availability of the link system, two radio-relay routes have been analyzed. The relay routes are used by an internet provider, one is between Niš and Pirot and the other is between Niš and Leskovac. The first route has three hops: Niš-Kamenički vis in the range of 18 GHz, Kamenički vis-Šljivovički vrh in the range of 7 GHz and Šljivovički vrh-Pirot in the range of 18 GHz. The second route has two hops: Niš-Seličevica in the range 18 GHz and Seličevica-Leskovac in the range of 7 GHz. The paper reviews Šljivovički vrh-Pirot and Niš-Seličevica hops only, both in the range of 18 GHz.

The analysis has been performed with Pathloss 5 software packet [4], while NASA SRTM3 digital base has been used in terrain description [5]. The algorithm given in ITU-R P.350-9/14 [1] recommendation has been implemented in Pathloss 5 software. Selective fading is calculated based on equipment signature for a given radio-relay equipment made by Ceragon [6]. The signature parameter definitions and specifications of how to obtain the signature are given in recommendation ITU-R F.1093 [7]. Rain calculation method is ITU-R P.530, and specific attenuation regression coefficients are given in ITU-R P.838 [8].

The results of the analysis are displayed in Figs. 1 and 2, which display the dependencies of annual 2-way multipath and rain unavailability on the transmitter link TX power. In Fig. 1 the above mentioned dependency for 7 MHz bandwidth and QPSK modulation, which correspond to 10 Mbps bit rate can be seen, while fig. 2 displays the same dependencies, but for 56 MHz bandwidth and 256QAM modulation, which correspond to 374 Mbps bit rate.

TABLE I

UNAVAILABILITY VS BANDWIDTH AND BIT RATE

UNAVAILABILITY HOP: NIŠ - SELIČEVICA, 18 GHZ BAND **7 MHZ** BANDWIDTH			
	Mbps	Annual 2-way multipath (sec)	Annual rain (min)
QPSK	10	0.02	3.44
8PSK	15	0.03	4.22
16QAM	20	0.06	5.45
32QAM	25	0.10	6.51
64QAM	31	0.17	8.18
128QAM	36	0.31	10.38
256QAM	42	0.61	14.02
	46	0.97	17.34
14 MHz bandwidth			
	Mbps	Annual 2-way multipath (sec)	Annual rain (min)
QPSK	22	0.03	4.22
8PSK	31	0.06	5.22
16QAM	41	0.12	7.12
32QAM	54	0.17	8.18
64QAM	62	0.31	10.38
128QAM	75	0.61	14.02
256QAM	88	1.22	19.35
	97	1.72	22.93
28 MHz bandwidth			
	Mbps	Annual 2-way multipath (sec)	Annual rain (min)
QPSK	41	0.06	5.45
8PSK	58	0.09	6.22
16QAM	82	0.18	8.18
32QAM	112	0.39	11.45
64QAM	138	0.77	15.57
128QAM	170	1.72	22.93
256QAM	180	2.43	27.37
	200	4.42	37.44
56 MHz bandwidth			
	Mbps	Annual 2 way multipath (sec)	Annual rain (min)
QPSK	82	0.11	6.81
8PSK	123	0.2	8.57
16QAM	165	0.32	10.38
32QAM	222	0.71	14.77
64QAM	275	1.4	20.46
128QAM	330	2.76	29.09
256QAM	374	5.05	39.98
	414	8.71	52.64

It can be concluded from 1 and 2 that the selective fading is more pronounced at wider channel (56 MHz) and higher bit rate, and that by reducing the transmitter power the multipath unavailability increases drastically, as is the case with rain unavailability.

Table I displays a systematic analysis of annual 2-way multipath and rain unavailability dependence on bit rate for bandwidths of 7, 14, 28 and 56 MHz and QPSK, 8PSK, 16QAM, 32QAM, 64QAM, 128QAM and 256QAM modulations.

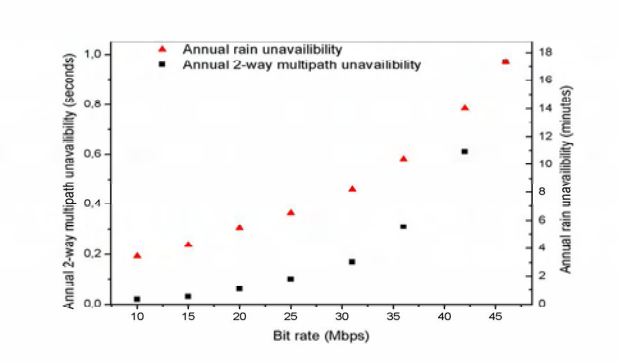


Fig. 3a. Unavailability, 7 MHz bandwidth

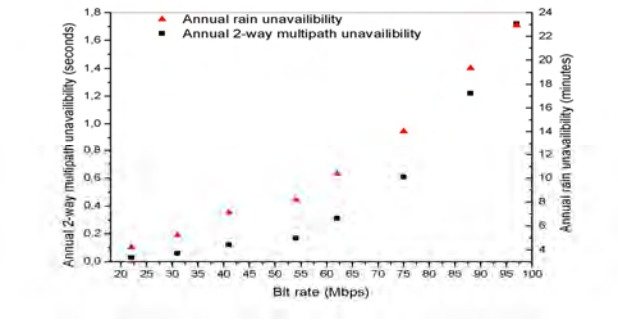


Fig. 3b. Unavailability, 14 MHz bandwidth

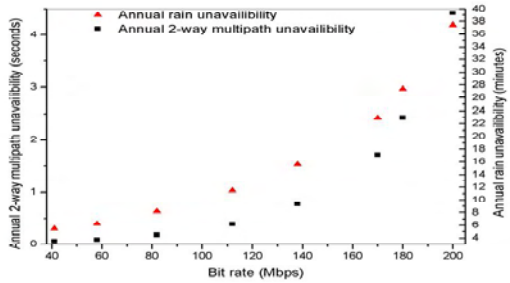


Fig. 3c. Unavailability, 28 MHz bandwidth

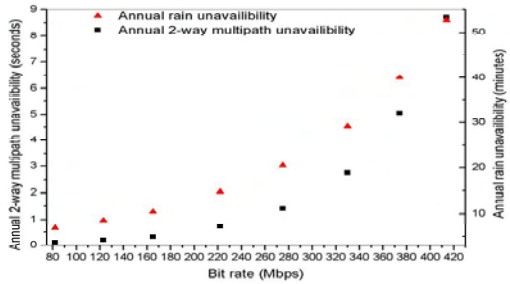


Fig. 3d. Unavailability, 56 MHz bandwidth

TABLE II
UNAVAILABILITY VS BANDWIDTH AND BIT RATE

UNAVAILABILITY HOP: ŠLJIVOVIČKI VRH – PIROT, 18 GHZ BAND **7 MHZ** BANDWIDTH			
	Mbps	Annual 2 way multipath (sec)	Annual rain (min)
QPSK	10	8.59	19.11
8PSK	15	15.27	23.72
16QAM	20	31.22	31.69
32QAM	25	50.14	39.16
64QAM	31	98.92	52.19
128QAM	36	212.57	74.67
256QAM	42	925.26	133.16
	46	885.8	160.26
14 MHz bandwidth			
	Mbps	Annual 2 way multipath (sec)	Annual rain (min)
QPSK	22	7.05	22.92
8PSK	31	16.32	28.93
16QAM	41	36.16	41.03
32QAM	54	54.1	48.67
64QAM	62	81.71	63.84
128QAM	75	202.25	97.15
256QAM	88	602.47	162.18
	97	1607.7	238.05
28 MHz bandwidth			
	Mbps	Annual 2 way multipath (sec)	Annual rain (min)
QPSK	41	8.34	29.21
8PSK	58	12.93	33.62
16QAM	82	41.28	47.21
32QAM	112	82.12	70.23
64QAM	138	224.9	108.86
128QAM	170	518.39	182.21
256QAM	180	890	253.19
	200	2451.09	407.53
56 MHz bandwidth			
	Mbps	Annual 2 way multipath (sec)	Annual rain (min)
QPSK	82	9.6	36.11
8PSK	123	17.06	45.98
16QAM	165	27.32	56.69
32QAM	222	60.4	84.75
64QAM	275	159.2	130.25
128QAM	330	461.47	217.27
256QAM	374	1430.19	371.77
	414	4719.29	643.25

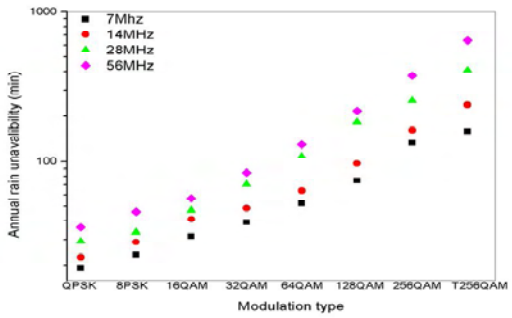


Fig. 4a. Unavailability vs. modulation

The analysis has been performed on Niš-Seličevica hop at 19 GHz frequency with vertical polarization. Hop length was 10.76 km. The results from Table I are displayed in Figs. 3a-d, and it can be noticed from them that there is a high dependency of unavailability on the bandwidth and modulation type, i.e. bit rate. These dependencies must be taken into consideration in order to enable reliable operation when links capacity is designed, since it is obvious that radio-relay availability strongly affects the operation routine. This is particularly the case with QAM modulation at high bit rates.

Table II displays the results of the analysis of multipath and rain unavailability for Šljivovički vrh-Pirot hop of length 18,76 km. The calculations have been performed at 19 GHz frequency with vertical polarization. Multipath and rain unavailability dependence on bandwidth, modulation type and

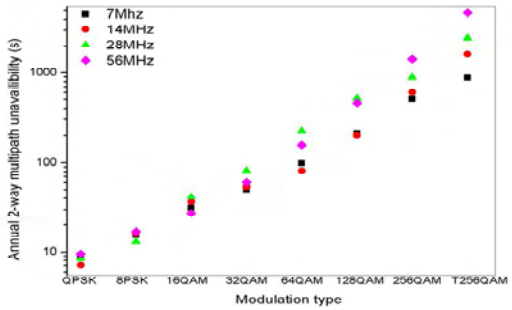


Fig. 4b. Unavailability vs. modulation

bit rate have been analyzed. Figures 4a and 4b show plots of unavailability dependence on modulation type for different bandwidths. It can be seen that by increasing bandwidths multipath and rain unavailability also increase. Multipath unavailability increase is significant at 256QAM modulation, i.e. at bit rates of several hundred Mbps, where unavailability is three orders of the magnitude higher than at QPSK modulation. As a consequence of this problem, special care must be taken when designing link connections of large capacities. The analysis of unavailability of both hops considered in this paper has shown the same dependence on bandwidth, modulation type and bit rate. This means that QPSK modulation has a lesser multipath and rain unavailability for the same bit rate, even though it uses a wider channel. This fact has to be taken into proper consideration when radio-relay connections of high reliability are designed.

IV. Conclusion

The reliability of radio-relay system due to the unavailability of link connection as a function of emitting antenna TX power, bit rate, bandwidth and modulation type has been analyzed, in purpose of study data transmission. Two link connections in the range of 18 GHz with vertical polarization have been discussed. The calculations were performed based on the algorithm defined in ITU-R recommendations with the use of a commercial software packet, while the characteristics of radio-relay equipment were given by manufacturer Ceragon for each bandwidth and modulation. During calculations of selective fading equipment signature has been used in order to determine precisely the 2-way multipath unavailability. The dependence of unavailability on the emitter TX power has been analyzed and it was shown that the unavailability drastically increases when TX power is reduced. Based on tables and figures displayed in the paper a large dependence of rain and multipath unavailability on the bandwidth can be noticed. The unavailability is significantly depended on the modulation type. The unavailability at QPSK modulation for the same bit rate and different bandwidths is considerably lesser than is the case with other types of modulation. Unavailability increases by three orders of the magnitude at 256QAM modulation in comparison to QPSK modulation. In conclusion, based on presented tables and figures, optimal equipment, which satisfies international standards, for a specific radio-relay internet connection can be designed. This can be useful for internet providing to the distance places.

References

[1] ITU-R P.530-14: Propagation data and prediction methods required for the design of terrestrial line-of-sight systems.
[2] R. M. Ramović, *Pouzdanost sistema – Elektronskih, telekomunikacionih i informacionih*, Beograd, Katedra za Mikroelektroniku i tehničku fiziku EF, 2005.
[3] ITU-R F.1703: Availability objectives for real digital fixed wireless links used in 27 500 km hypothetical reference paths and connections.
[4] www.pathloss.com
[5] www2.jpl.nasa.gov/srtm/
[6] www.ceragon.com
[7] ITU-R F.1093-2: Effects of multipath propagation on the design and operation of line-of-sight digital fixed wireless systems.
[8] Rec. ITU-R P.838-3: - Specific attenuation model for rain for use in prediction methods.

New Architectural Solutions to Improve the CATV System Performances

Lidia Jordanova[1], Dobri Dobrev[2], Kalin Dimitrov[3]

Abstract – **In this paper several conceptions for building the architecture of CATV systems with improved parameters are presented. They are based on the application of technologies DWDM, frequency stacking, PON, FTTH in order to increase the bandwidth efficiency of both the downstream and upstream paths, to reduce the existing asymmetry between them and to improve the quality of services. Criteria to choose the appropriate components (such as laser transmitters, DWDM multiplexers, optic amplifiers etc.) are given for each of the suggested architectures.**

Keywords – **HFC CATV, broadcast and targeted services, DWDM, upstream path, CWDM, PON, FTTH.**

I. INTRODUCTION

The new generation of CATV systems are the hybrid fiber/coax (HFC) distribution systems that consist of optical rings with additional hubs included along the rings. The signals are conveyed from the hubs to the optical nodes over the optical fibers. In the optical nodes the optical signals are transformed into electrical ones. After that the signals are distributed to the subscribers by a coaxial distribution system. Hence, previously built coaxial distribution systems were combined through the usage of optical rings and in this way the subscriber service was localized in one headend .

Cable distribution networks are bi-directional: that makes it possible for additional services (such as Internet access, VoD, VoIP etc.) to be provided to the subscribers. Two-way transmission of high-speed interactive services is performed by Cable Modem Terminal System (CMTS) that is located in the headend or the hub. Cable Modem or Set-Top-Box is used in order to receive the data packets addressed to the subscriber and to transmit the data to the CMTS.

Signals transmission over the cable network of a CATV system worsens the quality of service (QoS) due to noise and distortions inherent to the active devices in the HFC system such as laser transmitters, optical receivers, optical and RF amplifiers. The level of both noise and unwanted spurious signals depends on the parameters of the HFC network components, the dynamic range of RF signals, number of channels, optical modulation depth etc. Noise and distortions are mainly due to RF amplifiers used to compensate attenuation in the coaxial network. Hence, designers should aim at shortening the coaxial part of HFC networks.

The extension of the territorial range of CATV systems and the services package available, and the increasing of the number of their users demands new architectural solutions for building the cable distribution network. The aim is to increase the system's throughput and to improve the quality of the provided multimedia information.

Today, DWDM systems are being deployed to provide network segmentation and increased bandwidth. Additionally multiplexing in the RF domain is also being used in the upstream passband to increase bandwidth efficiency [1-3]. The best solution for reducing noise and signal distortion in the amplifiers used to compensate the attenuation in the cable is to move towards a fully passive cable distribution networks such as fiber to the curb (FTTC) and fiber to the home (FTTH) [4-6]. This paper presents architectures of CATV systems, based on the combination of these technologies.

II. APPLICATION OF DWDM FOR BUILDING THE DOWNSTREAM PATH OF HFC CATV SYSTEM

CATV systems differ by using RF carriers to transmit the information signals. Two frequency bands are provided for signal transmission from the headend to the subscribers: 112 MHz to 550 MHz (for analog video broadcasting) and 550 MHz to 862 MHz (for narrow casting services – data, voice and digital video). Analog video signals are transmitted by using VSB-AM while QAM methods (usually 256-QAM) are mainly used to transmit digital video programs and data. The system reverse paths make use of the 5 MHz to 65 MHz frequency band and subscribers' signals are transmitted by using QPSK or 16-QAM methods.

The RF signals are transferred over the optic fiber by means of optic carriers whose wavelength may be 1310 nm or 1550 nm while with DWDM the wavelengths can be chosen from the wave range recommended by ITU (from 1530 to 1565 nm). The step of the ITU grid is 0.8 nm but in the CATV systems it is selected greater (usually 1.6 nm). This is done to avoid the appearance of nonlinear distortions due to four-wave mixing (FWM).

In Fig. 1 and Fig. 2 two concepts for building the forward path of the CATV system are given. A specific feature of these architectures is that they implement DWDM technology to transmit interactive downstream channels from headend to hub over one optical fiber. To transmit 256-QAM signals for the interactive subscribers' service n wavelengths ($\lambda_1 ... \lambda_n$) selected from the ITU grid are used. The RF signals of the analog TV programs (VSB-AM signals) are transmitted over a separate fiber and to this end they modulate an optical carrier of wavelength λ_0 = 1550 nm.

[1]Lidia Jordanova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: jordanova@tu-sofia.bg

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

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

The analog transmitter and the ITU transmitters in the headend can be regarded as externally modulated sources that comprise a distributed-feedback (DFB) laser coupled to a Mach-Zehnder modulator. Outputs from the ITU transmitters are multiplexed and transported to the hub. To compensate for the loss within the optical channel erbium-doped fiber amplifier (EDFA) is used.

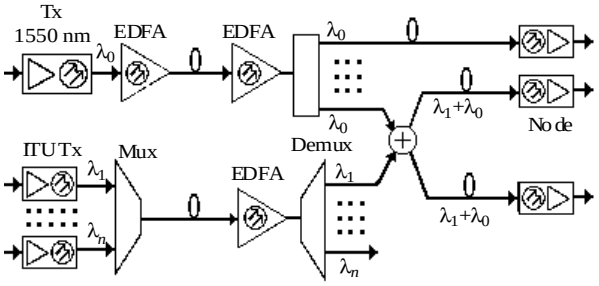


Fig. 1. Forward path configuration 1

The first DWDM architecture has been developed for the 1550-nm range where signals coming over the common-access and the interactive channels are combined at the hub in the optic range. The analog transmitter output is optically amplified to a saturated level of about 17 dBm and transmitted through a shared fiber to the hub, amplified again and split into a number of outputs that matches the number of targeted-services wavelengths. After splitting, the analog signal is combined with the QAM wavelengths and that combination is again split to serve the number of optical nodes for which the given wavelength is targeted. There may be multiple nodes targeted per wavelength, especially in the early deployment stages when subscriber take rates are low corresponding to a low bandwidth requirement per node.

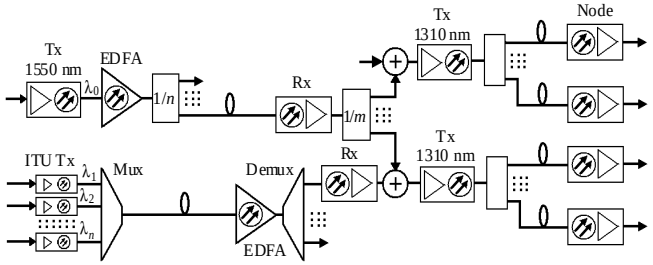


Fig. 2. Forward path configuration 2

The second DWDM architecture refers to cases when an infrastructure of a conventional HFC CATV system in the range of 1310 nm already exists. At the hub, the signals are returned to RF by means of optical receivers. Combined at RF, both broadcast and narrowcast signals drive the 1310 nm lasers. At the optical node (ON) a single detector converts optical signals to RF for distribution into the CATV plant.

Essential to the shared use of fibers is a means by which to combine incoming signals at the transmit end and to separate them at the receiving end. It is possible to use simple wideband splitter/combiners or directional couplers to combine optical signals at the transmit end. The trade-off is that a broadband optical combiner (e.g., an RF combiner), has a insertion loss of about 3.5 – 4 dB per two-way splitting level, whereas a 2-wavelength multiplexer may have a loss of under 2 dB, and a 20-wavelength multiplexer a loss of under 4 dB. At the current state of technology, the retail cost of a broadband combiner is about 10 to 15% of the cost of a 200-GHz-spaced DWDM multiplexer with an equivalent number of ports. Thus, the decision as to whether to use wavelength-specific or broadband combining at the transmit end of a WDM link must be driven by the consideration of the overall link design. At the receiving end, however, there is no alternative to using wavelengthspecific demultiplexers if the signals are to be detected separately.

III. SPLITTING THE OPTICAL NODE IN THE UPSTREAM DIRECTION

The demand for upstream bandwidth is increasing rapidly, driven primarily by the needs of DOCSIS modems. Each voice call requires about 128 kb/s in the upstream direction, while with 50 simultaneous conversations the necessary channel capacity increases up to 6.4 Mb/s. Peer-to-peer applications consume a lot of upstream bandwidth. Furthermore, the large picture, audio, and video files being uploaded today demand higher speeds. This pressure on upstream bandwidth has generated several technologies for splitting the node in the upstream direction, without using more fibers to get data back to the hub or headend. Splitting the node also helps with noise-funneling issues.

The return channel capacity may be increased by using a block conversion system, which takes several return paths and converts all except one to a unique block of frequencies. The blocks are combined and transmitted to the headend using one return optical transmitter. Single conversion is used for economy. For other applications, double conversion might be used if it is necessary to put many blocks close together, but the cost is higher.

Figure 3 illustrates block conversion in a common configuration that uses a 240-MHz return optical transmitter. Up to four return paths may be accommodated. The first is coupled directly to the optical transmitter. The other three are up-converted to other frequencies up to about 240 MHz and combined with the unconverted spectrum to supply signals to a return transmitter. At the headend, the process is reversed, developing four individual 5 to 65-MHz spectra to be supplied individually to receivers. Alternatively, it is possible to build return receivers that tune to 240 MHz, eliminating the need for down-conversion in the headend.

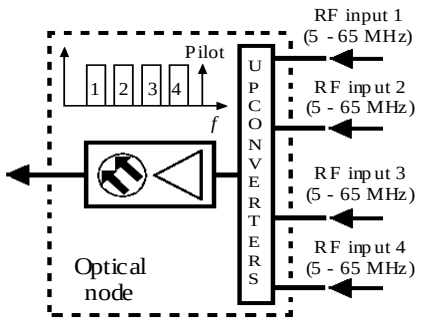


Fig. 3. Block conversion at an optical node

An alternative to block conversion is dense wave division multiplexing (DWDM). Each branch coming back to the node is supplied to a different optical transmitter operating on a different wavelength. A DWDM is used to combine the wavelengths on a single fiber for transmission to the headend or hub. A second DWDM is used at the headend or hub to separate the individual wavelengths before they are supplied to individual optical receivers.

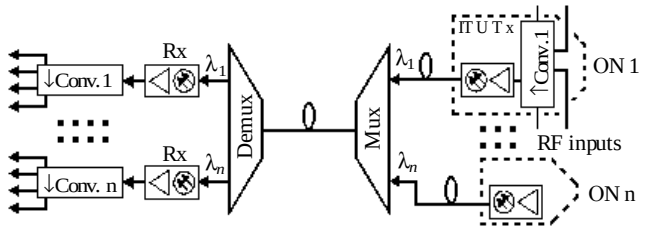


Fig. 4. DWDM reverse path configuration 1

Combining both DWDM and block up-conversion at an optical node it is possible to have 4*n* return bands on a single fiber. There are two configurations currently being investigated to combine these two technologies. The difference is the location of the ITU grid transmitters. In the first configuration these transmitters are located at the hub. In the second configuration, illustrated by Fig. 4, DWDM transmitters are located in the node. Fed by the RF spectrum from the upconverter the individual wavelengths are transmitted back to the hub.

A third technology for combining reverse path signals has also been developed: digitize the return band at the node and transport the digitized signals to the headend. At the headend, the digitized signals are converted again to RF in order to allow interface with legacy headend systems.

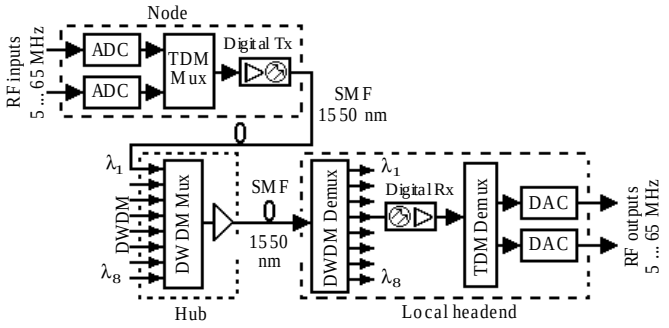


Fig. 5. DWDM digital return system

The implementation principle of a DWDM digital return system is shown in Fig. 5. In such a system, the outputs of the two diplexer low-pass sections are individually digitized in the two analog-to-digital converters (ADC), and then applied to a TDM multiplexer, which alternately passes data from one ADC and then from the other. The data are serialized and supplied to a digital transmitter for transport to the headend. At the headend, the data is demultiplexed into two signal streams, which are converted to RF in digital-to-analog converters (DAC). The RF signals can then be supplied to the normal headend upstream receiving equipment, such as DOCSIS CMTSs for data.

IV. FIBER-DEEP ARCHITECTURES

Several architectures are being used to drive fiber deep into the network, even as far as the home. Most involve a passive optical network (PON), which runs one feeder fiber from the central office to a passive terminal, then distributes the transmitted signals over distribution fibers to each of typically 16 to 32 optical nodes. Fiber-deep architectures, such as FTTC and FTTH, are driven by the improved quality of signals delivered over fiber-optic plant as compared to those delivered over coaxial plant, by improved reliability due to fewer devices in the network, and by improved bandwidth. Because the multiple services (analog and digital video, cable modem-based Internet access) are multiplexed using separate RF subcarriers, the delivered signals are compatible with existing consumer appliances, and demultiplexing the desired information from the full data stream is very simple. Furthermore, this RF subcarrier multiplexing gives the network operator the ability to gracefully evolve the service mix over time if, for example, the operator wishes to replace some or all of the analog channels with digital services.

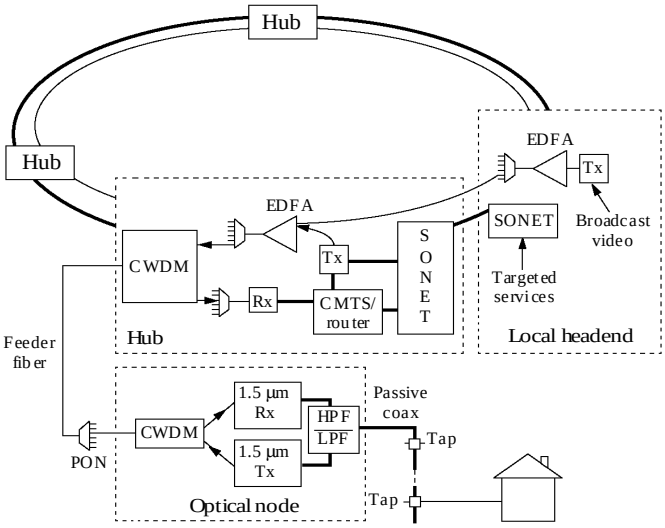


Fig. 6. Fiber-deep architecture

In Fig. 6 a concept for building a CATV system based on the FTTC/FTTH technology is presented. The local headend typically receives video signals in baseband format from a central primary headend, which receives them from a satellite or from local broadcasters. At the local headend, these signals are converted to AM-VSB format for analog video and to QAM format for broadcast digital video and transported to hubs, each serving roughly 20 000 subscribers, via a path-redundant supertrunking ring (that is, each output of the local headend optical splitter connects to a different hub on the ring via a dedicated fiber, and two connections per hub are made – one clockwise around the ring and one counterclockwise for redundancy). IP data and narrowcast video channels are typically carried together from local headend to hub over separate fibers using synchronous optical network (SONET).

Targeted services (TS) are passively combined with broadcast services at the hub and then the downstream signal

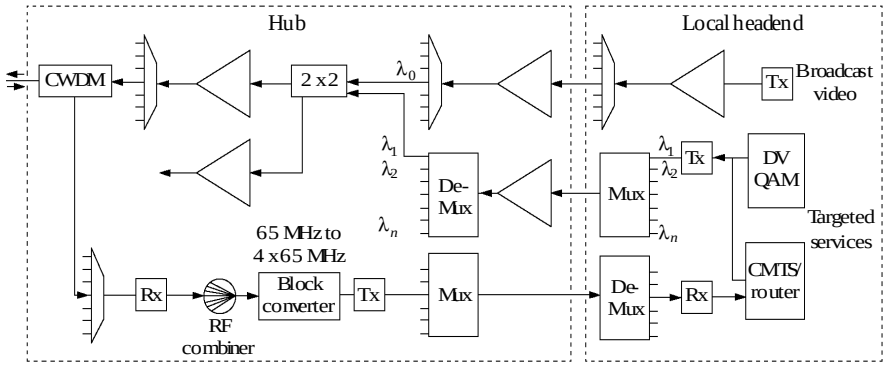


Fig. 7. DWDM-based fiber-deep architecture

is amplified and split multiple times. Amplifiers with up to +30 dBm are now commercially available and these high powers permit extensive optical splitting so that the amplifier cost and, indeed, the cost of all components from the splitter back can be shared over a large number of users. Inserting the TS channels at the hub reduces the sharing of the available TS bandwidth. Moving the TS insertion point toward the output of the hub reduces bandwidth sharing.

Unlike HFC, which uses dedicated downstream and upstream feeder fibers to connect the hub to remote optical nodes, this architecture employs PONs that branch out from each hub and terminate at ONs. Each PON carries bidirectional signals via 1.5 µm/1.3 µm coarse wavelength division multiplexing (CWDM). The FTTH ON operates from its location on the side of a house, while the FTTC ON serves a plurality of houses via one or more coax buses. Coax drops to subscriber homes connect directly to existing in-house coax so that existing customer premises equipment (cable-ready TVs, STBs, cable modems, and IP telephones) can be connected to the network.

In order to increase the system's capacity DWDM can be used to deliver targeted services. DWDM allows to move the TS interfaces (CMTSs, video modulators, video servers, telephony bandwidth managers, and advertisement-insertion equipment) back to the headend, thereby centralizing the operation and maintenance of the network and saving valuable floor space in hubs. Additionally, the use of DWDM between the local headend and the hub eliminates the need for the SONET ring.

The fiber-deep architecture with DWDM is shown in Fig. 7. Typical for this architecture is that the TS transmitters are located in the local headend and each is fixed at a controlled wavelength. These sources are multiplexed onto one or more fibers and transported to the hub. At the hub, the wavelengths are separated and each one is inserted onto a different branch of the network. Thus, by putting a channel on a specific wavelength or a set of wavelengths, it can be targeted to a specific segment of the network. In this case the TS insertion is done optically with a 2 x 2 combiner in the hub.

V. CONCLUSION

The concepts to build up the downstream and upstream paths of a DWDM-based HFC CATV system, suggested in this paper, allow delivery of independent signals to various end users, and the cost-effective collection of return signals from those users. These independent signals, or targeted services, include internet data streams, telephony, requests for and delivery of video-on-demand, near-video-on-demand, and analog channels reserved by franchise requirements for public, education, and government use.

The presented fiber-deep CATV architectures are superior to HFC in that they offer more total bandwidth, both upstream and downstream, which can be shared among fewer homes. By limiting the coax plant to a small passive run in FTTC, or eliminating it altogether in FTTH, minimizing of noises and nonlinear distortions is achieved, which leads to improvement of QoS. High-power EDFAs, CWDM, and PONs - technologies make the extension of fiber deeper into the network practical.

Optical signals on separate wavelengths interact as they travel through the fiber, and those interactions generate various levels of crosstalk. Depending on the parameters of a given link, these mechanisms can have a serious effect on recovered RF signal quality. As our analysis shows, the quality of discrete components in general, and of the wavelength demultiplexer in particular, is typically the limiting factor in achieving acceptable low levels of crosstalk interference.

ACKNOWLEDGEMENT

The research described it this paper is supported by the Bulgarian National Science Fund under the contract No ДДВУ 02/74/2010.

REFERENCES

[1] D. Large, J. Farmer, *Broadband access networks*, Elsevier, Inc., USA, 2009.
[2] J-P. Laude, *DWDM Fundamentals, Components, Applications*, Artech House, Inc., Boston, 2002.
[3] Dutta A., M. Fujiwara, *WDM technologies: optical networks*, Elsevier Academic Press, USA, 2004.
[4] Ciciora W., J. Farmer, D. Large, M. Adams, *Modern Cable Television Technology*, Elsevier, USA, 2004.
[5] Green P., *Fiber to the home – The new empowerment*, John Wiley & Sons, Inc., 2006.
[6] Prat J., *Next-Generation FTTH Passive Optical Networks: Research towards unlimited bandwidth access*, Springer Netherlands, 2008.

Measurement Site and Procedures for Experimental 2D DOA Estimation

Marija Agatonovic[1], Zoran Stankovic[1], Bratislav Milovanovic[1], Nebojsa Doncov[1], Leen Sit[2] and Thomas Zwick[2]

Abstract – **In this paper, a measurement site and experimental results of Two-Dimensional Direction of Arrival (2D DOA) estimation using 2D MUSIC algorithm are presented. Measurements are carried out in an anechoic chamber employing a horn antenna at the transmitting site and a rectangular 4 × 4 microstrip patch antenna array at the receiver. Measurement process is briefly described as well as the equipment and software used for that purpose. 2D MUSIC spectrum for different angular positions of source in azimuth and elevation is presented. Frequency dependency of DOA estimates is analysed.**

Keywords – **2D MUSIC, DOA estimation, frequency channel sounding, URA.**

I. INTRODUCTION

Direction of Arrival (DOA) estimation is well studied problem in recent years as it has numerous applications in wireless communication systems, military, acoustics, seismic, and medicine. Phase differences between signals collected by the array elements make it possible to calculate DOAs, and phase differences between the consecutive frequencies of the complex envelope at each element allow estimating the delays [1]. Using DOA estimation algorithms, both quality of received signals and capacity of mobile communication systems can be significantly improved. Further, there are a number of location-estimation techniques that are also based on DOA estimation. Employing two or more base stations, and adaptive arrays equipped with DOA estimation algorithms, they are able to determine the transmitter location by calculating the intersection of estimated DOAs from different base stations.

The most common algorithm for DOA estimation is MUSIC (MUltiple SIgnal Classification) [2]. This algorithm is well-known of its super-resolution capability as it provides highly-accurate DOA estimates. To do that, MUSIC needs "a priori knowledge" of signal characteristics as well as the total number of sources whose DOAs have to be estimated. Being a subspace-based algorithm, it performs eigen-decomposition of spatial covariance matrix that is formed from the received signals at different elements of an antenna array. In this paper, a system and procedures to provide these data experimentally are described as well as the results of 2D MUSIC algorithm. Since both azimuth and elevation angular positions are estimated, planar antenna array configuration is utilized due to its ability to scan three-dimensional (3D) space.

To provide hardware simplifications of a data acquisition system and to avoid requirements for a large number of receivers (one for each array element), frequency-domain channel sounding technique is employed. Measurements are based on the VNA (Vector Network Analyzer) that provides frequency response of the radio channel established between a transmitting and a receiving antenna [3]-[5]. There is only one receiver that is shared in time by use of RF switch matrix. In this way, signals to form a spatial covariance matrix are provided, and the matrix is further processed by 2D MUSIC algorithm. Measurements are done for three positions of the transmitting antenna in elevation, for a number of azimuth angles and frequency points. In our future research work, measurement data verified by MUSIC algorithm here will be used to further increase an efficiency of neural network models developed for fast and accurate DOA estimation [6].

The paper is organized as follows: Data model for 2D MUSIC algorithm is presented in Section II. System setup for measurements as well as the antennas and calibration procedure performed in the anechoic chamber are described in Section III. Measurement scenario and analysis of measurement results are given in Section IV. Section V contains conclusions remarks.

II. 2D MUSIC ALGORITHM

In this section, a description of the conventional approach to provide DOA estimates using the 2D MUSIC algorithm is given.

Let us suppose that signals of K sources illuminate a uniform rectangular array (URA) composed of $M \times N$ elements. Position of each element in the array can be denoted by its coordinates (m, n), where $m = 0, 1, 2, ..., M - 1$, and $n = 0, 1, 2, ..., N - 1$. Array element in the origin of the spherical coordinate system is taken as the reference one. Phase differences of signals received at all other array elements are defined relative to the reference. Further, it is supposed that the URA is placed in yz-plane, with distance between adjacent elements in y- and z-direction of dy and dz, respectively. Received signals at array elements can be written as

$$\mathbf{x} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_K \\ \alpha_1 e^{j\Phi_{(1,0)}^{(1)}} & \alpha_2 e^{j\Phi_{(1,0)}^{(2)}} & \cdots & \alpha_K e^{j\Phi_{(1,0)}^{(K)}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 e^{j\Phi_{(M-1,N-1)}^{(1)}} & \alpha_2 e^{j\Phi_{(M-1,N-1)}^{(2)}} & \cdots & \alpha_K e^{j\Phi_{(M-1,N-1)}^{(K)}} \end{bmatrix} \quad (1)$$

[1]Marija Agatonovic, Zoran Stankovic, Bratislav Milovanovic and Nebojsa Doncov are with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: [marija.agatonovic, zoran.stankovic, bratislav.milovanovic, nebojsa.doncov]@elfak.ni.ac.rs.

[2]Leen Sit and Thomas Zwick are with the Institut für Hochfrequenztechnik und Elektronik, Karlsruhe Institute of Technology, Engessestraße 5, 76131 Karlsruhe, Germany, E-mail: [leen.sit, thomas.zwick]@kit.edu.

where α_1, α_2, …, α_K are the amplitudes, and $\Phi_{(m,n)}^{(k)}$ represents the phase difference of the signal belonging to the k-th source, and received at array element with coordinates (m, n). $\Phi_{(m,n)}^{(k)}$ can be explicitly defined as follows

$$\Phi_{m,n}^{(k)}(\varphi_k,\theta_k)=\frac{2\pi}{\lambda}(d_y m\cos\theta_k \sin\varphi_k + d_z n\sin\theta_k) \quad (2)$$

where k=1, 2, ... , K, φ_k and θ_k are DOAs of the k-th source, and λ is the wavelength of incident signals. Spatial covariance matrix can be estimated using the following formula

$$\mathbf{R}=\mathbf{x}\mathbf{x}^H \quad (3)$$

If eigenvalues of matrix **R** are denoted by μ_i (i=1, 2, …, M*N), and keeping in mind that there are K incident signals, we obtain

$$\begin{aligned}&\mu_1>\mu_2>\mu_3>\ldots>\mu_K>\sigma^2\\ &\mu_{K+1}=\mu_{K+2}=\mu_{K+3}=\ldots=\mu_{M*N}=\sigma^2\end{aligned} \quad (4)$$

where σ^2 stands for the thermal noise. On the basis of Eq. (4), it can be concluded that there are total K eigenvalues larger than the thermal noise. To determine the angular positions of K sources, a particular property of the sub-spaces spanned by the eigenvectors related to the large and the small eigenvalues of the **R** matrix is exploited. The beam-steering vectors virtually pointing towards the sources are linear combinations of the eigenvectors related to the large eigenvalues. Hence they must be orthogonal to the noise sub-space spanned by the eigenvectors related to the small eigenvalues in the **R** matrix.

If we denote the eigenvector that corresponds to eigenvalue μ_i (i = 1, 2, ..., M*N) by $\mathbf{e}_i$ (i = 1, 2, ... , M*N), then steering vector **a**, steering matrix **A**, and 2D MUSIC spectrum, P_{MU}, for angular positions φ and θ can be determined using the following expressions

$$\mathbf{a}_{m,n}^{(k)}(\varphi_k,\theta_k)=e^{j\Phi_{m,n}^{(k)}(\varphi_k,\theta_k)} \quad (5)$$

$$\mathbf{A}(\varphi,\theta)=[\mathbf{a}_{(m,n)}^{(1)}(\varphi_1,\theta_1)\ \ \mathbf{a}_{(m,n)}^{(2)}(\varphi_2,\theta_2)\ \ \ldots\ \ \mathbf{a}_{(m,n)}^{(K)}(\varphi_K,\theta_K)] \quad (6)$$

$$\begin{aligned}P_{MU}&=\frac{1}{\sum_{i=K+1}^{M*N}\left|\mathbf{e}_i^H\mathbf{A}(\varphi,\theta)\right|^2}\times\mathbf{A}^H(\varphi,\theta)\mathbf{A}(\varphi,\theta)\\ &=\frac{\mathbf{A}^H(\varphi,\theta)\mathbf{A}(\varphi,\theta)}{\mathbf{A}^H(\varphi,\theta)E_N E_N^H\mathbf{A}(\varphi,\theta)}\end{aligned} \quad (7)$$

where

$$E_N=[e_{K+1}\ \ e_{K+2}\ \ \ldots\ e_{M*N}] \quad (8)$$

In other words, 2D MUSIC algorithm performs spectrum search for all angles in azimuth and elevation. In case the orthogonality condition is fulfilled, the Eq. (7) will give a strong peak since the denominator approaches zero.

III. System Setup

Vector network analysis of a signal is a method to accurately characterize its components by measuring their effect on amplitude and phase of the swept-frequency. Therefore, short distance measurement of radio channel characteristics, based on the use of VNA, represents very attractive measurement technique because of very simple implementation requirements, relative flexible system, and ability to track system error [1]. By the use of frequency sweep technique wide dynamic range can be provided for measurements. Further, it is possible to directly measure absolute losses due the path loss between two observed antennas. On the other side, disadvantage of this measurement method is slow measurement time and requirements for long cables to transmit the referent signal what limits its use on short-distance links with relatively stationary channel.

The measurement system shown in Fig. 1 is based on VNA (HP 8510) and frequency converter (HP 8511A). Frequency-channel sounding is carried out by sweeping a set of narrowband sinusoid signals through a frequency band. The VNA operates in the transfer function mode where one of its ports serves as the transmitting port and the other as the receiving port. Scattering S_{21} - parameter is used to express the complex frequency channel transfer function. VNA sends a frequency tone f through the channel and channel transfer function is represented as $S_{21}(f)$.

At the receiving site, each array element is through the coax cable connected to the appropriate input of RF switch matrix (Fig.1). The signal from the matrix is then guided to the second port of the VNA. Array elements are switched sequentially, where only one array element is active in a moment while other elements are present as dummy elements. Measurements at the network analyzer are averaged 16 times. At each angular position in the anechoic chamber 15 snapshots are recorded.

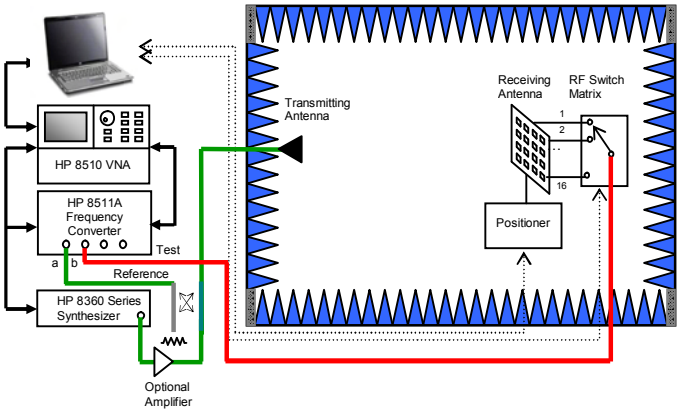


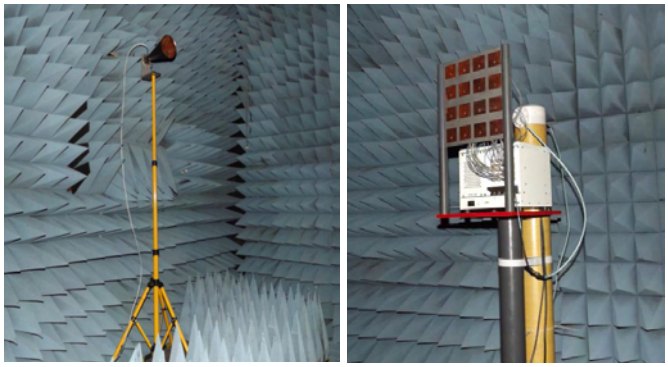
Fig. 1. Block scheme of the measurement system

To run the measurement routine automatically, VNA and the positioner in the anechoic chamber are controlled by the MATLAB software. Existing software for standard antenna measurements, SPAM 3D, is upgraded to implement RF switch matrix as a new instrument in the measurement setup. Code for the matrix control (selecting one of the 16 channels) is written in C, DLLs (Dynamic Link Libraries) are created and called from MATLAB. Previously, channels of RF switch matrix are properly configured and tested using MAX (Measurement & Automation Explorer). Positioner in the anechoic chamber is able to rotate in azimuth only, in the

smallest step of 1°. To make measurements for different elevation positions it is necessary to change the height of the transmitting antenna.

A. Antennas

As a transmitting antenna, a double-polarized quad-ridged horn antenna (model no. A6100), operating in the frequency band 2.18 - 20 GHz, is employed (Fig. 2 (a)). The gain of the antenna is around 7 dBi at 2.44 GHz. At the receiving site, a rectangular antenna array is positioned, composed of 16 (4 x 4) microstrip patch antennas with resonant frequency at 2.44 GHz (Fig. 2 (b)). Dimensions of microstrip patches are optimized in CST Microwave Studio, (L=39.6 mm, W= 49.4 mm) and realized on Rogers RT5880 Duroid substrate (epsilon=2.2, thickness=1.57 mm). Measured return loss of a single patch is -37 dB, and measured gain is 6.7 dBi at resonant frequency.



(a) (b)

Fig. 2. Transmitting antenna mounted on the tripod (a), receiving array with RF switch matrix behind placed on the antenna tower (b)

B. Calibration of a system

Calibration procedure is necessary to remove errors caused by the source, and more important, to eliminate all frequency dependent effects of a measurement system such as reflections in cables and connectors. The accuracy of the calibration procedure efficiently sets the dynamic range of the measurement system. To measure a transmission, as it is case with the characterization of a radio channel, response calibration is needed. Calibration results are then used to correct system errors of frequency response [1], [3].

Transfer function S_{21} of the complete system is measured in the anechoic chamber using reference antennas on mutual distance of 5.1 m and height of 1.90 m. Calibration data are recorded for each channel of RF switch matrix separately, and used later in the post-processing to normalize measured data. It is known that any change between the calibration and test setups can cause errors in measurement results such as multiple reflections on transmitting and receiving site due to the impedance change (when antenna elements are switched) and changes of cable characteristics due to flexures. To minimize such errors it is very important not to move any instrument or cable while the measurement is running.

IV. MEASUREMENT SCENARIO AND RESULTS

Measurements are performed as follows:

1) Transmitting antenna is set on a certain height on a tripod.
2) Receiving antenna array is rotated to a position in azimuth.
3) A channel of the RF switch matrix is selected and frequency sweep in the band from 2.41 to 2.47 GHz (51 pts) is performed. This procedure is repeated for all 16 channels.
4) Antenna tower is rotated to other position in azimuth, and array elements are switched. This is done for all required azimuth positions.
5) Results are saved in a .mat file.
6) Height of the transmitting antenna is increased by 10 cm (2.00 m) and set of data is recorded.
7) Height of the transmitting antenna is decreased by 50 cm (1.40 m) and set of data is recorded.

2D MUSIC spectrums, for three angular positions of the transmitting antenna, are plotted in Figs. 3, 4 and 5. Reference DOAs are (6°, 1.15°), (-30°, 0°) and (27°, -5.6°), and corresponding estimates are (9.11°, 1.12°), (-29.83°, 1.37°) and (31.83°, -4.11°). Figs. 6 and 7 present these values separately for azimuthal and elevation plane. Solid lines are plotted for DOA estimates and vertical dotted lines present the expected angular positions. It can be concluded that measured values are in good agreement with reference DOAs. For θ = 0°, estimation of azimuth angle is the most accurate. On the other side, there is an error of 1.37° for the elevation angle that could appear due to mutual coupling between antenna elements or non perfect alignment of the transmitting and receiving antenna during the measurement. The largest error is 4.83° for the azimuh angle, made for the elevation angle θ = -5.6°.

In Fig. 8, DOA estimates are plotted as a function of azimuth angles from -45° to 45°, and frequencies from 2.41 GHz to 2.47 GHz. As frequency shift corresponds to very small change of the wavelength of the incident signal (approximately $\lambda/50$), it can be observed that performances of URA for DOA estimation are not significantly deteriorated for lower and upper frequencies of its impedance bandwidth.

V. CONCLUSION

In this paper, measurement site for 2D DOA estimation is described, and performances of 2D MUSIC algorithm are evaluated. All measurements are carried out in an anechoic chamber taking into the consideration only the direct signal between the transmitting and the receiving antenna. As it is demonstrated, a 16-element microstrip patch array has good performance for 2D DOA estimation. To provide more accurate 2D DOA estimates, mutual coupling between array elements should be compensated performing careful calibration procedure for the array.

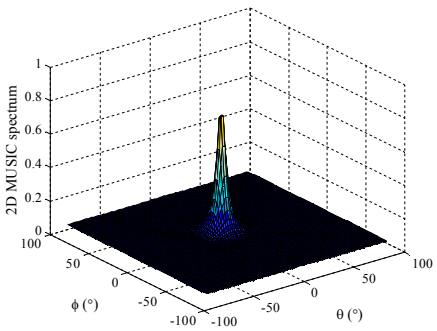


Fig. 3. 2D MUSIC spectrum for source position (φ, θ)= (6°, 1.15°)

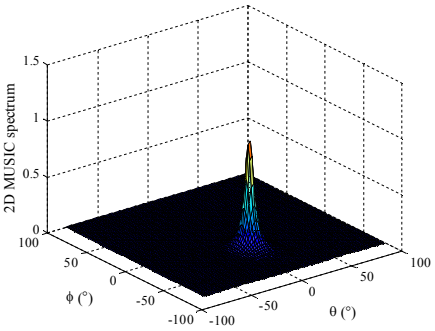


Fig. 4. 2D MUSIC spectrum for source position (φ, θ)= (-30°, 0°)

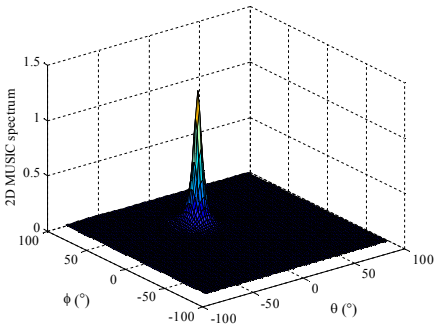


Fig. 5. 2D MUSIC spectrum for source position (φ, θ)= (27°, -5.6°)

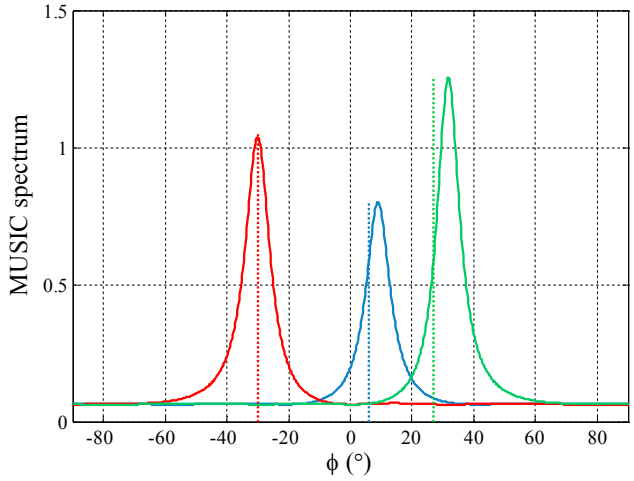


Fig. 6. Angular positions in azimuth (solid lines - DOA estimates, dotted vertical lines – actual positions)

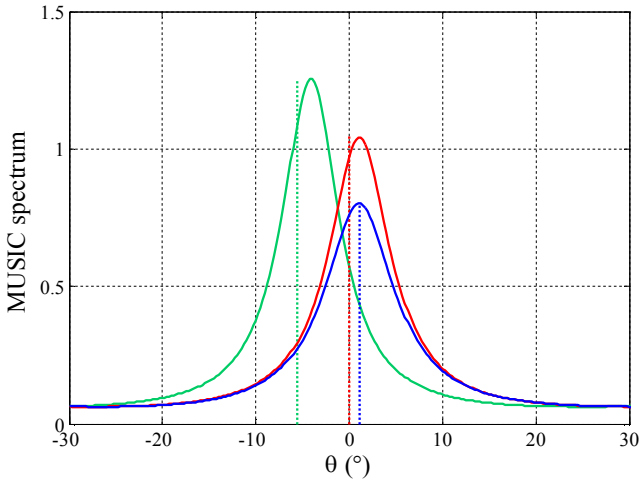


Fig. 7. Angular positions in elevation (solid lines - DOA estimates, dotted vertical lines – actual positions)

ACKNOWLEDGEMENT

This paper is a joint work of the project TR-32052 of the Serbian Ministry of Education and Science and the project CARE (Coordinating the Antenna Research in Europe). Measurements are performed at the Institut für Hochfrequenztechnik und Elektronik, Karlsruhe Institute of Technology, Karlsruhe, Germany.

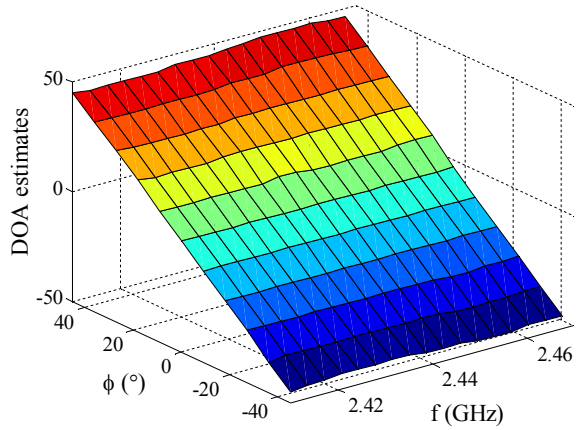


Fig. 8. Frequency dependency of DOA estimates for azimuth positions φ = [-45º, 45º] and elevation θ = 0º

REFERENCES

[1] A. M. Street, L. Lukarna and D. J. Edwards, "Use of VNAs for wideband propagation measurements," *IEEE Proc. - Commun.*, vol. 148, no. 6, December 2001.

[2] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276-280, March 1986.

[3] L. H. Macedo, M. H. C. Dias, R. D. Vieira, J. F. MacQdo and G. L. Siqueira, "Mobile Indoor Wide-Band 1.8 GHz Sounding: Measurement-Based Time Dispersion Analysis," *Proc. Vehicular Technology Conference*, vol.1, pp. 375 – 379, 2002.

[4] M. Hajian, C. Coman and L. P. Ligthart, "Comparison of circular, Uniform- and non Uniform Y-Shaped Array Antenna for DOA Estimation using Music Algorithm," *Proc.9th European Conference on Wireless Technology*, Manchester, UK, September 2006.

[5] A. Hafiizh, S. Obote and K. Kagoshima, "Multiple Subcarrier Indoor Geolocation System in MIMO-OFDM WLAN APs structure," *World Academy of Science, Engineering and Technology 58*, 2009.

[6] M. Agatonović, Z. Stanković, B. Milovanović and N. Dončov, "DOA Estimation using Radial Basis Function Neural Networks as Uniform Circular Antenna Array Signal Processor," *Proc. 10th IEEE International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services – TELSIKS 2011*, Niš, Serbia, vol.2, pp. 544-547, 2011.

Simulation influence of the thermal noise of PIN photodetector on performance DWDM optical network

Petar Spalević[1], Dejan Milić[2], Branimir Jakšić[1], Mile Petrović[1] and Ilija Temelkovski[2]

Abstract – **In this paper using simulation DWDM optical network with software OptiSystem is shown the influence of thermal noise PIN photodetector on the transmission quality. As a measure of quality used optical Q factor. On the basis of obtained results is discussed how to modify the transmission quality and behavior of the thermal noise with a variation of numbers amplifying section and the quantity of flow per channel. Graphically is shown change Q factor for different values of the thermal noise, the number amplifying section and length amplifying section.**

Keywords – **DWDM networks, thermal noise, Q factor, amplifying section.**

I. INTRODUCTION

WDM (Wavelength Division Multiplexing) is a technology which multiplexing multiple optical carrier signals on a single fiber using different wavelengths for the transmission of different information. At least attenuation in the optical fiber is achieved by applying the wavelength of 1550 nm, i.e. using the "third optical window" [1],[2]. WDM systems allow expansion of existing capacity without laying additional fiber optic cables. The capacity of the existing system is expanding using multiplexers and demultiplexers at each end of the system [3], [4].

DWDM (Dense Wavelength Division Mulitplexing) DWDM relates to optical signals multiplexed within the 1550 nm range in order to influence Erbium Doped Fiber Amplifier (EDFA), which are effective for wavelengths from 1525 to 1565 nm (C band).

Unlike a DWDM, CWDM (Coarse Wavelength Division Mulitplexing) uses a much larger spacing between channels. To enable 16 channels on one fiber, CWDM uses the entire frequency range between the second and third optical window (1300 nm and 1550 nm).

For successful transmission of optical signals over long distances is using the EDFA amplifiers. Weak signal enters in the erbium doped fiber in which light is injected using the laser to pump. This light excites erbium atoms to release stored energy as additional light of wavelengths around 1550 nm. How this process continues through the fiber signal becomes stronger. EDFA is available in C and L windows but quite narrow range (1530-1560 nm) [1], [2], [5].

[1]Petar Spalević, Branimir Jakšić and Mile Petrović is with the Faculty of Technical Sciences at University of Pristina, Kneza Miloša 7, Kosovska Mitrovica 38200, Serbia, E-mail: spalevicpetar@yahoo.com.

[2]Dejan Milić and Ilija Temelkovski is with the Faculty of Electronic at University of Niš, A. Medvedeva 14, Niš 18000, Serbia.

Thermal noise (also known as Johnson-Nyqvist noise) is electrical noise created thermal (Braun) the movement of electrons within the electrical conductors without any external influence. This noise occurs regardless of the applied external voltage, unlike other sources of noise.

The variance of thermal noise can be given as expression [1], [2]:

$$\sigma_T^2 = \frac{4k_B T}{R_L} F_n \Delta f \tag{1}$$

where k_B is the Boltzmann constant, T is the absolute temperature, R_L is the load resistor, Δf is the effective noise bandwidth, and F_n represents the factor by which thermal noise is enhanced by various resistors used in pre- and main amplifiers.

The performance of an optical receiver depends on the SNR (signal-to-noise ratio). The SNR of any electrical signal is defined as

$$SNR = \frac{\text{average signal power}}{\text{noise power}} = \frac{I_p^2}{\sigma^2} \tag{2}$$

where i_p is defined as $I_p = RP_{in}$ (R is the responsivity of the *PIN* photodiode defined as $R = \frac{\eta q}{h\nu}$). In most cases of practical interest, thermal noise dominates receiver performance ($\sigma_T^2 >> \sigma_S^2$). Neglecting the shot-noise (σ_S^2), Eq. (2) can be written as

$$SNR = \frac{R_L R^2 P_{in}^2}{4k_B T F_n \Delta f} \tag{3}$$

Optical Q factor is given with expression [1],[2]:

$$Q = \frac{I_0 + I_1}{\sigma_0 + \sigma_1} \tag{4}$$

Q factor is determined based on the average value of the signal I_1 and standard deviations σ_1 in the event of receipt" 1" i.e. based on the average value of signals I_0 and standard σ_0 deviation in case of receipt "0". In WDM networks, the transmission quality was achieved for the values of Q> 5.4.

II. SYSTEM MODEL

The appearance of WDM optical networks for the analysis in the OptiSystem [11] software is given in Fig. 1. It consists of an optical source, DWDM multiplexers, optical fibers, EDFA amplifiers, DWDM demultiplexers and optical receiver.

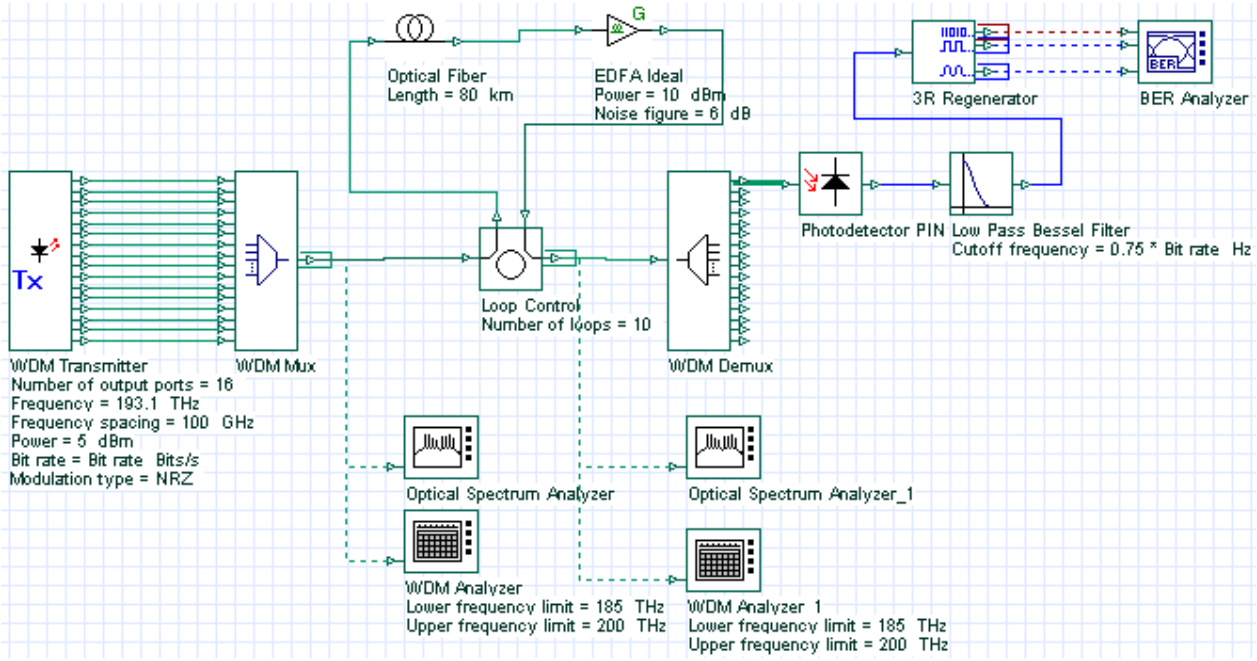


Fig. 1. The appearance of DWDM networks

Observe the optical WDM 16-channel network in the third optical window with the bit rate at channel 1 Gb/s and 5 Gb/s.

The power emitted by each source is 5 dBm. The system works in the third optical window and the frequency of each DWDM channel: first channel 193.1 THz, second channel 193.2 THz, third channel 193.3 THz, ... Channels operating at wavelengths: 1552,524 nm, 1551,721 nm, 1550,918 nm, ..., respectively. The frequency of each channel is separated by 100 GHz.

Signals are transmitted from the multiplexer through optical fiber amplified with EDFA amplifier. Gain EDFA is 10 dBm, noise that enters is the 6 dB. Amplifiers were set at 50 km, and the number of amplifying section changes from 1 to 10 in order to obtain the influence of the length of the section on the quality of signal transmission. System working in the third optical window so the weakening of the fiber is 0.2 dB/km.

On the receiver side, the signals coming to demultiplexers whose channels work on the same frequencies as the multiplexer. At the exit of the demultiplexers is optical receiver which consists of three elements: PIN photodetector, low-pass Bessel filter and 3R regenerator. Thermal noise is changing in the PIN photodetector in the range of 10^{-26} W/Hz up to 10^{-18} W/Hz. Optical receivers usually have a range of thermal noise in the bordered 10^{-24} W/Hz up to 10^{-20} W/Hz.

In addition to these network elements, include elements from which it is possible to read values that indicate the performance of the transmission of optical signals: WDM Analyzer, Optical Spectrum Analyzer and BER Analyzer [11].

III. Results Of Simulation And Discussion

In Tables I and II are given values of Q factor obtained by simulation, dependent on the different values of the thermal noise in the PIN photodetector for amplifying section of 50 km with a bit rate of 1 Gb/s and 5 Gb/s per channel, respectively. The obtained results shows that the increase in the number of amplifying section Q factor decreases for values of the thermal noise of 10^{-20} to 10^{-26} W / Hz. However, the thermal noise higher of 10^{-20} W/Hz leads to smaller fluctuations of Q factor, i.e. after his fall, to come up its growth in some amplifying sections.

If we consider the change of Q factors of the same sections, and for different values of the thermal noise we see that it increases with decreasing thermal noise to 10^{-24} W/Hz, when Q factor obtained approximately constant value.

TABLE I
CHANGING THE Q FACTOR FOR DIFFERENT VALUES OF THERMAL NOISE OF PIN PHOTODETECTOR FOR AMPLIFYING SECTION OF 50 KM ON DWDM NETWORK BIT RATE 1 GB/S PER CHANNEL

thermal noise W/Hz	10^{-18}	10^{-20}	10^{-22}	10^{-24}	10^{-26}
50 km	14.8916	27.026	28.0084	27.9900	27.9859
100 km	13.0851	23.9029	24.8721	24.8619	24.8592
150 km	11.8144	18.9094	19.5492	19.5555	19.555
200 km	10.2939	17.5041	18.1705	18.1860	18.1868
250 km	10.5957	16.6127	17.0582	17.0686	17.0691
300 km	9.46335	15.9481	16.4939	16.5027	16.5031
350 km	9.66262	14.8035	15.2027	15.2135	15.2142
400 km	9.97910	13.8202	14.0451	14.0504	14.0507
450 km	9.12860	12.9756	13.1663	13.1700	13.1702
500 km	8.54721	11.5364	11.7276	11.7351	11.7357

Dependence of Q factor on the fiber length for bit rates 1 Gb/s and 5 Gb/s is given in Fig. 2. Dashed red line shows the

limits of transmission quality on basis Q factor. From Fig. 2 we can see that the transmission quality can be achieved with a bit rate of 1 Gb/s per channel for all 10 amplifying sections if the value of the thermal noise PIN photodetector is 10^{-18} W/Hz and 10^{-22} W/Hz. In the case of transfer of 5 Gb/s per channel transmission quality can not be achieved in all 10 amplifying section (500 km). For value of the thermal noise 10^{-18} W/Hz distortion of quality transmission occurs after 250 km. For value of the thermal noise 10^{-22} W/Hz distortion of quality transmission occurs after 300 km.

Also, from Fig. 2 we can see that the drop Q factor is significantly more pronounced at lower values of the thermal noise than the higher values.

TABLE II
CHANGING THE Q FACTOR FOR DIFFERENT VALUES OF THERMAL NOISE OF PIN PHOTODETECTOR FOR AMPLIFYING SECTION OF 50 KM ON DWDM NETWORK BIT RATE 5 GB/S PER CHANNEL

thermal noise W/Hz	10^{-18}	10^{-20}	10^{-22}	10^{-24}	10^{-26}
50 km	8.02543	19.7139	20.3453	20.3177	20.3138
100 km	6.66421	17.0361	17.4225	17.3993	17.3963
150 km	7.05799	13.3352	13.3484	13.3309	13.329
200 km	5.95862	10.4932	10.517	10.5088	10.5079
250 km	5.41034	7.74074	7.75219	7.74878	7.7484
300 km	2.81884	4.42233	3.46192	3.46553	3.46588
350 km	3.06294	3.35063	3.3484	3.34782	3.34775
400 km	3.20253	3.40236	3.39184	3.39045	3.39031
450 km	3.1032	3.17188	3.15319	3.15109	3.15088
500 km	2.60183	2.95521	2.99387	2.99759	2.99796

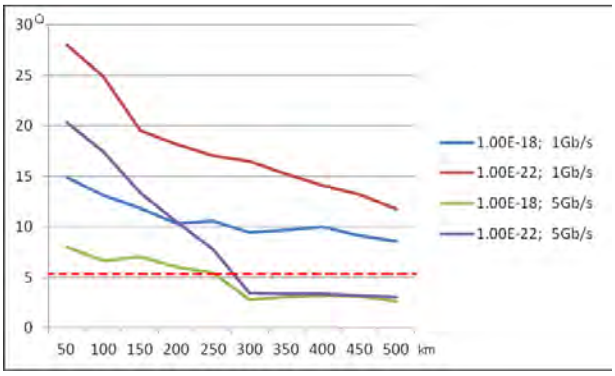


Fig. 2. Changing the Q factor at a bit rate of 1 Gb/s and 5 Gb/s per channel for the values of the thermal noise of 10^{-18} W/Hz and 10^{-22} W/Hz at amplifying sections of 50 km

The values of Q factor for the reduced value of the amplifying section of 25 km, for characteristic value of thermal noise of PIN photodetectors are given in Table III, for a bit rate of 5 Gb/s per channel.

In Figs 3, 4 and 5 shows the change Q factor when using amplifying sections 25 and 50 km and bit rate of 5 Gb/s per channel for the values of the thermal noise 10^{-18} W/Hz, 10^{-22} W/H and 10^{-26} W/Hz, respectively. Dashed red line shows the limits of transmission quality on basis Q factor.

In Figure 3 we see that whether we use the amplifying section 25 or 50 km distorsion of quality transmission occurs at the same length of 250 km. The behavior of Q factor for using amplifying sections 25 km and 50 km is approximately the same while there is a high quality transmission. When the Q factor drops below 5.4 occurs to abrupt change of its value in applying these amplifying sections. For the amplifying section of 50 km fall of Q factor is much more pronounced than in the case of a amplifying section of 25 km.

TABLE III
CHANGING THE Q FACTOR FOR DIFFERENT VALUES OF THERMAL NOISE OF PIN PHOTODETECTOR FOR AMPLIFYING SECTION OF 25 KM ON DWDM NETWORK BIT RATE 5 GB/S PER CHANNEL

thermal noise W/Hz	10^{-18}	10^{-22}	10^{-26}
25 km	6.98781	24.1337	24.2221
50 km	8.21292	21.9389	21.906
75 km	6.63424	18.9021	18.8769
100 km	6.47925	16.8889	16.8883
125 km	6.16904	13.2485	13.2831
150 km	6.9117	12.5919	12.5805
175 km	5.87067	10.7422	10.7382
200 km	5.97327	9.7244	9.71713
225 km	5.20897	8.40222	8.41579
250 km	5.39137	7.23923	7.23421
275 km	4.27537	6.62766	6.63298
300 km	4.45094	5.84338	5.83978
325 km	2.95336	3.21382	3.21286
350 km	2.87996	3.1264	3.12646
375 km	3.03281	2.9827	2.98049
400 km	2.87397	2.95531	2.95414
425 km	2.60157	2.94393	2.94562
450 km	2.85298	2.84018	2.83829
475 km	2.68203	2.73249	2.73183
500 km	2.33638	2.612	2.61413

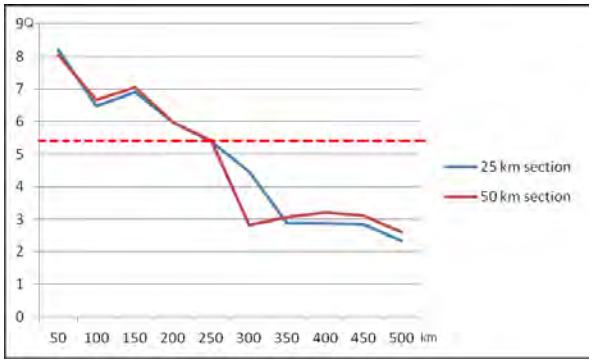


Fig. 3. Changing the Q factor at the thermal noise of 10^{-18} W/Hz for amplifying sections 25 and 50 km

In the case of the thermal noise of 10^{-22} W/Hz and 10^{-26} W/Hz curve behavior of Q factor is identical (Figs. 4 and 5 are

similar. When the value thermal noise falls below 10^{-22} W/Hz, Q factor tends constant value.

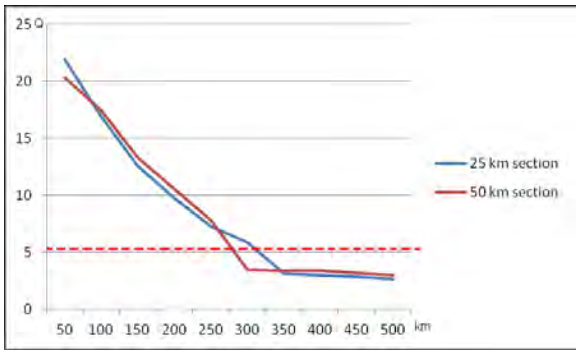


Fig. 4. Changing the Q factor at the thermal noise of 10^{-22} W/Hz for amplifying sections 25 and 50 km

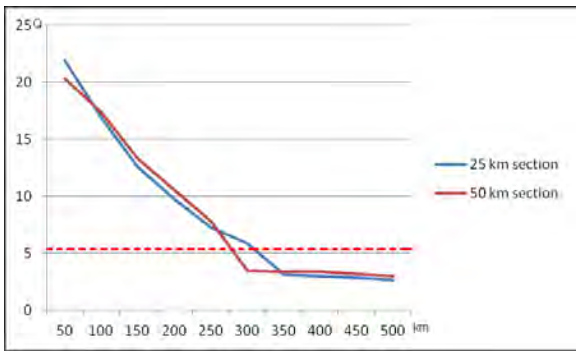


Fig. 5. Changing the Q factor at the thermal noise of 10^{-26} W/Hz for amplifying sections 25 and 50 km

From Figs. 4 and 5 we see that the behavior of the curve Q factor for amplifying section of 25 km and 50 km is not similar as the case for higher values of thermal noise. A faster decline in Q factor is achieved with amplifying sections of 50 km compared to the 25 km amplifying section. Better quality of transmission can be achieved by applying the amplifying section of 25 km with respect to section of 50 km.

IV. Conclusion

With simulation of the DWDM network is obtained the values of Q factors for different values of the thermal noise PIN photodetector and different number of amplifying sections. Based on these results we conclude that the increase in the number of amplifying section Q factor decreases for values of the thermal noise of 10^{-20} to 10^{-26} W/Hz. For higher thermal noise of 10^{-20} W/Hz Q factor also decreases but on the some amplifying sections there is a slight of his growth.

Quality transmission can be achieved with a bit rate of 1 Gb/s per channel for all 10 amplifying sections if the value of the thermal noise PIN photodetector is 10^{-18} W/Hz and 10^{-22} W/Hz. In the case of transfer of 5 Gb/s and value of the thermal noise 10^{-18} W/Hz distortion of quality transmission occurs after 250 km. For value of the thermal noise 10^{-22} W/Hz distortion of quality transmission occurs after 300 km.

Executed a comparison of Q factor and quality of transmission if the length amplifying sections cut from 50 km to 25 km, i.e. if you use twice many amplifiers. In the case of the thermal noise 10^{-18} W/Hz, regardless of whether use of the amplifying section 25 or 50 km distorsion of quality transmission occurs at the same length of 250 km. The behavior of Q factor for using amplifying sections 25 km and 50 km is approximately the same while there is a high quality transmission. When the Q factor drops below 5.4 occurs to abrupt change of its value in applying these amplifying sections.

In the case of the thermal noise of 10^{-22} W/Hz and 10^{-26} W/Hz the behavior of the curve Q factor for amplifying section of 25 km and 50 km is not similar as the case for higher values of thermal noise. Better quality of transmission can be achieved by applying the amplifying section of 25 km with respect to section of 50 km.

References

[1] Govind P. Agrawal, Nonlinear Fiber Optics, Academic Press, 2nd Ed., 2001.
[2] Govind P. Agrawal, Fiber-Optic Comunication Systems, Wiley, 3nd Ed., 2002. R. Ramaswami and K. Sivarajan, Optical Networks: A Practical Perspective, 2nd ed.,
[3] Morgan Kaufmann Publishers, San Francisco, 2002.
[4] J. Zyskind and R. Berry, in Optical Fiber Telecommunications IV, Vol. B, I. P. Kaminow and T. Li, Eds., Academic Press, San Diego, CA, 2002.
[5] A. Altuncu, L. Noel, W. A. Pender, A. S. Siddiqui, T. Widdowson, A. D. Ellis, M. A. Newhouse, A. J. Antos, G. Kar, and P. W. Chu, Electron. (1996).
[6] Mihajlo Stefanović, Dragan Drača, Aleksandra Panajotović, "The Common Influence of Time Shift and Appearing Place of Interference on Signal Propagation along Optical fiber", Electronics and Electrical Engineering, No. 1 (57), pp. 14-19, 2005.
[7] Mihajlo Stefanović, Dragan Drača, Daniela Milović, Aleksandra Panajotović, "Analytic Solution of Pulse Shape along the Fiber in the Presence of Interference and Third Order Dispersion", Journal of Optical Communications (accepted for publication #JOC 1095), 2006.
[8] Mihajlo Stefanović, Dragan Drača, Aleksandra Panajotović, Daniela Milović, "Performance of Optical Telecommunication System in the Presence of Chirped and Time Shifted Interference", Electronics, Vol. 8, No. 2, pp. 13-16, 2004.
[9] Mihajlo Stefanović, Dragan Drača, Aleksandra Panajotović, Daniela Milović, "Second and Third Order Dispersion Influence on Pulse Propagation in the Presence of the Interference", Conference Proceedings of ICEST 2006, pp. 91-94, Sofia, Bulgaria, 2006.
[10] Mihajlo Stefanović, Dragan Drača, Aleksandra Panajotović, "Influence of laser phase noise performance heterodyne optical PSK receiver", Proceedings of the conference "Noise and Vibration ", pp. 19-1-19-3, Nis, Serbia and Montenegro, 2000.
[11] http://www.optiwave.com/.

Study of ICI in PRS - OFDM systems

Stanio Kolev[1], Ilia Iliev[2], and Stoicho Manev [3]

Abstract – **In this work the inter carrier interference (ICI) in PRS-OFDM system with different polynomials is investigated. Carrier to interference ratio (CIR) is studied depending on the frequency offset (ε) for OFDM systems with any Correlative Coding (PRS) and without PRS.**

Keywords – **ICI, PRS, OFDM, CIR, Correlative coding**

I.INTRODUCTION

It is known that OFDM is widely used in digital communications in recent years. OFDM is successfully applied in mobile communications [4], such as: Wireless LAN (WLan) with very high speed transmission, ADSL, digital broadcasting (DAB and DVB) [3], data transmission in power line (PL) systems and in Satellite TV. In these systems arise some problems: ICI is caused by an error in synchronization of subcarriers, channel parameters that change over time and phase noise. OFDM is also sensitive to the frequency error between transmitter and receiver [2]. Carrier frequency shift causes many problems: Amplitude attenuation and phase shift of each subcarriers and interference between subcarriers (ICI) [1]. Impacts of frequency offset of the mobile communication systems performance are stronger. As a result of the relative displacement Doppler shift of carrier frequency occurs [5]. This causes synchronization errors between the bearing carriers in the receiver and transmitter, damage to orthogonality between the subcarriers, which causes rise to interference and ICI .

II. THEORETICAL BACKGROUND

In the presented paper basic relations are shown. These equations are used by the synthesis of the simulation model of the OFDM system, shown on fig.1. The main aim of the article is to obtain a global relation for the CIR estimation.

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

[2] Ilia Iliev ,Assoc.prof. at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: igiliev@tu-sofia.bg

[3] Stoicho Manev Assis. with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria

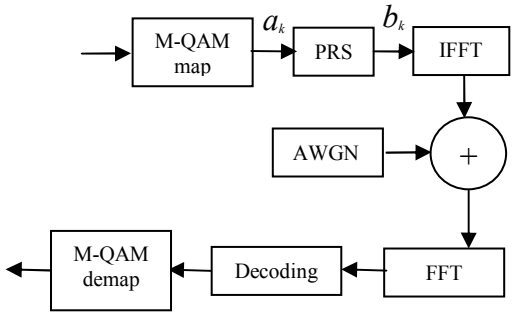


Fig.1 Block Diagram of PRS OFDM system

Here the following abbreviations have been used:
M-QAM mapping and demapping. Modulation in base band;
PRS: partial response signalling;
IFFT: inverse fast Fourier transformation;
FFT: fast Fourier transformation;
AWGN: additive white Gauss noise channel.

CIR is a key parameter in OFDM communication systems and it shows the effect of ICI, which is a random process. Several studies of the CIR value have been made. These values have been obtained by the investigation of OFDM system with and without PRS. A comparison has been made in [3]. The *k*-th subcarrier of the received signal in OFDM system with *N* subcarriers has the form [3]**:**

$$r_k = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{l=0}^{N-1} b_l \exp(j\frac{2\pi}{N}nl)\exp(j\frac{2\pi}{N}n\varepsilon)\exp(j\frac{2\pi}{N}nk) = \sum_{l=0}^{N-1} b_l S(l-k) \quad (1),$$

where:

$$S(l\text{-}k) = \frac{\sin(\pi\varepsilon + l)}{N\sin(\frac{\pi}{N}(\varepsilon + l - k))\exp(j\frac{\pi}{N}((N-1)\varepsilon - (l-k)))} \quad (2).$$

The parameter ε is the normalized frequency offset. The symbol, obtained at the output of the correlative coding block, is denoted by b_l. The received signal: r_k can be written as:

$$r_k = C_k + I_k,$$

where C_k is the information signal, and I_k is the interference in the *k*-th subcarrier. Basic relation for CIR estimation for M-ary modulation and any PRS polynomials is proposed:

$$CIR = \frac{E\left(|C_k|^2\right)}{E\left(|I_k|^2\right)} \quad (3).$$

$$E\left(|C_k|^2\right) = E\left(|b_k|^2\right)\left(\frac{\sin \pi.\varepsilon}{\pi.\varepsilon}\right)^2 = Eq.C.C^T.\left(\frac{\sin \pi.\varepsilon}{\pi.\varepsilon}\right)^2 \quad (4)$$

where:

$$Eq = \begin{cases} 1 & M = 2 \\ \frac{2(M\text{-}1)}{3} & M = 4,16... \end{cases} \quad (5)$$

M -ary modulation,

Eq- the mean of the energy of the sequence a_k.,

C - vector, which elements are the coefficients of the PRS polynomial.

The signal interference between the subcarriers is:

$$E[\left|I_k\right|^2]= Eq.C.C^T.\sum_{l=1}^{N-1}\left|S(l)\right|^2 - \left[\sum_{l=2}^{N-1} S(l)S^*(l-1)\mathrm{E}[b_l, b_{l+i}] + S(l-1)S^*(l)\mathrm{E}[b_l, b_{l-i}]\right] \quad (6)$$

Where: i =1,2..K-1, a K is the PRS polynomial elements number and $E[b_l, b_p]$ is the cross correlation function between b_l and b_p.

This expression is used to develop a simulation model, designed to measure CIR. The results of the experiments are given bellow.

III. EXPERIMENTAL RESULTS

The Diagram obtained as a result of simulations of PRS - OFDM systems using polynomial encoding of different classes is shown on fig.2.

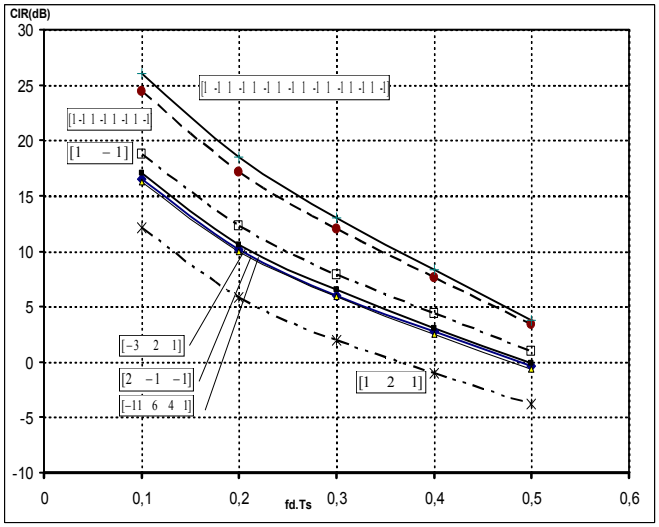


Fig.2. CIR in PRS - OFDM systems using polynomials with coefficients [-1 1 6 4 1], [-3 2 1], [2 -1 -1], [1 -1], [1 2 1] [1-1 1 -1 1 -1 1 -1].

On the x axis is ε - the normalized frequency offset and y-axis is the CIR in [dB]. The simulation environment is Matlab .

The simulation results are obtained under the following conditions:

the parameter ε has been changed

CIR has been measured.

The comparison between the obtained and the well known in the scientific literature results allows general estimation to be made and can be used as methodology for choice of appropriate coefficients for PRS polynomials, where the ICI has been minimized without decreasing the BER.

After number of experiments, selection of polynomials set, used in OFDM systems with PRS, in whom a significant improvement in CIR compared to conventional systems, has been carried out. The results of simulation studies have shown that PRS-OFDM system, using polynomials with coefficients (4, -3, -1) (-3, 2, 1) (2, -1, -1) , (1 -1 1 -1 1 -1 1 -1 1 -1 1 -1) in the PRS simulation model, give an improvement of CIR compared with the OFDM system without PRS.

IV.CONCLUSION

The given relations for CIR estimation for M-ary modulation and any PRS polynomials and simulations describe a possible methodology for the study of ICI in OFDM systems with PRS coding. It can be selected such polynomials that minimize ICI and keep the BER performance. This is particularly important for communications with fast moving mobile objects, where the Doppler shift is relatively large compared with the distances between the subcarriers .

REFERENCES

[1] P. Kabal and S. Pasupathy, "Partial response signaling," IEEE Trans. Commun., vol. COM-23, pp, 921-925, Sept. 1975.

[2] Proakis, J.G, "Digital Communications", 2"d Edition, McGraw-Hill,l989

[3] Zhao Y.,Jean-Damien L., and Sven-Gustav H.,"Intercarrier Interference Compression in OFDM Comminication Systems by using Correlative Coding",IEE Commun. Lett.,vol 2,pp.1089-7798,August 1998

[4] M. Russell and G. J. Stuber, "Interchannel interference analysis of OFDM in a mobile environment," in Proc. IEEE Vehicular Technology Conf., 1995, pp. 820–824.

[5] Hua Zhang and Ye (Geoffrey) Li "Optimum Frequency-Domain Partial Response Encoding in OFDM System" IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 51, NO. 7, JULY 2003

[6] F. Rusek and J.B. Anderson, "Maximal Capacity Partial Response Signaling," Communications 2007, ICC'07 IEEE International Conference 24-28 June 2007 pp821-826

[7] Kolev St. "A New Wireless PRS-OFDM Simulation Model" CEMA'10, Athens, 2010

BER simulation analysis of PRS - OFDM systems with MLSD

Ilia Iliev[1], Stanio Kolev[2] and Stoicho Manev [3]

Abstract – **In the presented paper investigations, concerning the correlation of one of the main parameters in the communication systems: BER from the SNR by fixed carrier frequency offset ε, have been carried out. Systems without PRS and systems with PRS have been studied. The using of MLSD is a way to improve BER performance by PRS signals detection.**

Keywords – ICI, PRS, OFDM, BER, Correlative coding, MLSD

I. INTRODUCTION

OFDM systems have been widely used in the digital communication systems. Due to the **frequency offset ε** the inter subcarriers interference (ICI) increases. This leads to negative effect to the BER performance. In this work comparison between the simulation manner obtained values for BER, concerning different polynomials for PRS coding, has been made.

II. SYSTEM MODEL

The block diagram of the investigated communication system is presented bellow [3]:

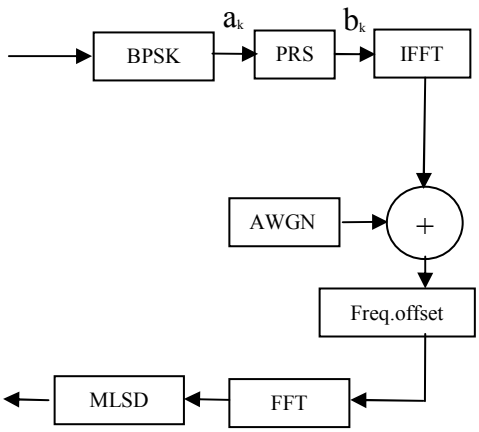


Fig.1. Block diagram of the communication system with 1- D coding block and MLS Decoder

Here the following abbreviations have been used:

BPSK: binary phase shift keying;
PRS: partial response signalling;
IFFT: inverse fast Fourier transform;
FFT : fast Fourier transform;
AWGN: additive white Gauss noise.

MLSD - Maximum Likelihood Sequence Detector.

The information has been coded by means of BPSK. After this operation PRS coding has been done. The following polynomials:[1], [4,-3,-1], [-3,2,1] and [3,2,-1] have been investigated. The first polynomial with coefficient [1] is equivalent to the case, when system without PRS coding is used.

The next operation in the system model is IFFT. The signal, obtained after the IFFT, is transmitted in base band through channel with additive white Gauss noise. Frequency offset ε has been considered.

The receiver includes FFT. The detection has been carried out by means of MLSD decoder. Vitterby algorithm has been used and the received sequence has been detected. In the detector channel estimate impulse response has been used. Traceback depth for equalizer is equal to 12.

III. EXPERIMENTAL RESULTS

Results are obtained by means of number of simulations. On the figures bellow the experimental results are shown.

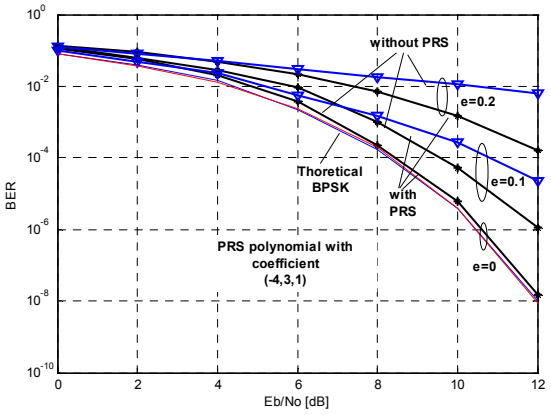


Fig. 2. Study of BER of PRS-OFDM system using a polynomial with coefficients (-4 3 1)

The simulation environment is Matlab .

The simulation results are obtained under the following conditions: the parameter ε has been changed BER has been measured.

[1] Ilia Iliev, Assoc.prof. in the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: igiliev@tu-sofia.bg

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

[3] Stoicho Manev Assis. with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria

The comparison between the obtained and the well known in the scientific literature results allows general estimation to be made and can be used as methodology for choice of appropriate coefficients for PRS polynomials, where the ICI has been minimized without decreasing the BER.

For Fig.2 when ε = 0,1 and SNR is greater than 7 dB there is an improvement of BER, if PRS coding is used. This is evident on the chart (the 4-th line from the bottom up (the blue line) refers to OFDM system without PRS) and the 3-th line from the bottom up (the black line) refers to OFDM system, using a PRS polynomial (4,-3,-1).

When ε = 0,2 and SNR is greater than 4 dB improvement is observed if PRS coding is used. This is evident on the chart - the third pair of curves (blue and black) from the bottom up.

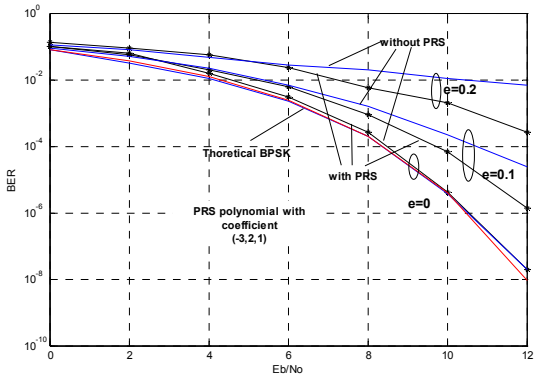


Fig. 3. Study of BER of PRS -OFDM system using a polynomial with coefficients (-3 2 1)

For Fig.3 when ε = 0,1 and SNR is greater than 6 dB there is an improvement of BER, if PRS coding is used. This is evident on the chart (the 4-th line from the bottom up (the blue line) refers to OFDM system without PRS) and the 3-th line from the bottom up (the black line) refers to OFDM system, using a PRS polynomial(-3,2,1).

When ε = 0,2 and SNR is greater than 5 dB improvement is observed if PRS coding is used. This is evident on the chart - the third pair of curves (blue and black) from the bottom up.

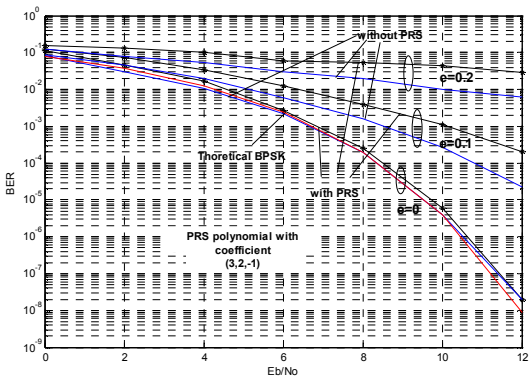


Fig. 4. Study of BER of PRS - OFDM system using a polynomial with coefficients (3 2 -1)

For Fig.4 when ε = 0,1 and SNR is greater than 7 dB there is an improvement of BER, if PRS coding is used. This is evident on the chart (the 4-th line from the bottom up (the blue line) refers to OFDM system without PRS) and the 3-th line from the bottom up (the black line) refers to OFDM system, using a PRS polynomial (4, -3, -1)

When ε = 0,2 and SNR is greater than 4 dB improvement is observed if PRS coding is used. This is evident on the chart - the third pair of curves (blue and black) from the bottom up

IV. Conclusion

The using of MLSD is a way to improve BER performance by PRS signals detection. After number of experiments, selection of polynomials set, used in OFDM systems with PRS is possible. A significant improvement in CIR, compared to conventional systems has been carried out. The results of simulation experiments have shown that PRS-OFDM system, using polynomials with coefficients (4, -3, -1), (-3, 2, 1), (3, 2, -1) in the PRS simulation model, give an improvement of CIR compared with the OFDM system without PRS.

References

[1] P. Kabal and S. Pasupathy, "Partial response signaling," IEEE Trans. Commun., vol. COM-23, pp, 921-925, Sept. 1975

[2] Proakis, J.G, "Digital Communications", 2"d Edition, McGraw-Hill, l989.

[3] Zhao Y.,Jean-Damien L., and Sven-Gustav H.,"Intercarrier Interference Compression in OFDM Comminication Systems by using Correlative Coding",IEE Commun. Lett.,vol 2,pp.1089-7798,August 1998

[4] M. Russell and G. J. Stuber, "Interchannel interference analysis of OFDM in a mobile environment," in Proc. IEEE Vehicular Technology Conf., 1995, pp. 820–824.

[5] Hua Zhang and Ye (Geoffrey) Li "Optimum Frequency-Domain Partial Response Encoding in OFDM System" IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 51, NO. 7, JULY 2003

[6] EE 379B - Digital Communication II: Coding; John M. Cioffi , http://www.stanford.edu/group/cioffi/

[7] Kyoung-Young Song, Jae-Dong Yang, Xianglan Jin, and Jong-Seon No "Quadrature Partial Response Signaling Based on Alamouti Code " ISIT 2009 Seoul Korea,June 28-Juli 3,2009

[8] Kolev St. "A New Wireless PRS-OFDM Simulation Model" CEMA'10, Athens, 2010

A New Modified Algorithm for Multi Constraint Routing

Yavor Tomov[1] and Georgi Iliev[2]

Abstract – **In modern telecommunications the necessity of supporting a large number of broadband services requires a guaranteed QoS routing. The basic task of routing is to find a path from source to destination. This path should satisfy these requirements. When the needed QoS depends on more than one parameter, the problem is known as Multi Constraint problem (MCP). Our goal in this article is to propose a new modified heuristic algorithm able to solve the problem.**

Keywords – **QoS, MCP, Routing.**

I. INTRODUCTION

To support broadband services such as the real time services, the QoS routing is the major problem. Routing has to deal two basic task – to support and update the information of the network and to find a path from source to destination. In cases when an application has no special requirements for the parameters (bandwidth, delay, jitter, packet loss) which define the QoS, finding the path from source to destination is known as The Shortest Path Problem. In other cases when two or more parameters have to be guaranteed from source to destination the problem is known as Multi Constraint Problem (MCP). MCP is NP hard [1] Depending on these parameters they may be classified as additive, multiplicative or concave. The paper is about the additive parameters, where the weight of this parameter (end to end) is equal to the sum of the weights of all links on the path. A path which satisfies all constraints is called a feasible path. There exist two basic routing strategies [2]: source routing and distributed routing. The source routing calculates the entire path locally in the source node. The distributed routing calculates the path in the intermediate note between the source and the destination.
The algorithms able to solve the MCP can generally be classified as heuristic and exact. The heuristic algorithms do not guarantee finding a feasible path(if it exists), while the exact algorithms do it. For this reason they are not applicable in practice. They are usually used to evaluate heuristic algorithms.
Our goal in this paper is to propose a new heuristic, modified, distributed routing algorithm, able to solve the MCP with two constraints.

Notation
Each real network can be presented as a graph G(V, E) where V is the set of nodes and E is the set of links. The nodes are routers , while the links are physical or logical connections between them. Each link $e \in E$ is associated with two – dimensional link vector $\vec{w}(w_1, w_2)$. The paths are noted as p, the source node as s, the destination node as t, any intermediate node as u. A path is feasible if (1) and (2) are correct.

$$w_1(p) = \sum_{e \in p} w_1 \le C_1 \qquad (1)$$

$$w_2(p) = \sum_{e \in p} w_1 \le C_2 \qquad (2)$$

where C_1 and C_2 are the constraints.

II. RELATED WORK

In the proposed algorithm, we present the two weights on each link as a linear combination. There exist many algorithms using this technique. They are known as algorithms with mixed metrics. Some of these algorithms use linear path length [3][4][5][6][7], others use non - linear path length[8][9]. The main advantage of linear path length algorithms is that they can implement Dijkstra's algorithm [10], while their major drawback is that they can return a path, which is not feasible (path outside the feasible region).
The major advantage of the algorithms that use the non - linear path length is that they can scan the feasible region precisely. Their drawback however is that subsections of shortest path are not always shortest paths. As a result of using Dijkstra's algorithm they are likely to fail. Fig.1 shows the way the two types of algorithms scan the feasible region. Fig 2 shows that the algorithm with linear path length works, Fig 3.shows that the algorithm fails.

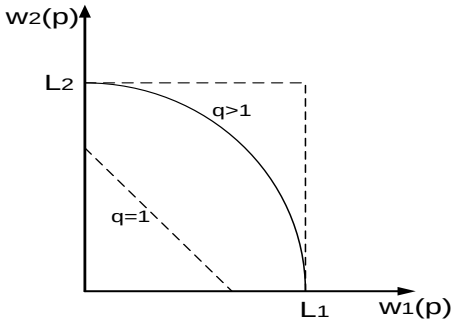


Fig. 1. Two types of algorithms scan the feasible region [3][8][9]

[1]Yavor Tomov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: qvor_tomov@abv.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

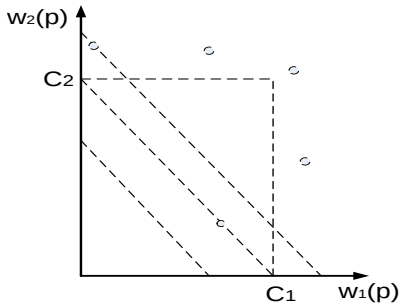


Fig. 2. The algorithm works [3]

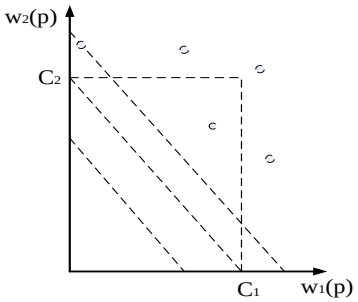


Fig. 3. The algorithm fails [3]

The first who proposed each link to be presented as linear combination was Jaffe[3].

$$w(e) = d_1 w_1 + d_2 w_2 \quad (3)$$

Where d_1 and d_2 are Lagrange multipliers.

Iwata [4] suggests a heuristic algorithm to solve MCP This algorithm computes the shortest path, based on the first parameter and checks whether all constraints are satisfied. If they are - the algorithm returns this path. If they are not - the algorithm finds the shortest path with respect to the second parameter and this is repeated until a feasible path is found.
In [5] the author proposed to reduce the search space increasing in this way the probability to find a solution.
Feng [6] proposed an algorithm based on [5],reducing the search space and implementing the k-th shortest path, suggested by Chong[11].
Based on these algorithms we worked out a heuristic algorithm able to deal with MCP in cases with two constraints.

III. OUR ALGORITHM

In this Section, we first prove an important theorem.
Theorem: If we use Dijkstra's algorithm to minimize the linear cost function on a graph that contains at least one feasible path, the algorithm returns a path where at least one $w(p) \le C$.

We assume that there is at least one feasible path p^* Using Dijkstra's algorithm, it returns the path p that minimizes the following linear cost function

$$g(p) = w_1 + kw_2 \quad (4)$$

where k is positive multiplier. This path however is not feasible

$$g(p) < g(p^*) \quad (5)$$

$$w_1(p) + kw_2(p) < w_1(p^*) + kw_2(p^*) \quad (6)$$

Since p* is feasible path it follows

$$w_1(p) \le C_1 \text{ and } w_2(p) \le C_2 \quad (7)$$

From (6) and (7) follow that at least one

$$w(p) \le C \quad (8)$$

With (8) we prove this basic theorem.

Our basic idea in this algorithm is to minimize two linear cost functions for each intermediate node – one from u to s and the other from u to t. First the algorithm finds the shortest paths from s to t with respect to w_1 and w_2. Then the algorithm defines k multipliers, as follows:

$$k_1 = \frac{C_1 - w_1(p_1)}{C_2 - w_2(p_2)} \quad (9)$$

The node s shares this value with all other nodes. Each of the intermediate nodes calculates two paths-one from u to s, and another – from u to t.These paths are calculated by the nodes using Dijkstra's algorithm to minimize the cost function

$$g(p) = w_1 + k_1 w_2 \quad (10)$$

If some of the nodes finds a feasible path, it returns it to s, if not according to our theorem each path contains not more than one $w > C$. In this case each of the nodes checks whether w_1 or w_2 does not satisfy the constraint. If it is w_1 - then the node decreases k_1 to k_2. If w_2 does not satisfy the constraint – the node increases k_1 to k_2.
In the next step each node computes the shortest path from u to s with respect to (10),and the path from u to t with respect to

$$g(p) = w_1 + k_2 w_2 \quad (11)$$

If the path

$$p(s \to t) = p(u \to s) + p(u \to t) \quad (10)$$

satisfies the constraint, the node returns this path. If not – the node changes the position of the two cost functions.

PSEUDO CODE

1. Dijkstra $w_1(s \to t) \to w_1(p_1)$

2. **IF** $w_1(p_1) \le C_1$ AND $w_2(p_1) \le C_2$
3. **RETURN path**
4. **END IF**
5. Dijkstra $w_2(s \to t) \to w_2(p_2)$
6. **IF** $w_1(p_2) \le C_1$ AND $w_2(p_2) \le C_2$
7. **RETURN path**
8. **END IF**
9. **IF** $w_1(p_1) > C_1$ OR $w_2(p_2) > C_2$
10. No Solution
11. . **END IF**
12. $k_1 = \dfrac{C_1 - w_1(p_1)}{C_2 - w_2(p_2)}$
13. **FOR** each link
14. $w(e) = w_1 + k_1 w_2$
15. **END FOR**
16. Dijkstra $p(s \to t)$
17. **IF** $w_1(p) \le C_1$ AND $w_2(p) \le C_2$
18. **RETURN path**
19. **END IF**
20. **FOR** each node u
21. Dijkstra $p(u \to t, k_1)$;
22. Dijkstra $p(s \to u, k_1)$
23. $p(s \to t) = p(s \to u) + p(u \to t)$
24. **IF** $w_1(p) \le C_1$ AND $w_2(p) \le C_2$
25. **RETURN path**
26. **ELSE**
27. **IF** $w_1(p) \le C_1$ AND $w_2(p) > C_2$
28. $k_2 > k_1$
29. **DO**
30. Dijkstra $p(u \to s, k_2)$,
31. Dijkstra $p(u \to t, k_1)$
32. $p(s \to t) = p(u \to s, k_2) + p(u \to t, k_1)$
33. **UNTIL** $w_2(p) \le C_2$
34. **IF** $w_1(p) \le C_1$ AND $w_2(p) \le C_2$
35. **RETURN path**
36. **ELSE**
37. Dijkstra $p(u \to s, k_1)$
38. Dijkstra $p(u \to t, k2)$
39. $p(s \to t) = p(u \to s, k_1) + p(u \to t, k_2)$
40. **IF** $w_1(p) \le C_1$ AND $w_2(p) \le C_2$
41. **RETURN path**
42. END IF
43. END IF
44. END IF
45. END IF
46. **IF** $w_1(p) > C_1$ AND $w_2(p) \le C_2$
47. $k_2 < k_1$
48. **DO**
49. Dijkstra $p(u \to s, k_2)$,
50. Dijkstra $p(u \to t, k_1)$
51. $p(s \to t) = p(u \to s, k_2) + p(u \to t, k_1)$
52. **UNTIL** $w_2(p) \le C_2$
53. **IF** $w_1(p) \le C_1$ AND $w_2(p) \le C_2$
54. **RETURN path**
55. **ELSE**
56. Dijkstra $p(u \to s, k_1)$
57. ,Dijkstra $p(u \to t, k2)$
58. $p(s \to t) = p(u \to s, k_1) + p(u \to t, k_2)$
59. **IF** $w_1(p) \le C_1$ AND $w_2(p) \le C_2$
60. **RETURN path**
61. END IF
62. END IF
63. END IF
64. END IF
65. END FOR

From line 1 to line 8 we implement Iwata's algorithm. On line 12 the algorithm calculates k_1 based on LARAC [5] algorithm. From line 20 to the last line the algorithm contains our basic idea which we exposed above.

EXAMPLE

Let we have a graph with six nodes. Each link is associated with two weights - w_1 and w_2. Our task is to find a path from source A to destination F with the given constraints $C_1 = 5$ and $C_2 = 10$ Fig. 4. There are four paths from A to F: ABDF=6, 3; ACEF=4, 11; ABCEF=5, 10; ACBDE=9, 6. If we apply only one linear cost function (like in most algorithms) the algorithm returns the path ABDF or the path ACEF(according to the choice of k)as solutions.

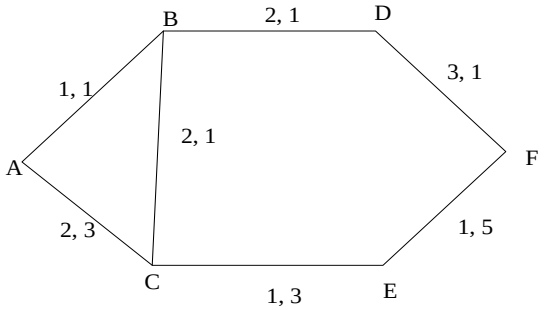


Fig. 4. Graph with two weights on each link

However these solutions are not correct.
Applying our algorithm it will first find the shortest paths from s to t with respect to w_1 and w_2 These paths are ACEF and ABDF. Both paths however are not solution. Next the algorithm will calculate $k = \frac{1}{7}$ based on (9). It will find again the shortest path based on (4). This path is ACEF, but it is not a solution either. In the next step the nodes C and B will calculate the path p based on (12), by $k = \frac{1}{7}$. For node B the path will be ABDF while for node C the path will be ACEF. The two nodes will check which one (w_1 or w_2) does not satisfy the constraint. Node B will establish that this is w_1, while for node C it is w_2. Then the node will decrease k and when $k \le \frac{1}{8}$ with respect to the path from B to F, it will find the path BCEF. The node will return the path ABCEF which is the feasible path. The procedure is the same concerning node C. When $k \ge \frac{11}{10}$ for the path from C to A the node will return the path ABCEF.

IV. Conclusion

We have created an algorithm that is based on five concepts:

1. The algorithm is distributive
2. Iwata's algorithm is implemented
3. Reduced search space[5]
4. Look – ahead approach[12][13]
5. Two linear functions for each intermediate node

In some algorithms with linear path length[6],as well as in other with non linear path length[8][9], a solution is searched applying the concept of k – th shortest path. This however leads to considerable increase in complexity of the algorithm.
We suggest another approach. Each of the intermediate nodes explores a set of paths that could be feasible. This idea is similar to the idea proposed by Korkmaz in his algorithm [14]. He uses one linear and one non linear function, while we use two linear function.
In the example looked through it becomes clear that Iwata's and LARAC algorithms are not able to solve the problem of finding the feasible path, while our algorithm finds it successfully

.Acknowledgement

For the acknowledgement use the unnumbered section layout.

References

1. M.R. Garey and D.S. Johnson. Computers and Intractability: A guide to the Theory of NP-Completeness. ISBN 0-7167-1044-7. W.H. Freeman and Company,San Francisco, 1979
2. S. Chen, K. Nahrstedt, An overview of quality of service routing for next-generation high-speed networks: problems and solutions, IEEE Networks 12 (6) (1998) 64-79
3. J.M. Jaffe. Algorithms for finding paths with multiple constraints. Networks,14:95—116, 1984
4. A. Iwata, R. Izmailov, D. Lee, B. Sengupta, G.Ramamurthy, and H. Suzuki. ATM routing algorithms with multiple QoS requirements for multimedia internetworking. IEICE Trans. Commun., E79-B(8):999—1007, Aug. 1996
5. A.Juttner, B. Szviatovszki, I. Mecs, Z. Rajko, "Lagrange relaxation based method for the QoS routing problem", Proceedings of the INFOCOM 2001 Conference, vol. 2, IEEE, 2001, pp. 859–868
6. Gang Feng, " Exact algorithms for multi-constrained QoS routing ," International Conference on Computer, Communication and Control Technologies (CCCT'03), Orlando, USA, Vol. 2, pp. 340 – 345, July 31 – Aug 2, 2003
7. P.Khadivi, S.Samavi, T.D.Todd, "Multi-constraint QoS routing using a new single mixed metrics", Journal of Network and Computer Applications, Vol. 31, Issue 4, pp 656-676, Nov. 2008.
8. H. De Neve and P. Van Mieghem. TAMCRA: a tunable accuracy multiple constraints routing algorithm. Computer Communications, 23:667—679, 2000.
9. Van Mieghem P. and F. A. Kuipers, "Concepts of Exact Quality of Service Algorithms," IEEE/ACM Trans. Net Vol. 12, pp. 851 - 862, 2004
10. E.W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, (1):269—271, 1959.
11. E.I. Chong, S. Maddila, and S. Morley. On finding single-source single-destination k shortest paths. Journal of Computing and Information, special issue ICCI'95, pages 40—47, July 1995.
12. S. Lin, "Computer solutions of the traveling salesman problem," Bell Syst. Tech. J., vol. 44, no. 10, pp. 2245–2269, 1965.
13. A. Newell and G. Ernst, "The search for generality," in Proc. IFIP Congr., vol. 1, 1965, pp. 17–24.
14. T. Korkmaz, M. Krunz, "Multi-constrained optimal path selection", Proceedings of the INFOCOM 2001 Conference, vol. 2, IEEE, Anchorage, Alaska, 2001, pp. 834–843.

A Schema Based Approach to Access Network Discovery and Selection in EPS

Ivaylo Atanasov[1]

***Abstract* – Access Network Discovery and Selection Function (ANDSF) defined in the architecture of the Evolved Packet System (EPS) allows the operator to control the user choice of access technology if several non-3GPP access networks are available. The 3GPP specifications define the management objects that can be used by ANDSF and the user equipment (UE), but do not define the structure and format of user capabilities that the UE may provide in order to support the rules generation by the ANDSF. The paper provides a structural approach to definition of the information that the UE may send to the ANDSF. The information is defined using XML schemes.**

***Keywords* – Evolved Packet System, Access Network Discovery and Selection, User equipment capabilities.**

I. INTRODUCTION

Evolved Packet System (EPS) is the name for the evolution of mobile networks [1]. EPS covers the radio access, the core network and the user equipments that comprise the overall mobile system. Along with the high-speed 3GPP radio access networks, EPS supports also non-3GPP accesses such as WLAN, WiMAX and wired access.

When a user equipment (UE) is registered in the home network and when both 3GPP and non-3GPP accesses are available, or when multiple non-3GPP accesses are available, the EPS network may provide the UE with assistance data/policies about available accesses. The EPS network allows the operator to influence the access that the UE shall handover to (when in active mode) or re-select (when in idle mode). The architecture that may be used for access network discovery and selection is based on a new network element called Access Network Discovery and Selection Function (ANDSF). The ANDSF contains data management and control functionality necessary to provide network discovery and selection assistance data as per operators' policy [2].

ANDSF is in an initial stage of standardization. Only the minimal interface to the mobile equipment is defined [3]. The definition on how the ANDSF should work is limited and there are no connections between ANDSF and other entities. Further, the ANDSF relies only on its own information for decisions. The following extensions are proposed in [4]. First, the ANDSF decisions should be based on subscription. Second, the ANDSF decisions should use dynamic access network information and the ANDSF should support from the always-best-connected enabler towards the services. Third, the ANDSF should provide the means to integrate femto-accesses for network discovery and selection.

The idea of using dynamic access network information is developed in [5], [6] and [7]. In [5], the authors propose a solution that combines ANDSF and MIH (Media Independent Handover (MIH) Service standard 802.21 from IEEE) for improving the inter-system handover behavior. In [6], the authors propose novel access reselection procedures which enable a network provider to optimize the allocation of the users on the different access networks available using as central concept that handovers can and should be triggered by the modifications on the resources required by the mobile devices in order to optimize the overall usage of the wireless environment. A multilink architecture with evolved ANDSF and Policy and Charging Rule Function is suggested in [7].

3GPP defines management objects that can be used by the ANDSF and the UE [3]. The Management Object (MO) is compatible with the OMA Device Management (DM) protocol specifications, version 1.2 and upwards. It is defined using the OMA DM Device Description Framework (DDF) as described in the Enabler Release Definition OMA-ERELD-DM-V1_2 [8]. The MO consists of relevant parameters for inter-system mobility policy and access network discovery information that can be managed by the ANDSF.

In this paper, we propose a structural approach to definition of information that the UE may send to the ANDSF in order to support the ANDSF decisions. The proposed information structure is based on the MO information for UE location as defined in [3] and on UE capabilities. The definitions are presented using XML schemes.

The paper is structured as follows. Section II provides an analysis on required information about UE capabilities and defines the respective XML schemes. In Section III, XML schemes are defined for the information related to UE location. Section IV presents an example of the content of a client initiated session alert message of code "Generic Alert" which UE may use to initiate the information provision by the ANDSF.

II. UE CAPABILITIES INFORMATION STRUCTURE

The UE needs to provide to the ANDSF information about its capabilities and its location. The UE capabilities describe the supported access technologies and the routing capabilities. The related XML scheme is shown in Fig.1.

The UE may obtain IP connectivity by attaching to 3GPP access technologies, 3GPP2 access technologies and non-3GPP access technologies. The UE may support more than one access technologies.

Each access technology is designed to operate in particular frequency bands. The UE routing capabilities define the mechanism for routing of different IP flows. The UE is classified in a given category depending on the supported

[1]Ivaylo Atanasov is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: iia@tu-sofia.bg

maximum data rates. The 3GPP access technologies that may be supported by the UE include GSM/GPRS, WCDMA/HSPA and LTE. The 3GPP technologies that may be supported by the UE include 1xRTT and HRPD. The non-3GPP2 technologies that may be supported by the UE include WiMAX and WLAN. Fig.3 shows the related XML scheme.

```
<xs:element name="UE">
 <xs:complexType>
  <xs:element name="capabilities">
   <xs:complexType>
     <xs:element name="technologies"
           type="supportedTchnologies"/>
     <xs:element name="routingCapabilities"
           type="routingCapabilitiesType"/>
   </xs:complexType>
  </xs:element>
   <xs:element name="UELocation" type="UELocationType"/>
 </xs:complexType>
</xs:element>
```

Fig.1 XML scheme for UE information

The XML scheme related to UE supported technologies is shown in Fig.2.

```
<xs:complexType name="supportedTechnologies">
  <xs:sequence>
     <xs:element name="supportedTechnology"
          type="supportedTechnologyType" minOccurs="1"
          maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="supportedTechnologyType">
  <xs:sequence>
    <xs:element name="technology" type="technologyType"/>
    <xs:element name="bands" type="bandsInfo">
    <xs:element name="category" type="xs:positiveInteger"/>
  </xs:sequence>
</xs:complexType>
<xs:simpleType name="technologyType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="WCDMA/HSPA"/>
    <xs:enumeration value="GSM/GPRS"/>
    <xs:enumeration value="LTE"/>
    <xs:enumeration value="WiMAX"/>
    <xs:enumeration value="WLAN"/>
    <xs:enumeration value="1xRTT"/>
    <xs:enumeration value="HRPD"/>
  </xs:restriction>
</xs:simpleType>
```

Fig.2 XML scheme for UE supported access technologies

IFOM (IP Flow Mobility) capable UE is an UE that is capable of routing different IP flows to the same PDN (Packet Data Network) connection through different access networks. The IFOM capable UE may simultaneously connect to 3GPP access and WLAN and exchange different IP flows belonging to the same PDN connection through different accesses.

Non-seamless WLAN offload capable UE is an UE that is capable of non-seamless WLAN offload. An UE supporting non-seamless WLAN offload may, while connected to WLAN access, route specific IP flows via the WLAN access without traversing the Evolved Packet Core (EPC).

MAPCON (Multi Access PDN Connectivity) capable UE is a UE that is capable of routing different simultaneously active PDN connections through different access networks.

A UE that is not capable of routing IP traffic simultaneously over multiple radio access interfaces (e.g. a non-IFOM or non-MAPCON capable UE, or an UE that has such a capability disabled, or an UE not capable of non-seamless WLAN offload) selects the most preferable available access network for inter-system mobility based on the received/provisioned inter-system mobility policies and user preferences and disregards the inter-system routing policies it may have received from the ANDSF. The XML scheme related to routing capabilities is shown in Fig.3.

```
<xs:simpleType name="routingCapabilities">
  <xs:restriction base="xs:string">
    <xs:enumeration value="IFOM"/>
    <xs:enumeration value="NonSeamlessOffload"/>
    <xs:enumeration value="MAPCON"/>
    <xs:enumeration value="NoInter-systemRouting"/>
  </xs:restriction>
</xs:simpleType>
```

Fig.3 XML scheme for routing capabilities

The access technology operating bands define the frequency allocation for uplink and downlink as well as the duplex mode of operation. The UE may support more than one frequency bands. The related XML scheme is shown in Fig.4.

```
<xs:complexType name="bandInfo">
  <xs:sequence>
    <xs:element name="low" type="xs:decimal"/>
    <xs:element name="high" type="xs:decimal"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="bandType">
  <xs:sequence>
     <xs:element name="bandID" type="xs:positiveInteger"/>
     <xs:element name="uplink" type="bandInfo"/>
     <xs:element name="downlink" type="bandInfo"/>
     <xs:element name="duplex" type="duplexMode"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="bandInfo">
  <xs:sequence>
    <xs:element name="low" type="xs:numeric"/>
    <xs:element name="high" type="xs:numeric"/>
  </xs:sequence>
</xs:complexType>
<xs:simpleType name="duplexMode">
    <xs:restriction base="xs:string">
      <xs:pattern value="FDD|TDD"/>
    </xs:restriction>
</xs:simpleType>
```

Fig.4 XML scheme for UE supported operating bands information

III. UE LOCATION INFORMATION

The UE location node acts as a placeholder for location description; an UE therefore inserts information regarding all of the access networks that it can discover into this node. If the UE is aware of its geographical position, it shall insert the coordinates of its position in this GeoLocation element. Fig.5 shows the XML scheme related to UE location.

```
<xs:complexType name="UELocationType">
  <xs:sequence>
    <xs:element name="3GPPLocation" type="3GPPLocations"
       minOccurs="0" maxOccurs="1"/>
    <xs:element name="3GPP2Location"
       type="3GPP2Locations" minOccurs="0" maxOccurs="1"/>
    <xs:element name="WiMAXLocation"
       type="WiMAXLocations" minOccurs="0" maxOccurs="1"/>
    <xs:element name="WLANLocation" type="WLANLocations"
       minOccurs="0" maxOccurs="1"/>
    <xs:element name="GeoLocation" type="GeoLocationType"
       minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
</xs:complexType>
```

Fig.5 XML scheme for UE location information

For 3GPP networks, the access network information includes: PLMN-id, Tracking Area Code, Location Area Code and Cell Global Identity. The PLMN-id is defined by Mobile Country Code and Mobile Network Code. Fig.6 shows the XML scheme for UE location information in 3GPP access.

```
<xs:complexType name="3GPPLocations">
  <xs:sequence>
    <xs:element name="location" type="3GPPLocationType"
minOccurs="1" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="3GPPLocationType">
  <xs:sequence>
    <xs:element name="PLMN" type="xs:string"/>
    <xs:element name="TAC" type="xs:unsignedShort"
          minOccurs="0" maxOccurs="1"/>
    <xs:element name="LAC" type="xs:unsignedShort"
          minOccurs="0" maxOccurs="1"/>
    <xs:element name="GERAN_CI" type="xs:hexBinary"
          minOccurs="0" maxOccurs="1"/>
    <xs:element name="UTRAN_CI" type="xs:hexBinary"
          minOccurs="0" maxOccurs="1"/>
    <xs:element name="EUTRA_CI" type="xs:hexBinary"
          minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
</xs:complexType>
```

Fig.6 XML scheme for 3GPP location

For 3GPP2 1xRTT networks, the SID (System Identification code), NID (Network Identification code) and Base ID (Base Station Identification code) for one particular 3GPP2 1xRAT location are used to indicate UE position. For 3GPP2 HRPD networks, the Sector_ID (Sector Identification code), Netmask (Netmask code) for one particular 3GPP2 HRPD RAT location are used to indicate UE position. Fig.7 shows the XML scheme for UE location information in 3GPP2 access.

For WiMAX networks, UE Location includes: NAP-ID and BS-ID. The NAP-ID (Network Access Provider) and BS-ID (BS identifier) for a particular WiMAX network are used to provide access network information. For WLAN networks, HSSID, SSID and BSSID are captured. Fig.8 shows the XML scheme for UE location information in non-3GPP access.

The GeoLocation node acts as a placeholder for one or more UE geographical location descriptions. Fig.9 shows the related XML scheme.

In order to illustrate the usage of the proposed information structure, the following section presents an example.

```
<xs:complexType name="3GPP2Location">
  <xs:sequence>
    <xs:element name="1xLocation" type="1xLocations"
          minOccurs="0" maxOccurs="1"/>
    <xs:element name="HRPDLocation" type="HRPDLocations"
          minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="1xLocations">
  <xs:sequence>
    <xs:element name="location" type="1xLocationType"
          minOccurs="1" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="1xLocationType">
  <xs:sequence>
    <xs:element name="SID" type="xs:hexBinary"/>
    <xs:element name="NID" type="xs:hexBinary"
          minOccurs="0" maxOccurs="1"/>
    <xs:element name="Base_ID" type="xs:hexBinary"
          minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="HRPDLocations">
  <xs:sequence>
    <xs:element name="location" type="HRPDLocationType"
          minOccurs="1" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="HRPDLocationType">
  <xs:sequence>
    <xs:element name="Sector_ID" type="xs:hexBinary"/>
    <xs:element name="Netmask" type="xs:hexBinary"/>
  </xs:sequence>
</xs:complexType>
```

Fig.7 XML scheme for 3GPP2 location

```
<xs:complexType name="WiMAXLocations">
  <xs:sequence>
    <xs:element name="location" type="WiMAXLocationType"
          minOccurs="1" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="WiMAXLocationType">
  <xs:sequence>
    <xs:element name="NAP-ID" type="xs:hexBinary"/>
    <xs:element name="BS-ID" type="xs:hexBinary"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="WLANLocationType">
  <xs:sequence>
    <xs:element name="BSSID" type="xs:hexBinary"/>
    <xs:element name="SSID" type="xs:hexBinary"
minOccurs="0" maxOccurs="1"/>
    <xs:element name="HESSID" type="xs:hexBinary"
minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
</xs:complexType>
```

Fig.8 XML scheme for non-3GPP location

```
<xs:complexType name="GeoLocationType">
  <xs:sequence>
    <xs:element name="AnchorLongitude" type="xs:decimal"/>
    <xs:element name="AnchorLatitude" type="xs:decimal"/>
  </xs:sequence>
</xs:complexType>
```

Fig.9 XML scheme for UE geographical location

IV. Example of User Data Provided to ANDSF

The UE may initiate information provisioning from the ANDSF, using a client initiated session alert message of code "Generic Alert". The "Type" element of the OMA DM generic alert message shall be set to "urn:oma:at:ext-3gpp-andsf:1.0:provision-multiple-if" as we assume that the UE is configured for IFOM. In our example, the UE supports both LTE and WiMAX access network technology. For LTE, the UE supports operating bands of 1447.8 MHz -1462.9 MHz for uplink and 1495.9 MHz -1510.9 MHz for downlink and FDD mode. For WiMAX the UE supports operating bands of 3527.5MHz -3562.5 MHz and TDD mode. The UE category in LTE is 5, that means that UE supports 4x4 MIMO, data rates of 300 Mbps for downlink and 75 Mbps for uplink, and the supported modulation format is 64QAM. The UE location in LTE access network is identified by PLMN-id of 28402, tracking area code of 0xD34F and E-UTRA cell identity of 0xAF52D10<EUTRA. In WiMAX, the UE location is identified by NAD-ID of 0xF4700A and BS-ID of 0xCA9912.

Fig.10 shows the XML description of information about UE capabilities and location used in the example for access network discovery and selection.

V. Conclusion

The paper provides a structural approach to definition of information that an UE may send to the ANDSF in order to support the decision for access network discovery and selection. The suggested information contains both UE capabilities and location. The UE capabilities include information about supported access network technologies and routing capabilities. The XML schemes defining this information can be used to express a set of rules to which the UE provided information must conform in order to be considered 'valid' according to that schema. The suggested XML schemes are designed with the intent that determination of the UE provided information validity would produce a collection of information adhering to specific data types.

Acknowledgement

The work is conducted under the grant of the Project 122pd0007-07 funded by Research and Development Sector, TU-Sofia, Bulgaria.

References

[1] 3GPP TS 23.002, Network Architecture, v11.1.0, 2011.
[2] 3GPP TS 23.402, Architecture enhancements for non-3GPP accesses, v11.1.0, 2011-12.
[3] 3GPP TS 24.312, Access Network Discovery and Selection Function (ANDSF) Management Object (MO), v11.1.0, 2011.
[4] M. Corici, J. Fiedler, T. Magedanz, D. Vingarzan. "Access Network Discovery and Selection in the Future Broadband Wireless Environment", Retrieved from http://www.slideshare.net/zahidtg/access-network-discovery-and-selection-in-the-future-broadband-wireless-environment, Accessed 2012.
[5] S. Frei, W. Fuhrmann, A. Rinkel, B. V. Ghita. "Improvements to Inter-system Handover in the EPC Environment", 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, Conference proceedings pp 1-5, 2011.
[6] M. Corici, T. Magedanz, D. Vingarzan, C. Pampu, Q. Zhou, "Access Network Reselection based on Momentary Resources in a Converged Wireless Environment", IEEE Globecom'2010 - Next Generation Networking Symposium, Miami, USA, Conference proceedings, 2010.
[7] H. Lonsethagen et al. Multilink network architecture, Report, Celtic project CP5-013, MARCH – Multilink architecture for multiplay services, http://projects.celtic-initiative.org/march/march/UserFiles/file/CP5-013-MARCH-D5_2-final.pdf Accessed 2012.
[8] OMA-ERELD-DM-V1_3: "Enabler Release Definition for OMA Device Management", 2009.

```
<UE>
 <capabilities>
  <technologies>
   <supportedTechnology>
    <technology>LTE</technology>
    <bands>
     <band>
      <bandID>21</bandID>
<uplink><low>1447.9</low><high>1462.9</high></uplink>
<downlink><low>1495.9</low><high>1510.9</high></downlink>
      <duplex>FDD</duplex>
     </band>
    </bands>
    <category>5</category>
   </supportedTechnology>
   <supportedTechnology>
    <technology>WiMAX</technology>
    <bands>
     <band>
      <bandID>2</bandID>
<uplink><low>3527.5</low><high>3562.5</high></uplink>
<downlink><low>3527.5</low><high>3562.5</high></downlink>
      <duplex>TDD</duplex>
     </band>
    </bands>
    <category>1</category>
   </supportedTechnology>
  </technologies>
  <routingCapabilities>IFOM</routingCapabilities>
 </capabilities>
 <UELocation>
  <3GPPLocation>
   <location>
    <PLMN>28402</PLMN>
    <TAC>0xD34F</TAC>
    <EUTRA_CI>0xAF52D10<EUTRA_CI>
   </location>
  </3GPPLocation>
  <WiMAXLocation>
   <location>
    <NAP-ID>0xF4700A</NAP-ID>
    <BS-ID>0xCA9912</BS-ID>
   </location>
  </WiMAXLocation>
 </UELocation>
</UE>
```

Fig.10 An example of UE capabilities and location required for access network discovery and selection

System for thermal comfort monitoring in working and living environment

Uros Pesovic[1], Dusan Markovic[2], Zeljko Jovanovic[1], Sinisa Randjic[1]

Abstract – **Thermal comfort of working and living space is dependent on a number of parameters, both personal and environmental parameters. Knowledge of index of thermal comfort, helps creating optimal micro-climatic conditions for unhindered work in a workspace, or customizes the way employees dress in accordance with the conditions prevailing in the environment. This paper introduces a system for monitoring thermal comfort in working and living space, based on the concept of smart transducers on IEEE 1451 standard.**

Keywords – **Thermal comfort, IEEE 1451, TEDS, NCAP, Web Service.**

I. INTRODUCTION

Thermal comfort represents set of micro-climatic conditions for which human feels comfortable in its work and living environment. Thermal comfort is dependant both form personal characteristics and environmental conditions [1]. Personal factors are dependent of human metabolic activity, age, sex and human physical condition, as well as type of clothing. Environmental conditions which affect feeling of thermal comfort are not directly related to human. These conditions include air temperature, relative humidity, air speed and radiant temperature. Air temperature is parameter, which is most usually related to human feeling of thermal comfort since human can easily feel it on their skin. On the other hand, humans cannot detect changes in relative air humidity, which can significantly affect human feeling of thermal comfort. Human body uses perspiration as mechanism for loss of excess heat in order to regulate body temperature. When air has high relative humidity, water cannot evaporate from skin and body cannot lose excess heat so humans feel hotter than usual. When air is dry, humans feel colder since water easily evaporates from skin and cools the human body. Also, air movement increases body heat loss due convection, so in windy conditions humans feel colder than usual. Radiant temperature represents temperature of skin exposed to some radiation source, such as Sun or some artificial heat source.

Humans can in advance adjust their clothing in order to feel thermally comfortable when conducting some activities, by knowing environmental conditions present in certain area. In this paper, we present system for thermal comfort monitoring in working and living environment. System is composed of network of smart transducers which monitors air temperature and relative humidity and web based user application, which is accessible worldwide. Users can select level of clothing and certain activity which they plan to conduct in some area. Based on measured air temperature and relative humidity, system will return information about expected level of thermal comfort which will be experienced in such area.

Paper [2] presents similar solution in thermal comfort monitoring, which collects thermal comfort data by ZigBee network and personal computer, which is later forwarded to handheld PDA device. Such system is intended for localized use because connection between personal computer and PDA device is established using Bluetooth network.

[1]Uros Pesovic, Zeljko Jovanovic and Sinisa Randjic is with the Technical Faculty at University of Kragujevac, Svetog Save 65, Cacak 32000, Serbia, E-mail: pesovic@tfc.kg.ac.rs,

[2]Dusan Markovic is with the Faculty of Agronomy at University of Kragujevac, Cara Dusana 34, Cacak 32000, Serbia

II. THERMAL COMFORT

Thermal comfort is defined as that condition of human mind which expresses satisfaction with its thermal environment. Thermal comfort scale, defined by international standard ISO7730 [3], has seven point thermal comfort states which are shown in Table I.

TABLE I
THERMAL COMFORT SENSATION SCALE

Value	Description
+3	Hot
+2	Warm
+1	Slightly
0	Normal
-1	Slightly
-2	Cool
-3	Cold

Standard defines PMV (Predicted Mean Vote), average vote of thermal comfort on seven point thermal comfort sensation scale. Average vote of thermal comfort is created when heat balance of human body is achieved, which represents difference of created metabolic heat MW and sum of heat losses $\sum H_i$ which human body emits into environment Eq. (1).

$$PMV = T_S \cdot \left(MW - \sum H_i \right) \quad (1)$$

In order to create PMV from heat balance, transformation coefficient T_S is used, which is obtained by Eq. (2).

$$T_S = 0.303 \cdot e^{-0.036M} + 0.028 \quad (2)$$

Metabolic heat is difference between power of metabolic activity M and external work W. Heat losses originate from set of thermodynamics processes which can be classified into six categories:

1. Heat loss by skin conduction H_1
2. Heat loss by sweating H_2
3. Latent respiration heat loss H_3

4. Dry respiration heat loss H_4
5. Heat loss by radiation H_5
6. Heat loss by skin convection H_6

Metabolic power M is dependent of level of physical activity. Standard ISO8996 defines typical levels of human physical activities where some of them are presented in Table II. It defines metabolic index Met, which represents metabolic power relative to sitting as reference metabolic activity.

TABLE II
LEVELS OF METABOLIC ACTIVITY

Metabolic activity	W/m^2	Met
Reclining	46	0.8
Sitting	58	1
Office work	70	1.2
Standing	93	1.6
Light work	116	2.0
Walking 3km/h	140	2.4
Walking 4km/h	165	2.8
Walking 5km/h	200	3.4

Clothing can significantly affect human thermal comfort sensation. Standard ISO9920 defines thermal insulation coefficient index (Table III) for various types of clothing. Thermal insulation coefficient index is referenced to thermal insulation coefficient of human body without clothes.

TABLE III
CLOTHING THERMAL INSULATION INDEX

Clothing	m^{2}K/W	CLO
Without clothing	0	0
Summer clothing (shorts, T-shirt, sandals)	0.05	0.3
Summer clothing (trousers, shirt, shoes)	0.08	0.5
Spring clothing (trousers, shirt, jacket, shoes)	0.11	0.7
Winter clothing (trousers, sweeter, jacket, shoes)	0.2	1.3

PMV present mean value of the votes of a large group of people exposed to the same environment. Votes are scattered around the mean value where certain percentage of group will feel uncomfortably hot or cold in given conditions. PPD (Predicted Percentage Dissatisfied) represents percentage of people who will vote hot, warm, cool or cold on the 7-point thermal sensation scale, and it's obtained by Eq. (3)

$$PPD = 100 - 95 \cdot e^{-0.03353PMV^4 - 0.2179PMV^2} \quad (3)$$

III. IEEE 1451 SMART TRANSDUCERS

Sensors are used in wide range of applications, such as industrial and home automation, military, healthcare, security, agriculture and environment monitoring. Their basic function was transforming physical value into the measuring signal. According to rapid technological development transducers in addition to its basic functionality got other functions like self-calibration, self-description, self-initialization and some signal processing. All these functions contribute to certain independent processing or intelligence so they get name smart transducers. Connecting these transducers in the network is very important because of the fact that, in this way, collecting of measured signal was facilitated. First networking of transducers had difficulty due to different communication standards that is used in different producer's devices. This problem is solved with standards of IEEE 1451 family which gave standardized way for implementing components of smart transducers.

Concept of smart transducers introduced by IEEE 1451 standard consists of Transducer Interface Module (TIM) and Network Capable Application Processor (NCAP) (Figure 1.). A TIM interacts directly with environments and contains sensors that are used to get measured signal or actuators to control specified activities. Part of the TIM is Transducer Electronic Data Sheet (TEDS) with data, which is used to define characteristics of measurements. These data in TEDS were specified according to standard IEEE 1451.0 [4] which is also used to define commands that could be executed on TIM. NCAP is module that has function of gateway between TIMs and user application. NCAP has interface TSI (Transducer Services Interface) which is used by user application for direct access to function defined by this standard (Figure 1).

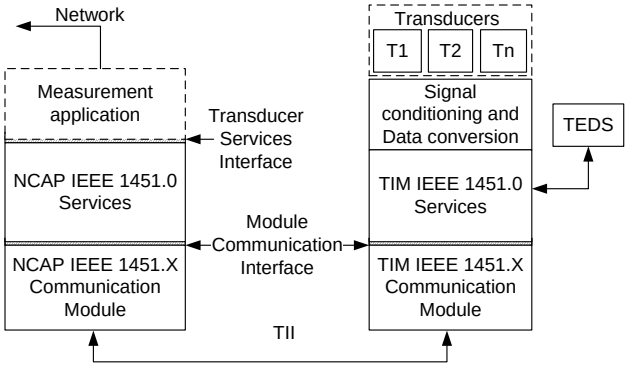


Figure 1. IEEE 1451.0 Reference model

Depending on the type of request from user application transducer services module is subdivided into five services: *TimDiscovery*, *TransducerAccess*, *TransducerManager*, *TedsManager*, *CommManager*, *AppCallback*. If there is a need for obtaining measured value from specific transducer channel *TransducerAccess* service could be used with his methods for accessing transducer channels, where the most common methods are read and write. TSI represents API used by user application to access IEEE 1451.0 layer via HTTP. User send HTTP request to NCAP where it is processed and call appropriate service of 1451.0 API. Service of TSI then call IEEE 1451.X API to establish communication with TIM and get result of measurement. HTTP server on the NCAP receives result from TIM and forms HTTP response that would be returned to end user. Response could be specified in HTML, XML or TEXT format.

IV. IMPLEMENTATION

System for thermal comfort monitoring is composed from smart transducer network and web based user application (Figure 2.). Transducer network, compliant to IEEE 1451 standard, is composed form two TIM modules and one NCAP module. Each TIM module has two sensors, one for air temperature and another for relative humidity. NCAP module acts as network gateway and provides communication interface for entire sensor network. It has integrated Web server, which is accessible via Ethernet network. User application is realized in form of web service and measured data represent to end user in adjustable form.

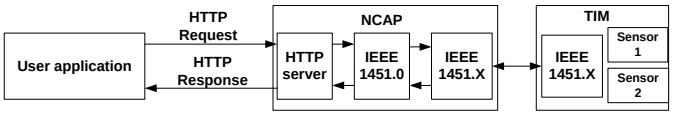


Figure 2. NCAP HTTP access

SHT15 digital sensor is used as a sensor for measurement of air temperature and relative humidity. It's composed form capacitive sensing element for measurement of relative humidity and bandgap sensor for temperature measurement. Each SHT15 sensor is calibrated in high precision humid chamber, where calibration coefficients are stored in sensor memory. SHT15 sensors communicate with microcontrollers using I^2C digital communication bus. It has low power consumption of 3 mW in operational mode and 5μW in sleep mode with wide range of supply voltage from 2.4 to 5.5 V [5]. Typical measurement characteristics of SHT15 sensor are presented in table IV.

TABLE IV
MEASUREMENT CHARACTERISTICS OF SHT15 SENSOR

Characteristics	Temperature	Relative humidity
Operational range	- 40 ÷ 124 ºC	0 ÷ 100 %
Resolution	± 0.01 ºC	± 0.05 %
Accuracy	± 0.3 ºC	± 2 %
Repeatability	± 0.1 ºC	± 0.1 %
Response time τ=0.63	5 s	8 s

NCAP module was implemented on *mBed* development board. This board uses NXP LPC1768 microcontroller, which is designed for low-power embedded applications. This microcontroller uses ARM Cortex-M3 uses 32-bit processor core, designed for wide range of application like industrial control systems, wireless networks and sensor integration. Processor has great performances in terms of processing power, fast response and also meets challenge of low power consumption. Besides ARM Cortex M3 core, this microcontroller contains 512KB FLASH memory, 32KB RAM memory, Ethernet, USB, CAN and other peripheral interfaces. *MBed* board was packaged in form of 40-pin DIP footprint which enables easy interaction to its environment. Programming and communication with computer is established via USB interface. Applications development for *mBed* board is possible using *mBed* Online compiler or using standard tools such as Keil uVision or Code Red [6].

Intention of this work was to develop NCAP module on the *mBed* platform for distribution of measuring data via HTTP protocol. In addition, Web service was created that could gather measured data, apply particular programming logic and offer complete results to end users. For NCAP, *mBed* board was equipped with Ethernet connector for connecting NCAP to local computer network. HTTP Web server was started on the *mBed* this HTTP request from outer computer network could be processed and response in one of the possible format could be returned. Method *ReadData* of the *TransducerAccess* interface was realized on the *mBed* board and these methods were used to send new request to TIM for reading measuring values. Measured results were obtained by two SHT15 sensors which are connected to the same *mBed* board, but represents independent logical component forming one TIM. Communication between NCAP and TIM is omitted because it is not part of the IEEE 1451.0 standard, and to facilitate prototype realization both logical module was situated on the same *mBed board.* Request for accessing NCAP on the *mBed* must be in the certain format defined by standard.

http://<host>:<port>/<path>?<parameters>
<host> - domain name of destination
<port> - port number (default value is 80)
<path> - IEEE 1451 path including specified commands
<parameters> - parameters link to command
HTTP request for our NCAP looks like this:
http://192.168.0.33/1451/TransducerAccess/ReadData
192.168.0.33 – IP address assigned to Mbed HTTP server
/1451/TransducerAccess/ReadData – path which indicate to appropriate method of 1451.0 API

When this request is received on the NCAP, measured data are read from TIMs sensors and then it was send back to the requester in structured XML format.

```
<sensor>
        <air_temperature>23.9</air_temperature>
        <relative_humidity>36.0</relative_humidity>
        <radiant_temperature>23.9</radiant_temperature>
        <air_speed>0.1</air_speed>
</sensor>
<sensor>
        <air_temperature>-2.4</air_temperature>
        <relative_humidity>78.0</relative_humidity>
        <radiant_temperature>-2.4</radiant_temperature>
        <air_speed>0.1</air_speed>
</sensor>
```

In presented XML structure, first two parameters, air temperature and relative humidity, are measured value from sensors. Last two values are not obtained by measuring; radiant temperature that is equalized to air temperature and air speed that is equal to constant value of 0.1 m/s.

The role of sender requester in our case has Web service (WS). WS gets the required data from *mBed* over the network, computes necessary values and returns them to end users. WS provide interoperating between software applications implemented in any language, running on a variety of platforms and frameworks. WS and its consumer exchange data using XML over HTTP or SOAP (Simple Object Access Protocol). Besides SOAP, WS uses Universal Description Discovery, and Integration (UDDI) and Web Services Description Language (WSDL) to publish their functionalities to clients.

In this paper, created Web service is used to compute and provide information about thermal comfort according to user selected parameters and measure values. User parameters are clothes (CLO) and metabolic activity (MET) passed to a web service from a client application, and four measured parameters (*air_temperature*, *radiant_temperature*, *air_speed*, *relative_humidity*) parsed from XML data obtained by NCAP. The code shown below represents part of the WSDL file that describes interaction with WS method getComfort and defines return values.

```
<element name="getComfort">
    <complexType>
        <sequence>
            <element name="CLO" type="xsd:double" />
            <element name="MET" type="xsd:double" />
        </sequence>
    </complexType>
</element>
<complexType name="ThComfort">
    <sequence>
            <element name="PMV" type="xsd:double" />
            <element name="PPD"  type="xsd:double" />
    </sequence>
</complexType>
```

J2EE web application, directly designed for client, was developed by Java Server Faces (JSF) framework. Development environment used in its realization was NetBeans 7.1 with Glassfish 3.1.

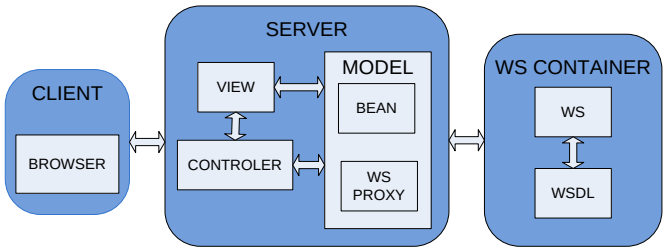


Figure 3. Web service position in client application

According to the structure of the client application (Figure 3.), WS can be easily inserted into logical model of the application. This means that interaction of application with WS was transparent for designer developers and they do not need to deal with that functionality.

Invoking WS functionalities in a web project was done by "*creating new web service client"* option in NetBeans, by pointing to WSDL URL. After this, source classes described in WSDL were created inside a web project. JSF page as a presentation page provides variable value visibility. This means that every presented value to a client in a JSF page is mapped in a managed Bean that creates its value. As a demonstration of WS functionalities, client application presents temperature comfort scale level as position of slider, ranging from COLD to HOT. Meteorological data are provided from two sensors, located indoor and outdoor, with values contained in NCAP XML response. Client application functionalities are shown in Figure 4.

Java framework was chosen for the web service since it offers build-in support by AXIS framework, which enables faster development of web services and web service clients.

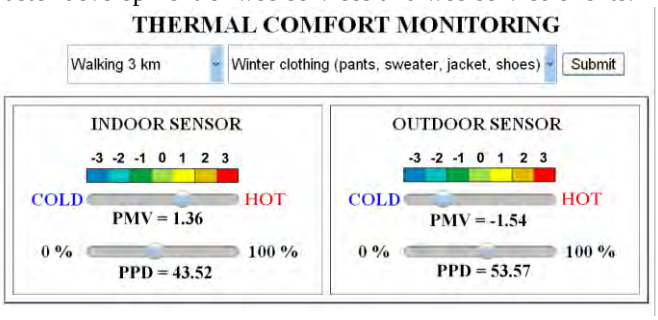


Figure 4. Client application

V. Conclusion

This paper presents system for monitoring thermal comfort in working and living space, based on the concept of IEEE 1451 smart transducers accessible via web service interface. Current system is based on Ethernet connectivity between smart transducer network and web service, which limits its usage to locations with existing Internet infrastructure. System can be upgraded with Wireless LAN or GPRS connection to NCAP, in order to be used in remote locations. Also, smart transducer network can be scalable in number of transducers which can be connected to NCAP either by wired of wireless connection.

Acknowledgement

Work presented in this paper was funded by grant TR32043 for the period 2011-2014, by the Ministry of Education and Science of the Republic of Serbia.

References

[1] Application Note: "Conditions of Thermal Comfort - Influence of Humidity and Temperature on Personal Well-Being", Sensirion – The Sensor Company, February 2010

[2] "A Comfort System Base on ZigBee Wireless Monitoring and Control in Smart Home", Neng-Sheng Pai Yao-Hong Ou, The 5th Imtelligent Living Technology Conference, Taiwan 2010.

[3] "Ergonomics of the thermal environment — Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria", International standard – ISO 7730**,** Third edition, 2005-11-15

[4] IEEE Standard for a Smart Transducer Interface for Sensors and Actuators – Common Functions, Communication Protocols, and Transducer Electronic Data Sheet (TEDS) Formats, IEEE Std. 1451.0-2007.

[5] "Datasheet SHT1x (SHT10, SHT11, SHT15) - Humidity and Temperature Sensor IC", Sensirion – The Sensor Company, December 2011

[6] mbed NXP LPC1768. http://mbed.org/handbook/mbed-NXP-LPC1768. Poslednja poseta: 04.02.2012

Some Integral Characteristics of MRC Receiver in Nakagami-*m* fading Environment

Hana Stefanovic[1], Dejan Milic[2], Dimitrije Stefanovic[3], Srdjan Milosavljevic[4]

Abstract – **In this paper Maximal Ratio combining (MRC) technique over Nakagami-*m* fading channel is described and some statistical characteristics of received signal envelope are analyzed. For analytical and numerical evaluation of system performance, the probability density functions (pdf) of received signal envelope after MRC are analyzed like particular solutions of corresponding differential equation, while the existence of singular solution is considered and analyzed under different conditions.**

Keywords – **Nakagmi-*m* fading channel, MRC receiver, probability density function (pdf), singular solution.**

I. INTRODUCTION

Many modern communication systems, like wireless cellular systems, operate in environments that are interference and bandwidth limited, where propagation characteristics are more complicated and multipath-induced fading and shadowing are a common problem [1]. A great number of channel models have been proposed to describe the statistics of the amplitude and phase of multipath faded signals [2]. As the result of multipath reception, the mobile antenna receives a large number of reflected and scattered waves. The rapid fluctuations of the instantaneous received signal power due to multipath effects are usually described with Rayleigh, Rician, Nakagami-*m*, Nakagami-*q* or Weibull model [1]. This paper discusses the case of Nakagami-*m* distribution, which models radio transmission in urban areas [3] where the random fluctuations of the instantaneous received signal power are very frequent and fast. It is shown that Nakagami-*m* distribution can model different propagation conditions, providing more flexibility and higher accuracy in matching some experimental data in comparison with the commonly adopted distributions [4]. Nakagami-*m* distribution is suitable for describing statistics of mobile radio transmission in complex medium such as the urban environment [5].

In the case of all distributions considered in the statistical theory of telecommunications, the pdfs of the received signal envelope are functions of several variables [1-3]. In the process of determination and analysis of integral characteristics, one of the variables is treated as a parameter, while the others are set to certain constant values of interest in practice. In this way, one obtains the received signal pdf curves family. The analysis of the position of the maximums for curves family can be performed analytically, using the first derivative of function, and also numerically. In such a case the same procedure is repeated for the situation where the second variable is treated as a parameter, and the others are set to constant values. This process gives a new family of curves, while the maximum position is determined by a new envelope. In papers [6-10] the position of maximums and the analytical expression for the integral characteristic of Nakagami-*m* distribution of received signal envelope are determined. Some integral characteristics of Rayleigh, Weibull and Rician distribution are presented in [11-13].

In order to combat multipath fading and shadowing effects, the complex receiver structures, using complicated synchronization schemes, demodulators, symbol estimators and diversity and MIMO techniques, are often applied [1-3]. An efficient method for mitigating fading effects by using multiple receiver antennas is called space diversity [3]. The main goal of space diversity techniques is to improve transmission reliability without increasing transmission power and bandwidth while increasing channel capacity. There are several types of space combining techniques that can be generally performed depending on the amount of channel state information (CSI) available at the receiver. The most frequently used schemes are selection combining (SC), equal-gain combining (EGC) and maximal ratio combining (MRC). EGC involves co-phasing of the useful signal in all branches, and summing of the received signals from all antennas. By co-phasing, all the random phase fluctuations of the signal that emerged during transmission are eliminated. For this process it is necessary to estimate the phase of the received signal, so this technique requires all of the amount of the CSI on received signal, and separate receiver chain for each branch of the diversity system, increasing the complexity of system. MRC output is a weighted sum of co-phased signals from all branches, also requiring all of the amount of CSI. Unlike pervious, SC technique processes only one of the diversity branches. Generally, SC selects the branch with the highest signal-to-noise ratio (SNR), that is the branch with the strongest signal [3], assuming that noise power is equally distributed over branches. In this paper we analyze MRC technique, which is chosen because of its benefits when compared to other techniques.

The remainder of this paper is as follows. After this Introduction, some characteristics of Nakagami-*m* fading

[1]Hana Stefanovic is with the College of Electrical Engineering and Computer Science Applied Studies, Vojvode Stepe 283, 11000 Belgrade, Serbia, E-mail: stefanovic.hana@yahoo.com

[2]Dejan Milic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: dejan.milic@elfak.ni.ac.rs

[3]Dimitrije Stefanovic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vule@elfak.ni.ac.rs

[4]Srdjan Milosavljevic is with the Faculty of Economics, Kolasinska 156, 38220 Kosovska Mitrovica, Serbia, E-mail: srdjan.milosavljevic@pr.ac.rs

channel model and MRC receiver operating over it are presented in Section II. The procedure of determining the integral properties in the case of MRC reception is described in Section III. For a fixed value of the fading depth parameter and average signal power, the pdf of signal envelope after MRC, depending on the received signal level, is analyzed for different number of diversity branches. Then the average signal power is treated as parameter, while the values of the received signal level and fading depth parameter are set to the constant values. In that case, the pdf of signal envelope after MRC is analyzed also for different number of diversity branches. In such a way, two series of families of curves are obtained and for each of them equation of the envelope of curves maximums is considered. In both cases these envelopes are straight lines, in the logarithmic scale, whose direction coefficients and values on ordinate-axis are determined analytically and numerically. Also, in both cases, the differential equations describing the complete dynamics of signal transmission process are determined, whereby the envelopes of the pdf curves families represent their singular solutions. Finally, some conclusions are given in Section IV.

II. General properties of Nakagami distribution

A. Nakagami-m Fading Channel Model

Since the Nakagami-*m* random process is defined as envelope of the sum of 2*m* independent Gauss random processes, the Nakagami-*m* distribution is described by pdf [3-4]:

$$p_z(z,\Omega)=\frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m z^{2m-1}\cdot\exp\left(-\frac{m}{\Omega}z^2\right), z>0, m\geq\frac{1}{2} \quad (1)$$

where z is the received signal level, $\Gamma(\cdot)$ is Gamma function, m is the fading depth parameter (fading figure), defined as:

$$m=\frac{\mathrm{E}^2[z]}{\mathrm{Var}[z^2]} \quad (2)$$

while Ω is the average signal power:

$$\Omega=\mathrm{E}[z^2] \quad (3)$$

The analytical expression for the envelope of pdf curves family maximums is determined [6-9] depending on the relevant parameters, like fading depth parameter and average signal power. The obtained results show that the position of the maximums of these functions is uniquely determined by envelope's equation, regardless of the values of the other parameters.

B. MRC Receiver over Nakagami-m Fading Channel

MRC is the optimal combining scheme, regardless of fading statistics, but it requires knowledge of all channel fading parameters [3]. In the case of MRC diversity technique, received signal envelope is described by [14]:

$$p_{MRC}(z,\Omega)=\left(\frac{m}{\Omega}\right)^{mM}\frac{z^{mM-1}}{\Gamma(mM)}\cdot\exp\left(-\frac{m}{\Omega}z\right) \quad (4)$$

where M presents the number of diversity branches.

III. Analytical and numerical results

Graphic presentation of dependence of received signal pdfs after MRC, versus received signal level, for fixed values of the fading depth parameter and the average signal power, in logarithmic scale, for different number of diversity branches, is shown in Fig 1.

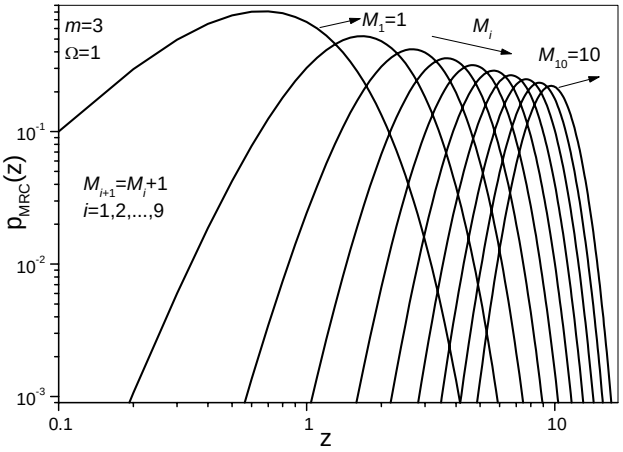


Fig. 1. The received signal pdf versus signal level, in logarithmic scale, for fixed values of *m* and Ω, with number of diversity branches taking values from 1 to 10.

Analysis of this dependence shows that the increase of number of diversity branches results in smaller values of the pdf maximums, reached for higher values of received signal level, as shown in Fig 1. It can also be concluded that all the maximums lie on a straight line, and that the envelope of maximums is also a straight line, in logarithmic scale. In order to determine the values of received signal level when the maximums are reached, as well as the values of maximums and direction coefficient of envelope, the first derivative of (4) relative to received signal level is determined. Equalization the first derivative with zero obtains:

$$z_{p\max}=\Omega(-1+mM)/m \quad (5)$$

Substituting (5) into (4) yields:

$$p_{MRC}^{z\max}=\frac{1}{\Gamma(mM)}\left(\frac{m}{\Omega}\right)^{mM}\left(\frac{\Omega(mM-1)}{m}\right)^{mM-1}\cdot\exp(1-mM) \quad (6)$$

From (6) we get:

$$\log\left(\max p_{MRC}(\Omega)\right)=k_1\log\Omega+n_1 \quad (7)$$

where the direction coefficient is:

$$k_1=-1 \quad (8)$$

and the value on the ordinate axis is:

$$n_1=\log m-\log\Gamma(mM)+(mM-1)\log(mM-1)+1-mM \quad (9)$$

The direction coefficient of envelope has the value -1, regardless of the value of the fading depth parameter and number of diversity branches.. The envelope determines a certain singular solution of differential equation which can describe the dynamics of this process, while the received signal pdf is its particular solution:

$$p'_{MRC}(z)+p_{MRC}(z)\left(m/\Omega-(mM-1)/z\right)=0 \quad (10)$$

The influence of fading parameter m is presented in Fig 2, showing the translation of the pdf curves family maximums envelope. For higher values of m, the value on the ordinate axis n_1 is greater, as shown in Fig 2, while the direction coefficient k_1 has the same value.

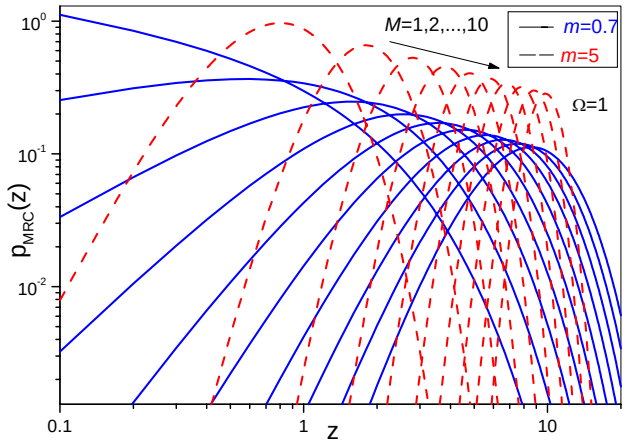


Fig. 2. The received signal pdf versus signal level, in logarithmic scale, for fixed value of Ω, and different values of m, with number of diversity branches taking values from 1 to 10.

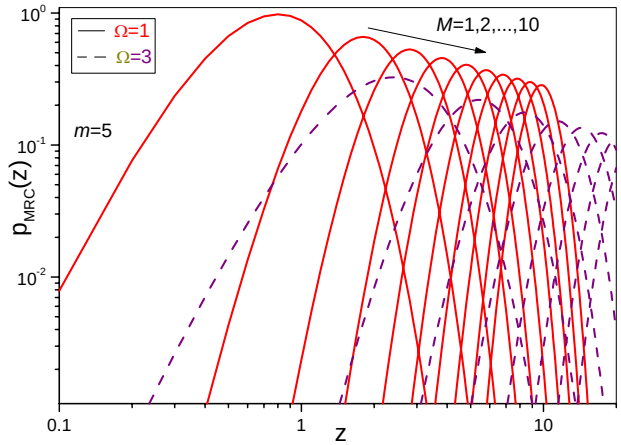


Fig. 3. The received signal pdf versus signal level, in logarithmic scale, for fixed value of m, and different values of Ω, with number of diversity branches taking values from 1 to 10.

The influence of average signal power is presented in Fig 3, also showing the translation of the pdf curves family maximums envelope. For higher values of Ω, the value on the ordinate axis n_1 is smaller, as shown in Fig 3, while the direction coefficient k_1 has the same value.

Graphic presentation of dependence of Nakagami-m pdfs after MRC, versus average signal power, for a fixed value of the fading depth parameter and received signal level, in logarithmic scale, for different number of diversity branches is shown in Fig 4.

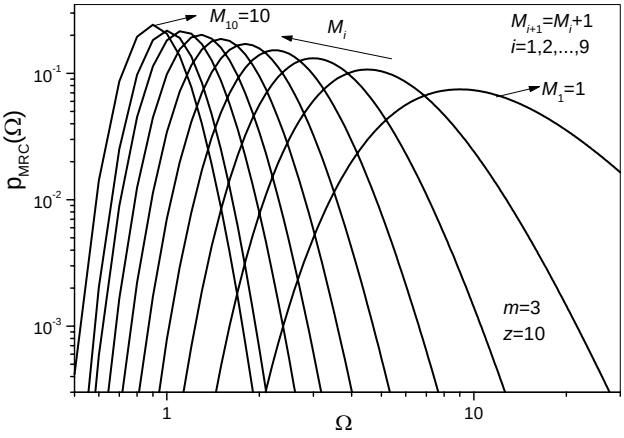


Fig. 4. The received signal pdf versus average signal power, in logarithmic scale, for fixed value of m and z, with number of diversity branches taking values from 1 to 10.

The analysis of this dependence shows that the increase of number of diversity branches results in higher values of the pdf maximums, reached for smaller values of average signal power, as shown in Fig 4. It can also be concluded that all the maximums lie on a straight line, and that the envelope of maximums is also a straight line, in logarithmic scale. In order to determine the values of average signal level when the maximums are reached, as well as the values of maximums and direction coefficient of envelope, the first derivative of (4) relative to average signal power is determined. Equalization the first derivative with zero obtains:

$$\Omega_{p\max}=z/M \quad (11)$$

Substituting (11) into (4) yields:

$$p_{MRC}^{\Omega\max}=\frac{(mM)^{mM}}{\Gamma(mM)}\cdot\frac{1}{z}\cdot\exp(-mM) \quad (12)$$

From (12) we get:

$$\log\left(\max p_{MRC}(z)\right)=k_2\log z+n_2 \quad (13)$$

where the direction coefficient is:

$$k_2=-1 \quad (14)$$

and the value on the ordinate axis is:

$$n_2 = mM\log(mM) - \log\Gamma(mM) - mM \qquad (15)$$

The direction coefficient of envelope has the value -1, regardless of the value of the fading depth parameter and number of diversity branches. The envelope determines a certain singular solution of differential equation which can describe the dynamics of this process, while the received signal pdf is its particular solution:

$$p'_{MRC}(\Omega) + p_{MRC}(\Omega)(M - z/\Omega)\cdot m/\Omega = 0 \qquad (16)$$

The influence of fading depth parameter m is presented in Fig 5, showing the translation the translation of the pdf curves family maximums envelope. For higher values of m, the value on the ordinate axis n_2 is greater, as shown in Fig 5, while the direction coefficient k_2 has the same value.

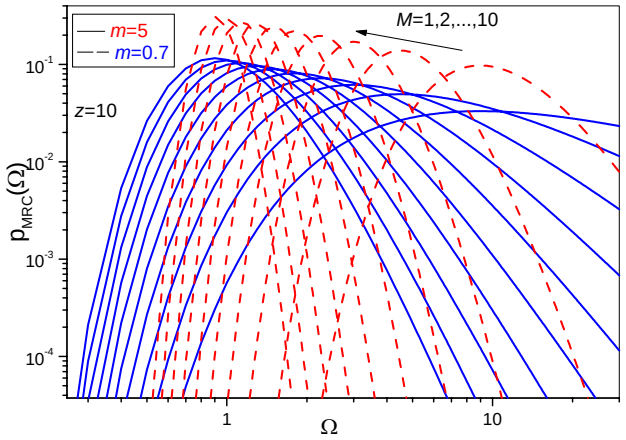


Fig. 5. The received signal pdf versus average signal power, in logarithmic scale, for fixed value of z, and different values of m, with number of diversity branches taking values from 1 to 10.

The influence of received signal level is presented in Fig 6, also showing the translation of the pdf curves family maximums envelope. For higher values of z, the value on the ordinate axis n_2 is smaller, as shown in Fig 6, while the direction coefficient k_2 has the same value.

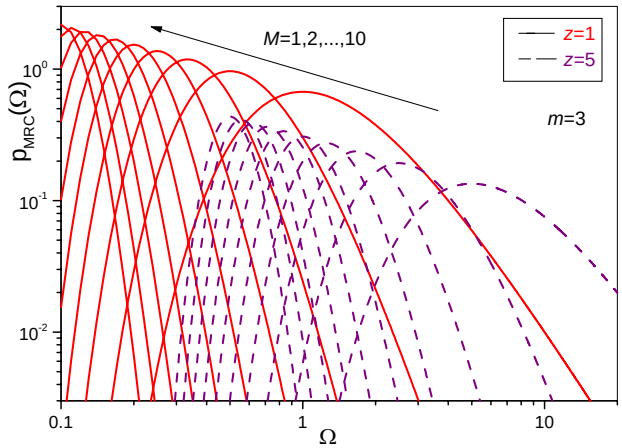


Fig.6. The received signal pdf versus average signal power, in logarithmic scale, for fixed value of m, and different values of z, with number of diversity branches taking values from 1 to 10.

IV. Conclusion

This paper presents some properties of the envelope of the pdf curves family in Nakagami-m environment when MRC receiver is applied. The obtained results show that the position of the maximums of these pdfs is uniquely determined by the maximums envelope's equation, which presents the singular solution of corresponding differential equation describing the complete dynamics of the signal transmission process.

In such a way, the boundary conditions for radio transmission, with given propagation conditions, could be defined, while some system performance measures could be evaluated, since they are all related to the received signal pdf.

References

[1] J. Proakis, *Digital Communications*, 3rd Ed., McGraw-Hill, 1999.

[2] W. Y. C. Lee, *Mobile Cellular Communications*, McGraw-Hill Book Co., New York, 1989.

[3] M. K. Simon, M.S. Alouini, *Digital Communication over Fading Channels*, New York: Wiley, 2000.

[4] M. Nakagami, *The m-distribution, a General Formula of Intensity Distribution of Rapid Fading in Statistical Methods in Radio Wave Propagation*, W. G. Hoffman, Ed.Oxford, England: Pergamon, 1960.

[5] G. K. Karagiannidis, "Moments-Based Approach to Performance Analysis of Equal Gain Diversity in Nakagami-*m* Fading", IEEE Trans. on Communications, vol. 52, no. 5, pp. 685-690, 2004.

[6] H. Stefanovic, A. Savic, "Integral characteristics of the Nakagami-*m* distribution of signal envelope", IPSI BgD Transactions on Internet Research, vol. 8, no. 1, 2012.

[7] H. Popovic, A. Mitic, I. Stefanovic, D. Stefanovic, "Some Integral Properties of Nakagami-*m* distribution", ICEST 2007, Conference Proceedings, pp. 299-302, Ohrid, Macedonia, 2007.

[8] H. Popovic, D. Stefanovic, A. Mitic, I. Stefanovic, D. M. Stefanovic, "Some Statistical Characteristics of Nakagami-*m* Distribution", TELSIKS 2007, Conference Proceedings, vol. 2, pp. 509-512, Nis, Serbia, 2007.

[9] H. Stefanovic, A. Savic, "Some general characteristics of Nakagami-*m* distribution", ISCIM 2011, Conference Proceedings, pp. 721-731, Tirana-Durres, Albania, 2011.

[10] H. Stefanovic, D. Stefanovic, Z. Popovic, V. Stefanovic, "Analiza integralnih karakteristika Nakagami-*m* raspodele", INFOTEH 2009, Conference Proceedings, vol. 8, ref. B-I-6, pp. 104-107, Jahorina, Bosnia and Herzegovina, 2009.

[11] H. Popovic, Z. Popovic, D. Blagojevic, V. Stefanovic D. Stefanovic, "Analiza integralnih karakteristika Rejlijeve raspodele", TELFOR 2007, Conference Proceedings, pp. 365-368, Belgrade, Serbia, 2007.

[12] H. Stefanovic, Z. Popovic, D. Stefanovic, D. Blagojevic, "Analiza integralnih karakteristika Vejbulove raspodele", TELFOR 2009, Conference Proceedings, pp. 500-503, Belgrade, Serbia, 2009.

[13] A. Mitic, D. Blagojevic, D. Stefanovic, S. Veljkovic, "The Analysis of Rician PDF Integral Properties from the DSRC System Viewpoint", ICEST 2007, Conference Proceedings, pp. 295-297, Ohrid, Macedonia, 2007.

[14] C. D. Iskander, P. T. Mathiopoulos, "Analytical Level Crossing Rate and Average Fade Duration for Diversity Techniques in Nakagami Fading Channels", IEEE Trans. Commun., vol. 50, no. 8, pp. 1301-1309, 2002.

Adaptive Filtering Algorithms Suitable for Real-Time Systems

Maria Nenova[1]

Abstract – **In this paper the main characteristics of adaptive systems are introduced. The vital role in the development of adaptive algorithms for the telecommunications is explained. Also, adaptive systems proved to be extremely effective in achieving high efficiency, high quality and high reliability of ubiquitous telecommunication services.**

Keywords – **adaptive filtering, adaptive filtering algorithms, IIR and FIR adaptive filters.**

I. INTRODUCTION

Adaptive digital filters can be realized on the basis of different structures. The choice of structure is the main factor influencing the computational complexity (number of arithmetic operations at each iteration), and hence the number of iterations to achieve the desired efficiency. Adaptive digital filters can be divided mainly into two main classes according to the received pulse shape characteristics: finite impulse response characteristic (FIR) and infinite impulse filter characteristics (IIR).

The FIR and IIR filters can be realized with implementation of an adaptive algorithm. The solution to successful adaptive signal processing is understanding the fundamental properties of the adaptive algorithms. The algorithms instead of structure (recursive or not) are also divided into two main classes.

The main characteristics for assessment the performance of adaptive systems are: stability, speed of convergence of the algorithm, missadjustement errors, robustness to both additive noise and signal conditioning (spectral colouration), least mean square error, numerical (computational) complexity, robustness, the order of the filter transfer function and the round-off error analysis of adaptive algorithms.

However, some of these properties are often in direct conflict with each other, since consistent fast converging algorithms tend to be in general more complex and numerically sensitive. Also, the performance of any algorithm with respect to any of these criteria is entirely dependent on the choice of the adaptation update function, that is the cost function used in the minimization process. A compromise must be than reached among these conflicting factors to come up with the appropriate algorithm for the concerned application.

[1]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. MEASUREMENT PARAMETERS

1. Cost functions

Before proceeding to discuss any adaptive algorithm, it is necessary to discuss the performance measure (cost function) which is used in adaptive filtering. The adaptive filter has the general form shown in Fig. 1 where the FIR filter of order N is considered here. The filter output $y(n)$ is given by:

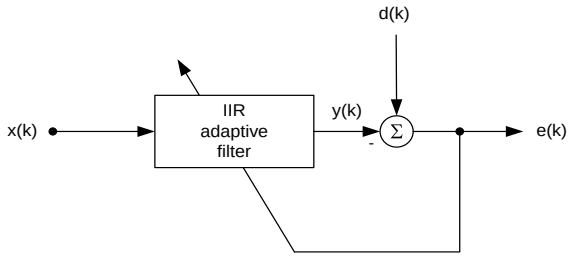


Fig. 1. Figure example

$$y(n) = \sum_{i=0}^{N-1} w_i(n)x(n-i) = \quad (1)$$

$$= w^T(n)x(n) \quad (2)$$

and $$x(n) = [x(n), x(n-1), x(n-2), \ldots, x(n-(N-1))], \quad (3)$$

where $$w^T(n) = [w_0(n), w_1(n), w_2(n) \ldots w_{N-1}(n)] \quad (4)$$

where T denotes transpose. In general, adaptive techniques have been classified under two main categories. In one category, the cost function to be optimised in a running sum of squared errors is given by:

$$J(n) = \sum_{j=0}^{n} e^2(j) \quad (5)$$

where the error $e(n)$ is defined to be the difference value between the desired response $d(n)$ and the output of the adaptive filter $y(n)$, that is, $e(n) = d(n) - y(n)$ The approach, defined by (5), is based on the method of least squares [3-5], which contains the whole class of recursive least squares (RLS) algorithms [6], [7].

In the other category, the cost function to be optimised is a statistical measure of the squared error, known as the mean squared-error (MSE) [7]. This cost function is given by

$$J(n) = E[e^2(n)] \quad (6)$$

where $E[]$ denotes the statistical expectation. This category contains the whole class of gradient algorithms, which includes the least mean-squared (LMS) algorithm [1], [6], [7].

All the functions presented in this section and others not mentioned in this work should be positive and monotonically increasing [35] for their corresponding algorithms to perform correctly.

2. Convergence Rate

The speed of convergence determines the rate at which the filter converges to the optimal solution. The main objective in the design of adaptive systems is to achieve fast convergence. The speed of convergence depends on all the other characteristics of the filter.

If some of the parameters are changed to receive good convergence of the adaptive digital filter, then this will lead to increase or decrease of the other characteristics. Very often when the speed of convergence increases, the stability characteristics will get worst, this make the system to diverge to a solution instead of converge to the optimal solution. This proves that the speed of convergence to the optimal solution can be considered and evaluated only in the context of assessment of the other key features.

3. Minimum Mean Square Error

The least mean square error (MSE) is a parameter indicating how exact the system adapts to the particular solution. If the MSE has a very small value, this means that the adaptive system exactly converges to the optimal solution of the system. If the size of the MSE is very large, it means that the

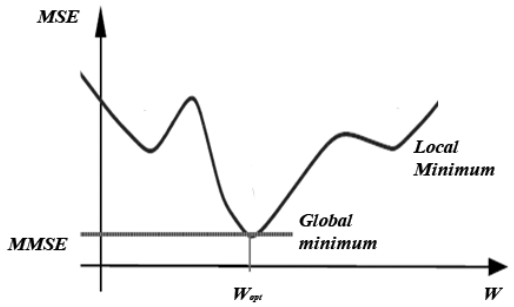


Fig. 2. Figure example

adaptive filter can not properly model the system or initial conditions that are set are wrong starting point is not correct and the filter converges, but not to the right/optimal solution. There are many parameters by which can be determinated the minimum of the MSE. Some of those parameters are the noise due to quantization effects, order of the adaptive system, measurement noise and the gradient error due to the final step size of adaptation.

If an adaptive FIR filter has two weights then the performance surface has the form of a paraboloid in 3 dimensions. If the filter has more than three weights then it cannot be drawn the performance surface in three dimensions. In mathematical point of view there is only one minimum point which occurs when the gradient vector becomes equal to zero. When the performance surface is quadratic with more than three dimensions is called a hyperparaboloid.

If a FIR filter is used, then the MSE is determined, leading to a point situated in the lowest part of a hyperboloid. The mathematical representation of this is the process of finding the zero of the gradient vector.

If the implementation is based on an IIR filter structure, then the MSE will have local minimum as well the desired global one. The graphical representation of this is:

4. Computational Complexity of adaptive algorithms

The computational complexity is very important parameter in real time applications of adaptive digital filters. If the applications are in real-time systems there are limitations which are introduced by the hardware, and they can affect the behavior of the whole system. If is used a very complicated algorithm it will require more complex hardware than it gets when a simpler ones are used.

The efficiency of the algorithm is in close relation with the computational complexity. The number of additions and multiplications per iteration are the limit of the adaptive system, because they take time on the processor to process the signals. The tendency is to develop more and more complex algorithms. One of the most interesting areas for researchers is the development of algorithms with lower computational requirements, due to the limitations of the hardware realizations.

Important feature for an algorithm is the time necessary for processing. This time is based on the number of operations in a single operation. Usually adaptive algorithms are iterative.

5. Stability

The stability is the next important feature which is important to be investigated during the process of design of adaptive filters. This is may be the most important characteristic. Because of their nature adaptive systems have very few completely asymptotically stable systems can be realized. In most cases, the systems used are marginally stable, which is predetermined by the initial conditions, the system transfer function and step at the entrance.

6. Robustness

The robustness of a system is directly related to the stability of a system. Robustness can be defined as the ability of the system to tolerate changes. The redundant, concurrent system models allow for a quick context switching on occurrences of abrupt changes and also for concurrent simulation and testing to continuously adapt to the environment or to the requirement.

Robustness is a parameter by which is measured how the system will work when are introduced the effects of the input noise, the noise due to the quantization and the insensitivity to external errors. Those analyses the behavior of a system against internal errors due to the effects listed above.

7. *Filter order*

The order of the adaptive filter system is inherently related to many of the other parameters involved in the assessment of the system. The order determines how exact a system can be modeled by an adaptive filter. It also affects the speed of convergence by increasing or decreasing the time necessary for calculations, the stability of the system at a fixed step size of adaptation and the minimum of the MSE. If the order of the filter increase then will increase the number of calculations, thus reducing the maximum speed of convergence. In order to achieve stability because of increasing the order of the filter can be added poles and zeros, which is less than it already has. In such cases, the maximum step size or the maximum rate of convergence will have to be reduced to ensure stability. Finally, if the system is under specified, meaning there are not enough pole and/or zeroes to model the system, the least mean square error will converge to a constant different from zero. Usually when the system is over specified, meaning that it has to many poles and/or zeros it will be possible to converge to zero, but the increasing number of calculations will affect the maximum of possible speed of convergence.

This division according to different types of algorithms is depicted on the next graph.

III. ADAPTIVE ALGORITHMS

Types of adaptive algorithms

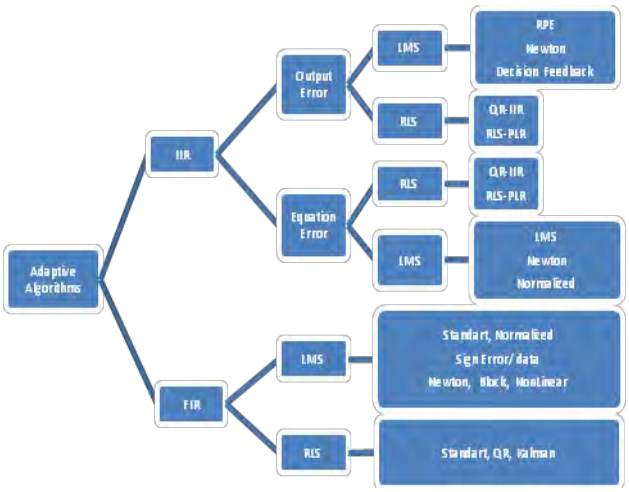


Fig. 3. Adaptive algorithms tree

The adaptation of the filter parameters is based on minimizing the mean squared error between the filter output and a desired signal. The most common adaptation algorithms are the Recursive Least Square (RLS) and the Least Mean Square (LMS). The RLS algorithm has higher convergence speed compared to the LMS algorithm. If the main characteristic is the computation complexity, then the LMS algorithm is much faster than the RLS. Due to the computational simplicity, the LMS algorithm is most commonly used in the design and implementation of integrated adaptive filters. The LMS digital algorithm is based on the gradient search according to the equation (1).

3. 1. Finite Impulse Response (FIR) Algorithms

The adaptive algorithms can be divided in two main types according to the mathematical formulation used. As it was previously mentioned – RLS and LMS and the second point of division – IIR or FIR. The implementation depends on the specific requirements in each realization. One of the most widely used is:

3.1.1 Least Mean Squares Gradient Approximation Method

If is given an adaptive filter with an input signal $x(n)$, an impulse response $w(n)$ and an output signal $y(n)$ can be derived a mathematical relation for the transfer function of the system:

$$w(n+1) = w(n) - \mu\Delta_{E^2}(n) \tag{7}$$

$$y(n) = w^T(n)x(n) \tag{8}$$

and $$x(n) = [x(n), x(n-1), x(n-2), \dots, x(n-(N-1))], \tag{9}$$

where $$w^T(n) = [w_0(n), w_1(n), w_2(n) \dots w_{N-1}(n)] \tag{10}$$

are the time domain coefficients for an N-th order FIR filter.

In the above equation $w(n+1)$ represents the new coefficient values for the next time interval, μ is called scaling factor, and $\Delta_{E^2}(n)$ is the ideal cost function with respect to the tap weight $w(n)$. From the above formula can be derived the estimate for the ideal cost function

$$w(n+1) = w(n) - \mu e(n)x(n) \tag{11}$$

where $e(n) = d(n) - y(n)$ and $y(n) = x^t(n)w(n)$.

In the above equation the coefficient μ is very often multiplied by 2, but here we will assume it is in the μ factor.

The Least Mean Squares Gradient method, usually presented as the Method of Steepest Descent, an investigation based on the current filter coefficients is made, and the gradient vector, the derivative of the MSE with respect to the filter coefficients, is calculated from the investigated. Secondly is made tap-weight vector estimation by making a change in the present guess in a direction opposite to the gradient vector. This process is repeated until the derivative of the MSE is zero.

3.1.2 Quasi-Newton Adaptive Algorithms

The quasi-Newton adaptive algorithms are based on the implementation of second order statistic in order to reduce the speed of convergence of the adaptive filter, according to the Gauss-Newton method. The fames one quasi-Newton algorithm is the RLS algorithm. It is important to be notet that

even the speed of convergence is increased the RLS requires a great amount of processor power, which will lead to difficulties in their implementation in real-time systems.

In the family of quasi-Newton algorithms are some having good convergence properties and are alternative to process information signals in real-time. In the paper [7] and [9] it is well described.

3.1.3 Adaptive Lattice Algorithms

The main reason for the use of lattice structures is to reduce quantization noise introduced by the filter coefficients in systems with limited word length. The purpose of developed lattice adaptive algorithms is to reduce the effects of quantization noise and thus to try to reduce the length of the register maintaining good behavior. In [7] and [9] this is well described.

3.2. Infinite Impulse Response (IIR) Adaptive Filters

The most important advantage of IIR filters is that they are the basis to receive an equivalent amplitude frequency response of a FIR filter, but with a lower number of coefficients. This theoretically decreases also the number of adders, multipliers and other mathematical operations to perform filtering. This is the main reason for implementing them. The lower number of coefficients leads to less number of undesired sources of noise (for example due to the limited length of the digital registers-finite word length). The recursive filters, however, lead to many problems in their use, due to their instability issues.

The main problem with the use of adaptive recursive filters is the possible instability of the poles position of the transfer function. In some cases they could get out of the unit circle, during the process of training the system which means that the system will become unstable. Even if the system is stable at the beginning and in the end, there still is a possibility the system to be destabilized during the process of convergence. One good solution to this is to introduce restrictions on the position of the poles (in order to limit them within the unit circle) but this method requires a small step size, which will significantly reduce the speed of convergence.

Due to the interplay between the movement of the poles and zeros, the convergence of IIR systems tends to be slow [3]. The result is that even though IIR filters have fewer numbers of coefficients, therefore fewer calculations per iteration, the number of iterations may increase and this will cause a change and loss of time in processing time to reach the convergence. This, however, is not a problem when implementing all pole filters.

In IIR system the MSE surface can have a local minimum, which can lead to convergence of this system to a local minimum and not to a global one. It must be considered also the initial conditions for adaptive IIR filters.

The IIR filters are more susceptible to quantization errors of the coefficients of the FIR, which is due to the presence of a feedback.

There have been a number of studies done on the use of IIR adaptive filters, but due to the problems stated above, they are still not widely used in industry today.

IV. Conclusion

In conclusion in this contribution was made an outline of the main adaptive algorithms and characteristics important in investigations of an adaptive system. The choice of the mathematical algorithm for IIR and FIR filters depends on the concrete realization and performance desired. In some cases adaptive RLS algorithms are preferred, in other LMS as the most widely used for both types of filters because of its very good properties: fast convergence and good performance stability.

Investigations on different structure realizations are also important in order to avoid finite word length effects, possible instability when IIR filters are implemented. A crucial point for researchers is to develop structures with canonical number of multipliers and with low sensitivity to all those undesired effects.

The role of adaptive systems is wide spread covering almost all aspects of telecommunication engineering, but perhaps most notable in the context [3] of ensuring high-quality signal transmission over unknown and time varying channels.

References

[1] M. Bellanger, *Adaptive Digital Filters* (Second edition). Marcel Dekker, ISBN 0-8247- 0563-7, New York, 2001

[2] A. Shoval, D. Johns, W. Snelgrove, "Comparison of DC Offset Effects in Four LMS Adaptive Algorithms, IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing; Volume 42, No. 3, (March 1995), pp. 176- 185.

[3] K. Murano, "Adaptive Signal Processing Applied in Telecommunications," IFAC Adaptive Systems in Control and Signal Processing, pp. 431-441, 1992.

[4] H. W. Sorenson, "Least-Squares Estimation from Gauss to Kalman," *IEEE Spectrum,*vol. 7, pp. 63-68, July 1970. [2]

[5] C. F. N. Cowan and P. M. Grant, *Adaptive Filters*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[6] G. Carayannis, D. G. Manolakis, and N. Kalouptsidis, "A Fast Sequential Algorithm for Least-Squares Filtering and Prediction," *IEEE Trans. Acoust., Speech, Signal Processing, vol.* ASSP-31, pp. 1394-1402, Dec. 1983.

[7] J. M. Cioffi and T. Kailath, "Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering," *IEEE Trans. Acoust., Speech, Signal Processing, vol.* ASSP- 32, pp. 304-337, 1984.

[8] Haykin, Simon, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey, 2003.

[9] Jenkins, W. Kenneth, Hull, Andrew W.,Strait, Jeffrey C., Schnaufer, Bernard A., Li, Xiaohui, *Advanced Concepts in Adaptive Signal Processing*, Kluwer Academic Publishers, Boston, 1996.

Presentation of a model to study facsimile coded signals from a fourth group

Todorka Georgieva[1]

Abstract - **The paper presents investigation of the advantages and disadvantages of the coding methods and the compression level, error sensitivity and selecting an eventual compromise set of coding methods have made MRCII (Modified READ Code II)**

Keywords – **coding, Modified READ Code, images**

I. INTRODUCTION

Coding of facsimile signals is used for shortening the amount of transmitted messages (data). Using these codes ensure not only efficiency but also the compatibility of different groups of terminals [1]. Essential requirement for the development of various codes is that of economy, ie code combinations equivalent to certain messages should have shortest length possible. This length is defined primarily from basis of the selected code. If source generates various messages, the length of a code combination is :

$$N = \{ \mathrm{Log}_m N \} \quad (1)$$

where "m" is basis of the selected code, and "N" is length of the codeword.

Facsimile devices of fourth group use two-dimensional coding line by line. Before transmission errors are corrected by control procedure at a higher level. Scanned lines in Group 3 are coded two-dimensional and every K lines adds a one-dimensional coded line to prevent the spread of errors.

Signal for end of line (EOL) and filling bits are not used. Signal for end of facsimile coding is indicated by the end of facsimile block (EOFB) with length of 24 bits instead of previously used RTC 78 bits. As a result coding economy grew by almost 40% compared with Group 3's two-dimensional coding.

MRCII code (Modified READ Code II) is able to encode lines with more than 2623 items. This option is needed to achieve higher resolution. Series lengths from 0 to 2623 elements are coded in the same way as in group 3 with main and ending codeword for the relevant color. Lengths longer than 2623 are coded first with main word for 2650, and if the balance exceeds this number code words are added until the balance remains lower than 2650[2]. Then the balance is encoded with code for completion or a combination for Group 3.

[1]Todorka Georgieva is with the Faculty of Electronic, TU Varna,. Telecommunication Dep., Studentska 1, 9010 Varna, Bulgaria,
E-mail: tedi_ng@mail.bg

II. ANALYTICAL REZULTS

Coding scheme uses the method for two-dimensional coding line by line, where location of each changed element of the current code line is coded as per location of the reference standard element situated on the same line or line above it. After coding of this line it becomes reference for the next line.

To reduce the error probability two-dimensional (K - 1) consecutive lines are coded after each one-dimensional coded line. Value of the parameter K is "2" for standard vertical resolution and "4" for additional (higher) resolution.

The coding is done line by line. The position of each changed element of the image is encoded as per the position of the relevant reference standard element situated in the coding line or reference line above it. After encoding, each line is reference for the follow [3] .

Coding procedure:

The following algorithm is performed after selecting appropriate mode:

Step 1: If you select operating mode, that is indicated by the codeword "0001". "A0" element, located under "b2", is established for an initial (reference standard) element for the next coding. If operating mode is not selected, step 2 followed.

Step 2: Absolute value of the relative distance "a1b1" is determined. If $| a1b1 | \leq 3$, then "a1b1" is encoded in vertical mode. Then the "a1" position is regarded as a position of the new reference standard element "a0" in the next encoding.

If $| a1b1 | > 3$, then horizontal coding mode is used. "A0a1" and "a1a2" are coded one-dimensional, the position of "a2" is used as a reference for further encoding.

Coding procedure determines the coding mode which will be used for encoding of each changed item in coded line. The code of selected coding mode is taken from Table 1 end Fig.1.

TABLE I
CODE OF SELECTED CODING MODE

Regime	Elements Coding	Indication	Codeword
PASS mode	b1, b2	P	0001
Horizontal mode	*A0, A1, A1A2*	H	001 + M(a0a1) + M(a1a2)
Vertical mode	a1 in b1 a1b1=0	(0)	1
	a1b1=1	Vd(1)	001
	a1 right of b1 a1b1=2	Vd(2)	000011
	a1b1=3	Vd(3)	0000011
	a1 left of b1 a1b1=1		
	a1b1=2	Vl(1)	010
	a1b1=3	Vl(2)	000010
		Bl(3)	0000010
Extension			0000001xxx

Identification of modified element

Changed element is an element whose color is different from that of previous element on the same scan line [4].

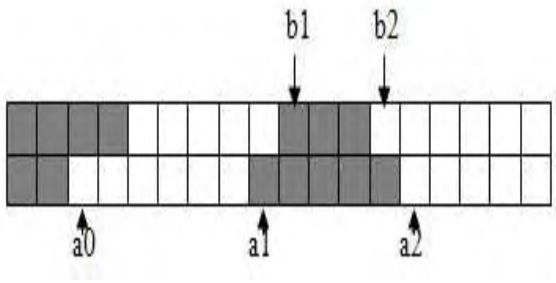


Fig.1. Coding Procedure

“a0” - reference standard or starting changed element of the code line. During encoding of the current line, “a0” position is determined by the coding mode;

“a1” - next changed item on the right of “a0”;

“a2” - next changed item on the right of “a1”;

“b1” - first changed element of the reference line on the right of “a0” and with color different from “a0”;

“b2” - next changed element on the right from “b1” on the reference line [5].

The algorithm for investigation of facsimile codes is shown in Fig. 2.

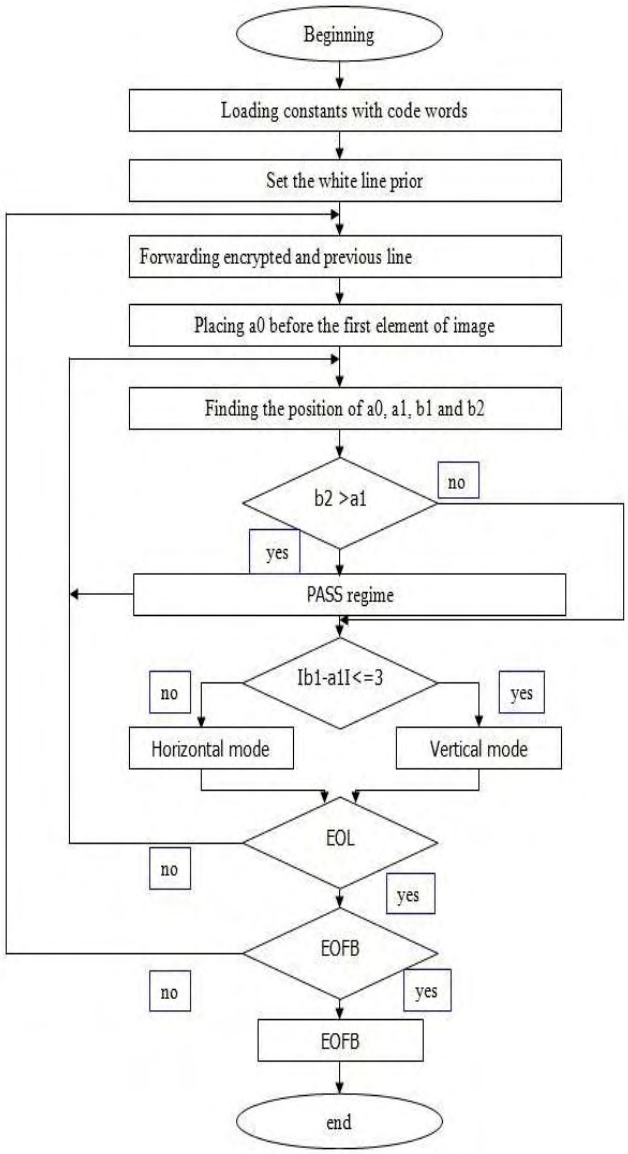


Fig.2. Block diagram of the algorithm for investigation of facsimile codes

PASS mode

It is initialized with codeword 0001. In this mode, "b2" and "b1" must be located to the left of "a1". After coding, "a0" is placed under the position of "b2".

Vertical mode

This mode has seven sub modes : VR3, VR2, VR1, V0, VL1, VL2, VL3 and is indicated when “I a1-b1 I” is equal to or less than 3. “R” indicates that "a1" is on the right of "b1", and “L” – on the left.

Code words are :

a1-b1=0 (V0)……1;

a1-b1=1(VR1)……..011;

a1-b1=2(VR2)……000011;

a1-b1=3(VR3)……..0000011;

a1-b1=-1(VL1)…….010;

a1-b1=-2(VL2)……000010;

a1-b1=-3(VL3)……..0000010;

After coding, "a0" replaces "a1".

Programming model parameters

- **Image [T Bitmap] –** coded image
- **Out Image [T Bitmap] –** decrypted image
- **Out Stream [T Memory Stream] –** stream for encoding / decoding information
- **Out Bits [String] –** presents OutStream in text 1010 .....
- **Coding Method [T Fax Coding Method] –** type of encoding algorithm
- **Pure Bytes [Long Int] –** correct number of bytes that image contains
- **Compressed [Long Int]** – number of compressed bytes

TFaxCoddingMethod = (fcmBlock, fcmHuffman, fcmMHC, fcmREADMHC);

Class Tfax Custom Coding (in file u coding.pas)

- **Code –** code "Image" in Out Stream

Decode – decode Out Stream

Test results are shown in Fig. 3 end

Fig.4 :

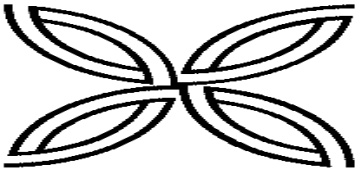


Fig. 3. Original image 1

Fig. 4. Test result with noise intensity 30%;

Coefficient of contraction:

K = 2 : **74,52%**;

K = 4 : **81,88%;**

Performance 0,062 sec (62 ms);

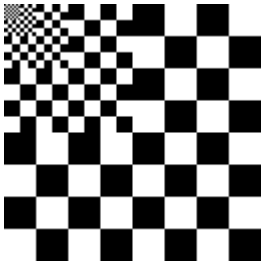


Fig. 5. Original image 2

Experiment with an array: dimensions 128/128 and size 2048byte. [6]

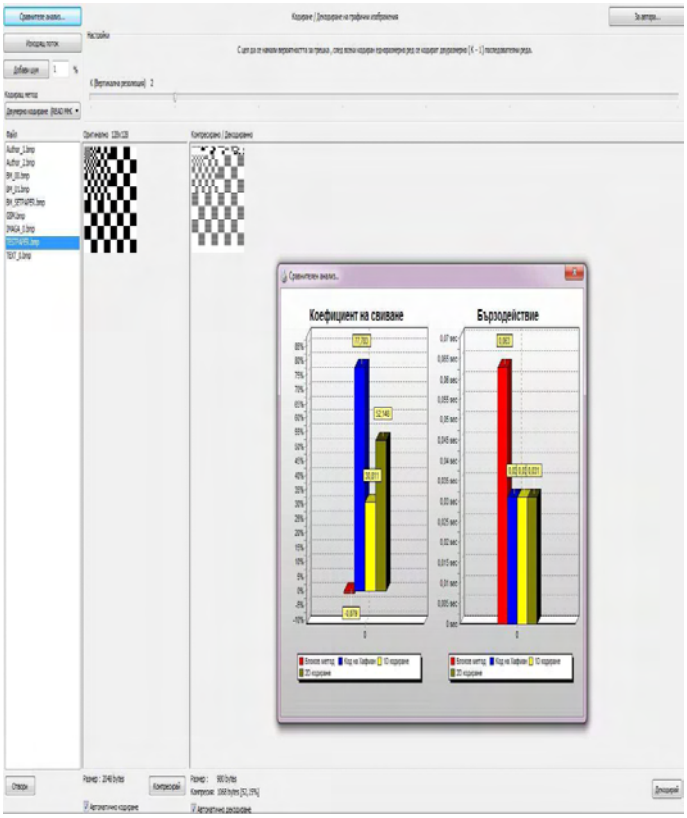


Fig. 6. Experimental results. Coefficient of contraction and performance

III. Conclusion

Different results for the investigated methods are obtained with noise importation in the data channel. Increasing the error probability in data channel, the total number of errors in image reconstruction also increase. At the same time the number of transitions from black to white element are increased. [7]

Modified READ code (MRCII), used for fax machines group 4 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

References

[1] ITU-T Recommendation T4.Standardization of Group 3 facsimile terminals for document transmission

[2] ITU-T Recommendation T6. Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus

[3] ITU-T Recommendation SG8

[4] Long D., Jia W. Optimal Maximal Prefix Coding and Huffman Coding Proceedings of The Seventh International Conference on Distributed Multimedia Systems, Taipei, Taiwan, Sept. 26-28, 2001, pp. 101-107

[5] Chowdhury R.A., Kaykobad M. An Efficient Decoding Technique for Huffman Codes Information Processing Letter, Vol. 81, N. 6, pp.305--308, March 2002

[6] Blelloch G. Introduction to Data Compression, JPEG, MPEG,Computer Science Department Carnegie Mellon University, October 16, 2001.

[7] В.И. Шульгин. Экономное кодирование - Учеб. пособие. – Харьков: Нац. аэрокосм. ун-т «Харьк. авиац. ин-т», 2003.

Design and implementation of a device for a cloudiness measurement

Cvetan Kitov[1]

Abstract – **In the paper are investigated the algorithms for measurement of cloudiness. First is the HYPERION algorithm, second is presented the process of development of a real one device, based on Thermopile sensors. The requirements for such an apparatus for hardware and software development are fulfilled.**

Keywords – **HYPERION algorithm, thermopile sensor, cloudiness measurement, cloud cover.**

I. INTRODUCTION

Analysis of weather conditions is an important part of our daily life. In the busy air traffic, clouds affect the visibility, and hence the security of flights and satellite researches. For these reasons, a development of algorithms and devices to measure the density of the clouds is necessary.

One of the most important features in their development are compact of the device, low cost, ease of use.[1] For this reason are conducted both scientific and practical research in this area. One of the leading organizations working in this direction is the European Organization for the Exploitation of Meteorological Satellites.

Cloud monitoring has been developed for use with cosmic ray air shower fluorescence detectors, the High Resolution Fly's Eye and the Pierre Auger Observatory. This is based on an infrared thermopile device which, unlike previous such monitors, requires no moving chopper and is suitable for unattended operation over long periods of time.[2]

There are various devices and ways to measure cloud cover. One of the easiest ways is the photo. Unfortunately, the data obtained from them is difficult to be processed. In this method, the human factor is crucial and there still remain difficulties in the process of evaluating the data to be automated. The equipment required to perform the algorithms for image processing are expensive, but also dependent on the environmental conditions.

The analysis of cloud characteristics is very important because it gives the method to differ the cloud form the rest of the environment. Tracking them is the basis to make a correct assessment of their parameters (such as density, height, etc.). The main characteristic is the reflection of light. If the reflection of the light rays and the characteristics of reflection are tracked, then any change in the background indicates the presence and type of cloud cover. [3]

[1]Cvetan Kitov is former PhD student in the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: mvn@tu-sofia.bg.

II. ALGORITHM FOR MEASUREMENT OF CLOUDYNESS IMPLEMENTED ON HYPERION

A. The Hyperion

A cloud cover detection algorithm was developed for application to EO-1 Hyperion hyper spectral data. The algorithm uses only bands in the reflected solar spectral regions to discriminate clouds from surface features and was designed to be used on-board the EO-1 satellite as part of the EO-1 Extended Mission Phase of the EO-1 Science Program. The cloud cover algorithm uses only 6 bands to discriminate clouds from other bright surface features such as snow, ice, and desert sand. The code was developed using 20 Hyperion scenes with varying cloud amount, cloud type, underlying surface characteristics and seasonal conditions. Results from the application of the algorithm to these test scenes is given with a discussion on the accuracy of the procedure used in the cloud cover discrimination. Compared to subjective estimates of the scene cloud cover, the algorithm was typically within a few percent of the estimated total cloud cover.[4][5]

The Hyperion devise is a satellite to the Earth with embedded tools for monitoring. On the meeting of EO-1 SVT on Nov 21, 2002 was introduced algorithm for image capture, according to which clouds are excepted from the rest of the earth objects. [6][7] The reflected light from the cloud is used. The image is processed pixel by pixel, and is defined which pixel is covered and which is not. That is the method for assessment of cloud cover as a percentage of the taken image. On the basis of the algorithm is the image capturing in 6 frequency channels:

0.56 μm - used to calculate the index of snowing,

0.66 μm - main channel for testing the effect of clouds reflection,

0.86 μm – it is used with 0.66 μm for testing NDVI,

1.25 μm – differentiation between desert and sands,

1.38 μm – channel testing of high clouds position,

1.65 μm - used to calculate the index of snowing.

For each frequency band i using the following formula to convert the calibrated data in Hyperion L_i in reflections ρ_i:

$$\rho_i = [\frac{\pi}{\mu_0 S_{0,i} / d^2_{earth-sun}}]L_i \qquad (1)$$

The result is a value for each reflection spectrum of each pixel of the image. Based on the values of the reflection ρ_i is developed algorithm, which determines whether a pixel is a reflection of clouds (low, medium high or high), whether or desert sand, snow, ice, water or plant. [7]

B. Algorithm implemented on the device developed and tested during the experiments

In the present paper is chosen for cloud cover measurement the effect of a thermal radiation.

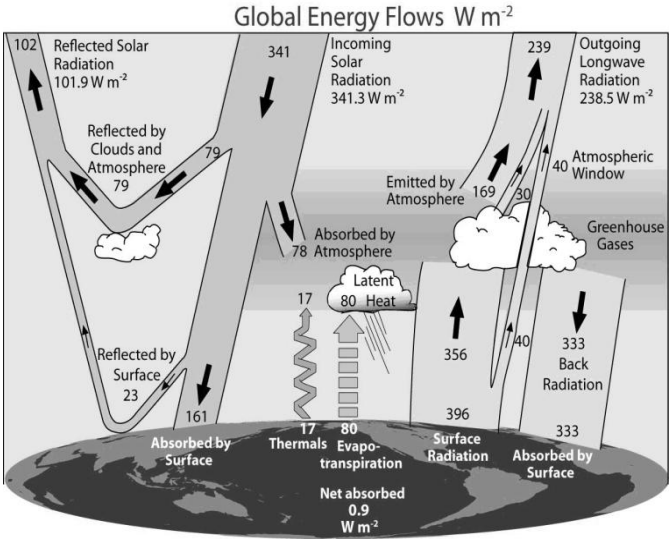


Fig. 1. Global energy flow on earth. Return of greenhouse gases and clouds of energy.[8]

The sun emits heat that reaches the ground. During the process of reaching the Earth the energy is accumulated, depicted on the right part of Fig.1. The Earth itself begins to radiate heat, which is directed towards the sky - left side of Fig.1 On crisp, clear sky, the earth radiated energy is not absorbed and is send back to earth. The more the emitted from the sky energy is, the more dense of the cloud is. It is measured by the difference in radiated from the ground and returned from the objects in the sky heat.

The quantitative assessment of these effects is performed by thermopile sensor.

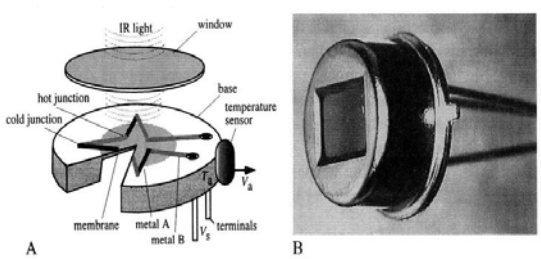


Fig.2. Thermopile sensor

The thermopile sensor is a set of thermocouples connected in series, rarely parallel. The sensor measures the thermal radiation of the objects. When thermal radiation reaches the sensor one of each pair of thermocouples is heated. This couple is called active. Another thermocouple is thermally isolated from the environment.

Between heated and protected thermocouple is created an electric potential which is proportional to the difference in temperatures between the two transitions.

Since the paired thermocouples are connected in series the generated voltages are collected, and as a resultant tension between the outputs of the two sensors.[8]

Extremely favorably during the measurements is the falling of drops of water on the sensitive surface. In this case the water acts as a lens and focuses the sensor at a point in the sky that can be not correct to the overall cloud cover. To avoid such errors in the measurements must be foreseen a system for removal of droplets and drying of the sensors.

C. Requirements for the device

The device for measuring cloudiness is a part of the weather system that also measures temperature, humidity, barometric pressure, rainfall and location of lights and their power. The main function of the device is to measure proven the amount of clouds in a specified perimeter. Assessment of the cloud is made according to a scale consisting of 8 units called octaves. This is an international scale standard METAR [1]. The lowest level of clouds is in the first octave, and the highest in the octave with number 8. The information from the device is collected in a centralized system that processes the results of all instruments and transmits the information to the upper levels of the system for visualization. The protocol for conducting communication between the device and the system is RS485. The message to be generated by the clod measure device has a specific format specified by the system which will received it and will have to process it.

III. THE DEVICE

The device structure is also presented here. The main building blocks are depicted on Fig.2 .

1. *Interfaces*

Because the device is mounted outdoors, the block interfaces, which is RS485 interface is implemented in a weather-proof box. This ensures low contact resistance and high reliability of connection. The communication lines are secured and strengthened against wind and mechanical effects.

2. I/O transmitters block

It consists of interface circuits that convert the signals. The CMOS levels, with which the microcontroller works are not suitable for establishing communication links over long distances. Therefore, are used interface circuits, level transformers that ensure reliable data transmission. A communication channel is build with defined bandwidth and noise protection.

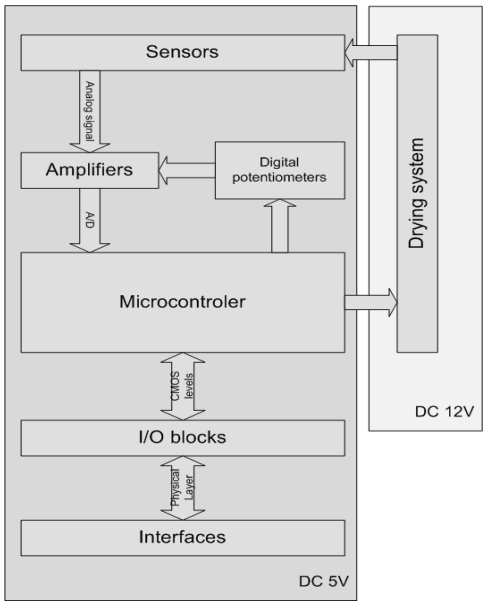


Fig.3. Block Scheme

3. *Microcontroller*

Microcontroller is the main component of the device. It is responsible for the connection between all the blocks of the device, and communication with the centralized system. The microcontroller measures the signals from the sensors, uses built-in A/DC to digitize the data. Then process and prepares them for transmission in a convenient, easy to understand format. Instrument calibration is also performed by the microcontroller in two stages:

1. The sensors are established to generate one and the same signal level. This means that under the same conditions each of the four sensors will show equal values. This represents unification of "zero" on the sensors and ensure that readings are equally relevant for the purposes of measurement.
2. The reporting of the results of the cloud measurement is presented in 8 octaves. To define in which octave is the current value of the cloudiness are defined ranges. The correct values (thresholds) are set, which are the base to pass from one octave to another according to initial levels of the sensors and model device.

The "zero" levels are adjusted on amplifiers with digital potentiometers. The calibrated thresholds are stored in an array of values embedded in the microcontroller EEPROM. In the EEPROM is stored data and device as ID, speed of data transmission and more.

Communication with the centralized system is done using a built-in microcontroller USART interface. It works with 8 bit register; it is possible to configure the start bit, parity check and one or two stop bits. The speed is up to 76,800 boudps.

4. *Amplifiers*

Since the signal level which is provided after the measurement is low, the sensors will produce very similar values for different clouds. This is a prerequisite for a wrong switching between the upper and the lower range. For these reasons, the signal is amplified. Thus, achieving levels similar to those the microcontroller works, used is the full range of the A/DC, which is 10 bits. In this way the thresholds for switching different octaves, are moved away from each other with which reduces the probability of error acquiring in measurement.

For the practical purposes is realized electrical scheme for the digital part of the device.

The electrical circuit scheme is based on the microcontroller ATMega8L. [2] For the realization of interface RS-485 is chosen the DS75176. The element is interface scheme which from the side of the microcontroller operates with CMOS levels, and on the side of the channel to connect a differential amplifier. The amplifier eliminates the noise in the communication channel and amplifies only the desired signal. This feature gives the possibility data to be transmitted on long distances without loss in the payload. The DS75176 is suitable for the purposes of the studied realization, because the source power is 5V, which is the same as the voltage of the processor. The connections scheme and the terminating resistor values are taken from the specification of the integrated circuit.

The signal flow in transmission of the information is as follows: the microcontroller sends bits coded with CMOS levels, which are serially transmitted from the seventh otput of the microcontroller, which is a part of the USART protocol, bits are taken from an agreed scheme on the fourth output D, which is an input working with CMOS. If DE and RE are with a high level then the data is transmitted. If they are with lower level will have to wait for an acceptance from the receiver. The outputs RE and DE are conducted by the microcontroller. There is ability to work in full-duplex or half duplex mode. This option is selected during installation of the device from jumper. In half-duplex mode, DS75176 is in receiving mode on both sides of the line. To start, transmission is checked that the transmission line is free, if so, can begin transmission. In full duplex mode one the scheme is in receiving mode and the other is in transmission mode. The schemes on the other site of the channel are in opposite mode. In receiving mode, the receiving buffer will decode the differential signal from the channel in CMOS levels, which are transmitted to the RXD output of the microcontroller. The bits are stored in 8-bits long register. In the formation of a word a flag is raised in the register of the microcontroller that there is information to be processed.

The power of the digital and the analog parts of the scheme is divided. The reason is the high sensitivity of operational amplifiers.

5. *Software*

Software for implementation in the device is also developed.

The algorithm for management of the cloud measure device can be conditionally divided into two parts:

1. Command interpreter – its purpose is to process information that is entered by the user via RS-485. This information is a string of symbols that have meaning of a command. After the acceptance of the string, it is processed and the command is recognized, it is executed and again goes into standby mode for reception of a new command. The cloud measurer remains in this mode until a command will be entered, get into a break or the power is not switched off.
2. Process of measurement – implemented in hardware interruption by time. The interruption may occur at any time from the work of the command interpreter. The current values of the registers are saved in a stack of interruptions. Runs the code in the breaking. Then from the stack of interruptions are recovered the old values of the registers and the system continues the execution of the command interpreter.

The design of the device includes not only implementation of the defined algorithm, but also a hardware realization.

6. Hardware design

The instrument measures the local cloud cover by Thermopile sensors, which records the thermal radiation The device is working on a comparative basis.

1. The device is a part of the weather system. Successfully communicate with the central module of the system. The format of the data is correct, consistent with the system and it recognizes successfully results.
2. The device is compact. The integrated shell gives possibility of an easy change of the parameters of the device and fast calibration.
3. With the embedded mechanisms for self diagnosis the device reports dropped power of the amplifiers or system for drying, decisions for automatic power on of the system.
4. With the use of calibration arrays of the measurement results are accurate and are not influenced by changes in the environment.
5. The design of the device includes the most modern components. The microcontroller has a built-in mechanisms to monitor power and low power consumption. The appliance has been tested in a specialized laboratory for electromagnetic compatibility and has passed the tests successfully.

IV. CONCLUSION

After the calibration of operational amplifiers and proper setting of arrays according to which is determinate the octave of the cloud, measurements are accurate and report the actual cloud cover. The device has been tested several months. During the test period is checked the behavior of the device in climate changes from winter to spring. The changes of atmospheric environmental conditions were successfully ignored by the measurement and it remains within the seasons change. To achieve this, several arrays were used to determine the octave of the cloud. The unit automatically determines which array to work with on the basis of the reference sensor. The drying system was also tested in real conditions. In rainy weather the system works successfully, and the events for the selection of automatic activation of the system are appropriate and work.

After the tests the device was approved for use.

The command interpreter provides excellent opportunities to configure all device parameters. This makes it more flexible and convenient for use in different conditions. Using Remote Desktop, interpreter allows remote configuration, which is especially useful for system maintenance. Thus saving much time and effort to remove the smaller problems.

REFERENCES

[1] http://en.wikipedia.org/wiki/METAR, Cloud_reporting
[2] http://www.atmel.com/Images/doc2486.pdf, pp. 1-302
[3] Microchip Corp., "*Single/Dual Digital Potentiometer with SPI™ Interface*", MCP42010, pp.1-32
[4] M. Griffin, Su M. Hsu, H. Burke, S. Orloff, and C. Upham, "Examples of EO-1 Hyperion Data Analysis", LINCOLN LABORATORY JOURNAL, VOLUME 15, NUMBER 2, 2005, pp.271-298
[5] D. Tran, S. Schaffer, G. Rabideau, A.G. Davies, T.Doggett, R. Greeley, F. Ip, V. Baker, J. Doubleday, R. Castano, D. Mandl, S. Frye, L. Ong, F. Rogez, B. Oaida, "Onboard classification of hyperspectral data on the Earth Observing One mission", IEEE Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009
[6] National Semiconductor Corp., DS75176 - Multipoint RS-485/RS-422 Transceivers - DS75176
[7] Vaisala Corp., CL51 Ceilometer for High-Range Cloud Height Detection, datasheets
[8] US Patent 4355896, 26 Oct 1982, datasheet

Classification and comparative analysis of localization approaches for Wireless Sensor Networks

Ivanka Tsvetkova[1], Plamen Zahariev[1], Georgi Hristov[1] and Mihail Iliev[1]

Abstract – **The aim of this paper is to investigate and study the various localization approaches, which are used in the modern wireless sensor networks (WSNs). In the Introduction we shortly describe what the WSNs are and what their main characteristics and some of their purposes are. In the second section of the paper we investigate and analyse the radio-based approaches for localization in the WSNs. Later we present the main characteristics of the satellite-based approach for localization, and its implementation for the needs of the WSNs. In chapter four we shortly present the sound-based approaches for localization. In the next chapter we conduct a short comparative analysis of the localization approaches and we highlight the main disadvantages and issues with them. The paper is then completed by the conclusion section, followed by the acknowledgment and references sections.**

Keywords – **Wireless sensor networks, localization approaches, radio-based localization, satellite-based localization, sound-based localization.**

I. INTRODUCTION

The rapid rate of development in the fields of telecommunication and computer sciences has led to the emerging of several new technologies and paradigms for networking. One of these new ideas was the development of tiny mobile devices with sensing capabilities and with the possibility for wireless data delivery. Since the initial introduction of these devices, they have been found as suitable for a variety of different purposes - from animal movement and population monitoring, trough warning systems and systems for detection of hazardous agents and radiation to the latest military purposes - as vehicle and troops tracking and monitoring systems [1, 2, 3]. Despite the variety of sensor devices and their many purposes, there are several disadvantages of the networks they form. One of the current issues is the limited sensor lifetime due to the battery capacity of the devices. Another open issue, which is a consequence of the demand for smaller and more compact sensor devices, is their limited storage and computation capabilities [4]. The sensor motes are equipped with wireless communication interface and are using the 433 MHz, 868-915 MHz and the 2.4 GHz bands [1]. There are two modulation formats available - two-tone Frequency-Shift-Keyed (FSK) at 433 and 868-915 MHz and direct sequence spread spectrum (DSSS) at 2.4 GHz supporting the 802.15.4/ZigBee standard. All radio interfaces are commonly bidirectional, and support a range of 10 to 100 meters [2, 3]. There are several different data routing approaches in the WSN. The direct routing approaches are most easy to use, but they suffer from many disadvantages and are not suitable for large scale WSNs. The flat routing approaches provide more efficient ways for energy consumption, which they achieve by retransmission of the data towards the bases station using sensor motes in close proximities and by aggregation of the information. The hierarchical routing approaches are the most widely used ones. They define that the network has to be organized into clusters, which are controlled by cluster heads. One of the largest advantages of the WSNs is the possibility to locate and track different objects. This is also a prerequisite for the efficient work of the directed and the hierarchical routing approaches, since they rely on the location of the different neighbouring sensor devices in the sensor field.

[1]Ivanka Tsvetkova is a PhD student at the Department of Telecommunications of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: itsvetkova@uni-ruse.bg.

[1]Dr. Plamen Zahariev is a principal assistant professor at the Department of Telecommunications of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: pzahariev@uni-ruse.bg.

[1]Dr. Georgi Hristov is an associate professor and Vice Head of the Department of Telecommunications of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: ghristov@uni-ruse.bg.

[1]Dr. Mihail Iliev is professor at the Department of Telecommunications and Vice Rector of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: itsvetkova@uni-ruse.bg.

II. RADIO-BASED LOCALIZATION APPROACHES FOR WIRELESS SENSOR NETWORKS

Some of the most widely used approaches for localization in the wireless networks are the radio-based approaches. There are two main classes, but they both use the information from the radio interfaces to calculate the range to a certain device. These approaches do not require additional hardware components and instead use the build-in radio interfaces of the devices.

Localization using the Radio Hop Count

The first of the radio-based localization approaches, which we will investigate, is the radio hop count approach. This is one of the most inaccurate approaches [5], but yet it is a usable approach in certain cases and based on the purpose of the network. As it is widely known the radio interface of every device has a limited range, let's say *R* meters. So when sensor device *A* is communicating with device *B*, then *A* has to be no more than *R* meters away from *B*. As it is to notice by this statement there is an error margin of *0*R* to *1*R* meters, since in the two border conditions the nodes can be either on top of each other (error of *1*R* meters) or they can be on R

meters away from each other (error of *0*R* meters). Even though that this approach provides low level of localization accuracy it still can be used for a number of purposes like for instance in the routing processes for the WSNs. There is a number of routing approaches which rely on the radio hop count to find a shortest path to a given target device or to decide, if it is effective to transmit data between two devices [5, 6]. Similar to the routing protocols, which are used in the IP networks, there are routing protocols in the WSNs, which use the local connectivity information provided by the radio interfaces to form a connectivity graph. In this graph of the network the vertices are the sensor devices and the edges are representing the radio links between them (Fig.1).

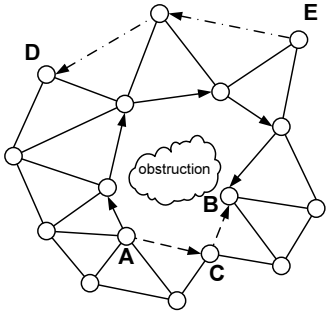


Fig.1. Examples of most common issues with the hop count localization based approach

Based on this, the hop count H_{AB} between any two sensor devices *A* and *B* can be calculated using Dijkstras shortest-path algorithm [7]. By knowing the value of the hop count, we can define that *A* and *B* are at the most $R*H_{AB}$ meters away. As seen in Fig. 1 this statement is correct, but the value of the Euclid distance between both devices is many times smaller than $R*H_{AB}$. This is the first of the three major problems of this approach. The second one is the fact that two different in length routes can have the same value for hop counts. This can be seen in Fig.1 for the routes between devices *E* and *D* and between devices *A* and *B* ($H_{ED}=H_{AB}=2$). This problem is specifically to be taken into consideration when there are multiple equal in hop count routes between two devices. As a result of this, the inappropriate route can be selected, which will lead to a significant delay or to an ineffective energy consumption (both due to the larger distances the data will have to travel). The third and last problem with this approach is the fact that different obstacles can prevent edges from appearing in the connectivity graph. This can be demonstrated by removing device *C* (due to the obstacle) from Fig.1. This action will lead to the removal of the path between *A* and *B* trough *C*. The new path will have a hop count of *5*, but again this will not necessary mean that the devices are *5*R* meters away.

Received Signal-Strength Indication

The RSSI approach is significantly more complex than the radio hop count approach, but can provide more accurate estimation of the location of the devices [8]. The main idea behind this approach is to use the widely known fact, that the energy of the radio signals decreases with the distance travelled. The receiver can then use this statement to calculate backwards the distance to the transmitter based on the strength of the received signal. This approach would work perfectly in ideal conditions, but in the real world the propagation of the radio signals is not uniform and this leads to a certain degree of errors [9]. Additionally the environment, in which the sensor nodes are placed, can contribute significantly to the localization errors, due to the fact that many physical objects and surfaces can absorb or reflect the radio waves. In the real world, when this approach is used for localization in the network, two devices, which are centimetres away, but separated by a concrete wall, can appear to be positioned many meters away one from another.

III. SATELLITE-BASED LOCALIZATION APPROACHES FOR WIRELESS SENSOR NETWORKS

The idea behind every localization technique is to determine the coordinates of a given object. Unlike some of the localization methods (like RSSI and Radio Hop count), which determine the location of the sensor devices relatively and based on some local known coordinates, there is also an approach, which can determine the global coordinates of the devices. This means that the coordinates of every device can be compared towards already known global points, obtained by any satellite-positioning system, or even better - all of the devices in the sensor field can be equipped with a satellite-positioning sensor boards meaning that they can obtain their global coordinates by themselves [10].

In order to implement the satellite-based localization approach in a 2-D based sensor network, there must be at least 3 noncollinear sensor devices, which will act as referent points. If the localization is performed in 3-D then there must be at least 4 such noncollinear referent devices [11]. These sensor devices are regular sensor motes, but they have the means to obtain their global position (GPS/GLONASS modules) or they know their global coordinates a priori (hard copied coordinates). The consequence of using satellite-based localization approach is the decreased time for localization of the devices and increased device performance (no need to perform calculations to obtain coordinates or to store RSSI or Hop count data), since the global coordinates are directly received. However, this approach is rarely implemented since GPS receivers are very expensive, which is in direct controversy to the idea to have very cheap sensor devices (less than 1 USD). Another major issue with this approach is the fact that the GPS receivers can rarely be used indoors and their reading are sometimes influenced by environmental obstacles. Additionally the satellite receivers require huge amounts of energy to operate, which is a big problem in the energy-limited wireless sensor networks. A solution for this issue is to use predetermined data about the global coordinates of the devices. This alternative is sometimes very impractical or impossible. Placing a large number of nodes on specific coordinates is a difficult and time-consuming problem, and sometimes the nature of the environment, where the sensor networks is to be deployed does not allow it (like for instance

when placing the sensor devices on the bottom of a river or when the devices have to be mobile).

IV. SOUND-BASED LOCALIZATION APPROACHES FOR WIRELESS SENSOR NETWORKS

Localization using sound is not a new paradigm and there are many animals, which use sound to locate they pray or to draw they movement plan. In the modern wireless sensor networks there are several approaches, which use sound for localization, and here we will discuss the two most used ones – Time Difference of Arrival and Angle of Arrival.

Time Difference of Arrival (TDoA)

The idea of TDoA is to have all sensors equipped with microphones and speakers. As already known, radio waves travel substantially faster than sound waves [12, 13], so the idea is for the source to transmit a radio signal and the sink to mark the time of the reception, then the source to transmit a sound pattern and the sink again to mark the time when it is received (Fig. 2). After this the sink can compute the distance to the source using the following equation:

$$d = (s_{radio} - s_{sound}) * (t_{sound} - t_{radio} - t_{delay}), \quad (1)$$

where d is the distance between the sink and the source, and t_{sound}, t_{radio} and t_{delay} are correspondingly the time interval, which determines the reception of the sound pattern, the time interval, which determines the reception of the radio signal and the time interval, which the source has waited before transmitting the sound pattern (t_{delay} can be zero).

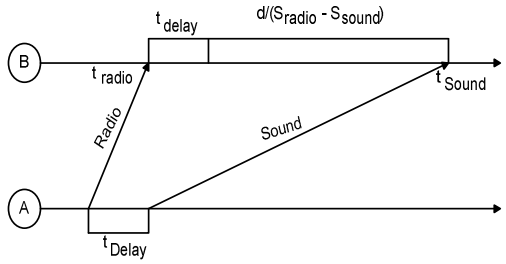


Fig.2. TDoA example, where node A is sending a radio signal and a sound pattern to node B and it is using the difference in the reception times to determine the distance to the signals source

There are several disadvantages of the TDoA approach. In order to implement TDoA localization in a WSN, all motes need to have a speaker and a microphone, which will increase the price of the devices [12]. Additionally the approach requires allocation of extra computational and storage resources. TDoA systems may require also special calibration, since speakers and microphones have different characteristics [13]. Finally there are some environmental requirements for the TDoA to be functioning correctly - line-of-sight between the source and the sink, constant humidity and air temperature and etc. This all has proven to be a holdback for the further development of this approach for its use in the wireless sensor networks.

Angle of Arrival (AoA)

The AoA approach is an amendment to the TDoA approach [14]. The idea is to have not one, but several spatially distributed microphones, which are to hear a single transmitted signal and by analyzing the phase or time difference between the signal's arrival at the different microphones to determine the AoA of the signal [15].

This method has proven to provide accuracy down to a few degrees, but unfortunately suffers from the same disadvantages as the TDoA.

V. OPEN ISSUES WITH THE LOCALIZATION APPROACHES IN THE WSNS

Since the wireless sensor networks are a special type of telecommunication networks there are some additional factors (compared to the traditional communication networks), which have performance impact on the localization approaches. There are three major categories of issues with the localization approaches for wireless sensor networks – environmental-based, device-based and network-based issues.

Environmental-based localization issues in the WSNs

The environmental obstacles are something, that the designers of the WSNs are rarely taking into consideration, but can be something that can have a huge impact on the performance of the localization approaches [16]. Large obstacles can interfere with radio interfaces and signals, introducing errors when the RSSI or the Radio hop-count approaches are used. Additionally they can shadow the GPS receivers and prevent the direct line of sight for many of the localization approaches. The deployment of the devices on various surfaces and in various atmospheric conditions can affect the radio-based and sound-based localization approaches as well. Placing the devices indoors and outdoors can also have a huge impact on the localization processes. As a overall conclusion the environmental-based localization issues are not something that the networks can have a impact on, but the sensor devices have to be designed or configured in such a way, so that the effects of the environment on the localization performance has to be reduced to a minimum.

Device-based localization issues in the WSNs

The idea about sensor networks is to have thousands of tiny inexpensive devices which are to detect certain events or to monitor environmental parameters and to transmit the sensed data using a wireless network. This means that due to the restrictions for the size of the devices they will have to be equipped with low size processors, memory chips and other components, which makes extensive computation and data storage impossible [17]. Moreover, sensor nodes are typically battery powered. This means that the wireless communication is to be performed on short ranges in order to conserve power. Since the localization is often not the primary task of a WSN, but is nevertheless used so that the network can perform its

normal operations, it should be performed at the lowest possible power cost, hardware cost and deployment cost.

Network-based localization issues in the WSNs

There are several parameters of the networks which are having an impact on the localization approaches and their accuracy. One of those parameters is the density of the nodes. The radio hop count and sometimes the satellite based approach (when a given number of devices have GPS receivers they have to be in communication range so that multilateration can be used) require high node density for their normal operations. Other approaches can be affected negatively by the increased density of the devices, since interference (both radio and sound) can occur.

An additional problem, which is defined by the network topology, is the unequal distribution of the nodes. This can lead to uneven levels of localization accuracy for certain areas of the network compared to the localization accuracy in areas with denser sensor mote population. Another issue is how to obtain the accurate coordinates of the edge nodes in the network. These devices are the last of the nodes in the sensor field and can be observed only from the inside of the network, which means that they are localized using partial information.

The data delivery and processing mechanisms of the wireless sensor networks can also affect the localization processes. Based on the network topologies or organization methods there are couple of approaches for localization. In the hierarchical cluster based WSNs the cluster heads can obtain their coordinates using the data sent from the base station. After that the cluster heads are being used by the sensors in the clusters, so that they can obtain their coordinates [17].

Different approach is to use an algorithm to roughly determine the coordinates of the nodes in the network. After that the location data can be improved either by the use of a different localization approaches or by a second run of the same approach, but with a different filters and settings.

There is also an approach in which the localization of the sensors is performed locally for a given group of sensors (like in a cluster). Finally the data for all of the sensor groups is merged together to form a map with the coordinates of all of the devices in the network.

VI. CONCLUSION

As seen by the analysis in this paper there are a number of localization techniques available for implementation in the WSNs. All of these approaches have their advantages and disadvantages and are suitable for specific network applications. Typically the sound based localization approaches achieve better accuracy than the radio-based localization approaches. The highest localization accuracy is provided by the satellite-based localization approaches, but nevertheless, this accuracy is achieved at the expense of higher equipment cost. The general conclusion is that when designing a wireless sensor network, one has to take into consideration all of the factors that will impact on the accuracy of the localization approaches - the characteristics of the environment, the hardware constrains of the devices and the particular specifics and features of the network itself. After that the appropriate approach can be selected based on the purpose and the requirements of the network.

ACKNOWLEDGEMENT

The work presented in this paper is completed as a partial fulfilment of 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 and Science.

REFERENCES

[1] I. Stojmenovic, Handbook of Sensor Networks: Algorithms and Architectures, Wiley, ISBN-13: 978-0471684725, 2005
[2] Mikhov M., L. Pachemanov, Two-Coordinate Driving System with Wireless Control, Annals of the University of Targu-Jiu, Vol. 10, No. 3, pp. 180-188, Romania, 2010, ISSN 1842-4856.
[3] Georgiev Ts., M. Mikhov, A Sensorless Speed Control System for DC Motor Drives, Scientific Journal of RTU, Vol. 25, No. 4, pp. 155-158, Riga, Latvia, 2009, ISSN 1407-7345
[4] Arnaudov R., G. Stanchev. Pulse Current Controller, Proceedings of ICEST’2002, Nish, Yugoslavia, October 2002.
[5] S. Dulman, M. Rossi, P. Havinga, M. Zorzi, On the hop count statistics for randomly deployed wireless sensor networks, Int. J. Sensor Networks, vol. 1, Nos. 1/2, 2006
[6] S. Yang, J. Yi and H. Cha, HCRL: A Hop-Count-Ratio based Localization in Wireless Sensor Networks, in Proceedings of SECON‘07, San Diego, California, pp 31-40, 2007
[7] E. Dijkstra, A note on two problems in connexion with graphs, Numerische Mathematik, vol. 1, pp.269-271, Springer, 1959
[8] U. Ahmad, A. Gavrilov, S. Lee, Y.Lee, A modular classification model for received signal strength based location systems, Journal on Neurocomputing, vol. 71, Elsevier, 2008
[9] X. Wang, Sh. Yuan, R. Laur, W. Lang, Dynamic localization based on spatial reasoning with RSSI in WSNs for transport logistics, JSA, vol. 171, Issue 2, Elsevier, pp 421-428, 2011
[10] Hung-Chi Chu, Rong-Hong Jan, A GPS-less, outdoor, self-positioning method for wireless sensor networks, Journal on Ad Hoc Networks, vol. 5, Issue 5, Elsevier, 2007
[11] N. Bulusu, J. Heidemann, D. Estrin, GPS-less Low Cost Outdoor Localization For Very Small Devices, IEEE Personal Communications Magazine, vol. 7, No. 5, pp. 28-34., 2000
[12] K. W.K. Lui, H.C. So, A study of two-dimensional sensor placement using time-difference-of-arrival measurements, Journal on Digital Signal Processing, vol. 19, Issue 4, 2009
[13] K. Dogancay, A. Hashemi-Sakhtsari, Target tracking by TDoA using recursive smoothing, JSP, vol. 85, Issue 4, Elsevier, 2005
[14] P. Kulakowski, J. Vales-Alonso, E. Egea-López, W. Ludwin, J. García-Haro, Angle-of-arrival localization based on antenna arrays for wireless sensor networks, Journal on Computers & Electrical Engineering, vol. 36, Issue 6, pp 1181–1186, Elsevier
[15] S.B. Kotwal, R. Abrol & S. Nigam, Sh. Verma, G.S. Tomar, Analysis of One Anchor RSSI and AoA Based Iterative Localization Algorithm for WSN, International Journal of Simulation Systems, Science & Technology, vol. 11, pp 35-41
[16] M. Iliev, I. Tsvetkova, N. Bencheva, D. Radev, Simulation of an algorithm for positioning of objects in sensor networks, Proceedings of the University of Ruse, vol. 48, pp 42 – 47, 2009
[17] I. Tsvetkova, Y. Aleksandrov, G. Hristov, P. Zahariev, M. Iliev, Comparison of target tracking algorithms in hierarchical WSNs, World Conference on Information Technology WCIT 2011, Antalya, Turkey

Comparative analysis of routing approaches for wireless sensor networks

Plamen Zahariev[1], Georgi Hristov[1], Ivanka Tsvetkova[1] and Mihail Iliev[1]

Abstract – **The aim of this paper is to present a simulation study of the three most common approaches for routing of information in the wireless sensor networks (WSNs). In the introduction of the paper we shortly describe some of the characteristics of the WSNs, and then we present some of their main implementation areas. In the second section we present the theoretical aspects of the approaches for flat, directed and hierarchical routing. Later we present a comparative analysis of the three approaches and we highlight their main advantages and disadvantages. In the fourth section we present the results of the conducted simulation experiments with the three approaches and we discuss and analyse them. The paper is then completed by the conclusion section, followed by the acknowledgment and references sections.**

Keywords – **Wireless sensor networks, hierarchical, directed and flat routing.**

I. INTRODUCTION

The wireless sensor networks are combining the benefits of the modern technologies for detection and sensing with the possibility to transmit the data using the wireless medium. These networks are found suitable and are being used for numerous of different purposes including animal and insect monitoring, intelligent and autonomous housing, vehicle and people tracking, intrusion detection and prevention, military purposes and other [1]. There are many different issues currently open with these networks, and they all are a consequence of the limited resources and the small size of the sensor devices [2]. One of the biggest challenges is to present a routing approach, which is corresponding to the needs of the sensor networks and is suitable to the hardware and software capabilities of the devices [3, 4]. There are several major approaches for routing of data in wireless sensor networks – direct routing, flat routing and hierarchical routing [5]. These three approaches are used separately and interchangeably and are dependent on the architecture of the network and its purpose. The first type of architecture, which we will investigate, is characterized by the fact that all of the devices in the network (except for the base station) are performing the same tasks and functions. This means that all of the devices are able to perform sensing tasks, to receive information from other devices and to send data to the other sensor motes or to the base station. These are the necessary prerequisites to implement the direct and flat routing approaches [6]. The second type of architecture is characterized by the fact that the devices, which are forming it, are not performing the same tasks and functions. The networks, which implement this type of architecture, are formed by two types of sensors. The first class of sensor devices is capable of performing the basic sensing and communication tasks while the second class of devices is performing specialized actions and functions, like cluster formation, data aggregation and other [7].

[1]Dr. Plamen Zahariev is a principal assistant professor at the Department of Telecommunications of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: pzahariev@uni-ruse.bg.

[1]Dr. Georgi Hristov is an associate professor and Vice Head of the Department of Telecommunications of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: ghristov@uni-ruse.bg.

[1]Ivanka Tsvetkova is a PhD student at the Department of Telecommunications of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: itsvetkova@uni-ruse.bg.

[1]Dr. Mihail Iliev is professor at the Department of Telecommunications and Vice Rector of the University of Ruse "Angel Kanchev", Studentska str. 8, Ruse 7013, Bulgaria, e-mail: itsvetkova@uni-ruse.bg.

II. DIRECT, FLAT AND HIERARCHICAL ROUTING IN THE WIRELESS SENSOR NETWORKS

The direct routing approach (Fig. 1) is one of the oldest approaches for data transmission in the wireless sensor networks. This approach defines, that the sensor devices are to be accessed using data about their coordinates or their location [8]. By using this approach the Base Station can request information about the occurrence of an event from a given area on the sensor field instead to request the data from all of the sensor devices in the network.

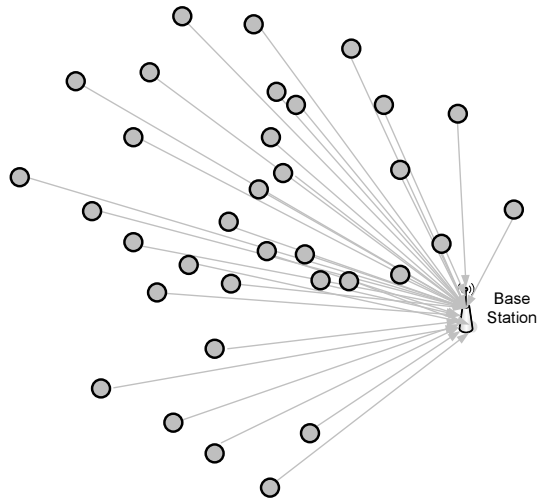


Fig.1. Example of the direct routing approach in the wireless sensor networks

The most common approach to determine the location of the sensor motes is to use the received signal strength indication (RSSI). Alternative approach for localization of the devices is the direct communication with a satellite, trough a GPS interface (if the mote has one).

The direct routing approach defines that the sensor devices should communicate direct with the data sink (either the Base Station or a neighbor sensor mote). Unlike this approach the flat routing defines that the sensor devices can communicate with the nearest device towards the sink thus reducing the communication distances and by that also decreasing the amount of energy required for the communication processes (Fig. 2). Additionally, when this approach is implemented, the intermediate devices can use the data from the transmitting device and if possible can aggregate it with the data, which it has personally collected from the sensed area. The flat routing approach, similar to the direct routing approach is ideal for implementation in information-orientated wireless sensor networks [8]. Additionally this approach is much more suitable for implementation in large scale WSNs, since due to the retransmission of the data, the sink will receive information from a smaller number of devices compared to the direct routing approach. This will provide also better quality of service and smaller delays.

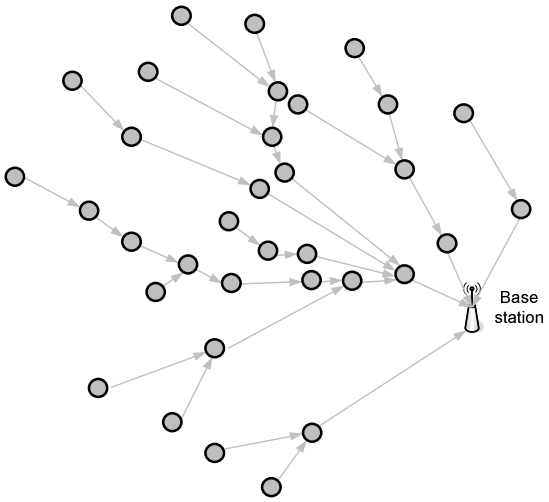


Fig.2. Example of the flat routing approach in the wireless sensor networks.

The hierarchical (also known as cluster based) routing approach (Fig.3) is initially proposed for implementation in the traditional Ethernet networks and is widely used because of its unmatched communication efficiency and because of its high level of tolerance towards network scalability [9, 10].

These are the main reasons for the exceptional results, which this approach shows in terms of effective and balanced consumption of the energy of the sensor motes. With the hierarchical architecture, the motes, which are having more energy, are used for data processing, aggregation and communication, while the sensor devices with less energy are used for low energy processes, like sensing and environmental monitoring [11, 12]. This idea is accomplished by organizing the network into clusters and by defining roles to the specific devices in the cluster. The network can be formed into several levels where the lowest level is consisting of sensor motes, which communicate with a cluster head, and the higher levels are formed by cluster heads, which are communicating either with other cluster heads or with the base station [13].

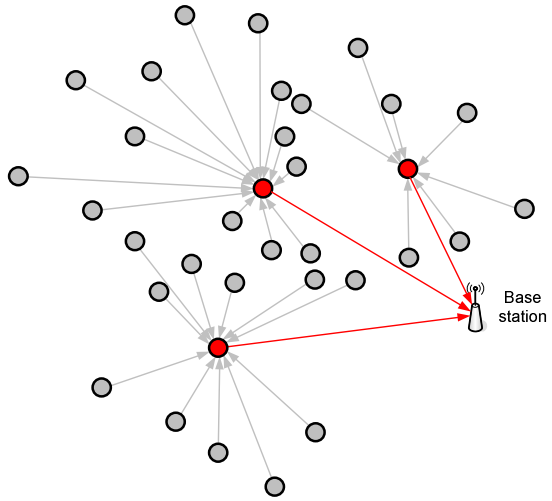


Fig.3. Example of the hierarchical routing approach in the wireless sensor networks

Typical representative of the hierarchical cluster based routing approaches is the LEACH protocol [5, 10]. The approach for hierarchical organization of the network, which is defined by this protocol, states that the first layer of the network will consist of devices, which are in communication range with the base station. The devices of every layer will be able to communicate with the sensors from the previous one, but not with the devices from the layers before that.

After the conclusion of this process the total number of the clusters N_{CL} in the system can be divided by the number of the layers N_L, and thus obtaining the average number of clusters per layer $N_{CL}{}^{avg}$. Assuming that the devices are distributed equally on the sensor field, then the average number of devices per layer n^{avg} can be calculated by dividing the total number of sensor nodes n by the number of the layers N_L:

$$n^{avg} = n / N_L \tag{1}$$

After the average number of devices is determined, they are distributed equally in the corresponding clusters for every layer.

By using these rules, the possibility for a certain module P_{ni} from a given layer to become a cluster head for the communication round t can be given by:

$$P_{n_i}(t)\begin{cases} \dfrac{N_{CL}{}^{avg}}{n^{avg} - N_{CL}{}^{avg}\left(R_e \bmod \dfrac{n^{avg}}{N_{CL}{}^{avg}}\right)} & when\ C_{n_i}(t)=1 \\ 0 & when\ C_{n_i}(t)=0 \end{cases} \tag{2}$$

In the equation above, R_e is the number of elapsed communication cycles from the start of the communication round and $C_{ni(t)}$ is a variable, which has value of 0 when the device has already been a cluster head during this communication round and 1 otherwise. After the completion of the process for organization of the hierarchical wireless sensor network, the system is divided into clusters of equal number of sensor modules.

III. Comparative Analysis of the Routing Approaches for WSN

In order perform a comparative analysis between the direct, the flat and the hierarchical routing approaches, we have decided to collect the main advantages and disadvantages of these approaches and present them using a table.

Table 1. Comparative analysis of the advantages and disadvantages of the three routing approaches for WSNs

	Advantages	Disadvantages
Direct routing approach	• Effective dissipation of the energy needed for communication, when the network is formed by small number (less than 100 nodes) of sensor devices. • Accurate estimation of the length of the communication round. • The amount of the traffic in the network is proportional to the number of the devices. • It is possible to use addressing schemes to distinguish the devices	• Requires accurate localization of the sensor devices. • Large delays during data transmission in networks, which consists of thousands of sensor devices. • Large possibilities for collisions accuracies in systems with shared transmission mediums (especially in large networks with thousands sensor motes). • Unequal energy dissipation (due to the unequal distances to the sink).
Flat routing approach	• Effective dissipation of the energy needed for communication, when the network is formed by large number (>100 nodes) of sensor devices. • Possible data aggregation, when the data is send to the base station. • Smaller amounts of network traffic (due to the data aggregation).	• Use of only information-oriented architecture. • Impossible to determine accurate the length of the communication round. • Delays in networks with large number of sensor devices (due to the larger number of intermediate hops). • Unequal energy dissipation and processing loads (due to the unequal number of hops for the devices further away from the Base Station).
Hierarchical routing approach	• Effective dissipation of the energy needed for communication, when the network is formed by large number (>100 nodes) of sensor devices. • Data aggregation in the cluster and between the clusters (higher data integrity and reliability). • The amount of the data traffic in the network depends on the amount of the cluster head devices. • Possible implementation of mechanisms for more effective and balanced energy dissipation. • Possible use of both information-oriented and addressable architectures.	• Impossible to determine accurate the length of the communication round. • Somewhat unequal energy dissipation and processing loads (due to the different roles of the devices).

IV. Simulation Experiments and Evaluation of the Routing Approaches for WSNs

In order to prove the statements from Table 1 and also to further investigate the routing approaches, we have conducted a series of simulation experiments with them. For this purpose we have developed several MatLab models for all of the approaches. We have then implemented and tested the model for a sensor network, which is deployed on a sensor field with dimensions of 100x100 meters. The network consists of 1000 random distributed nodes and a base station at x=50 m and y=50m (center of the field). In order to evaluate the approaches we use the values for the dependence between the communication rounds and the total amount of the energy in the networks. Additionally we evaluate the number of active sensor devices per communication round for the three approaches. The initial amount of energy for all sensor devices is the same and is equal to 0.1 J.

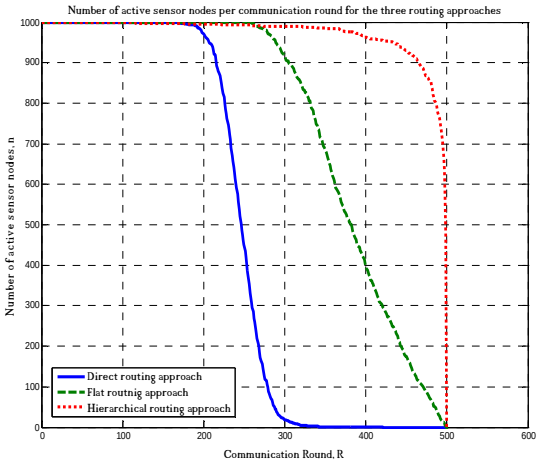


Fig.4. Number of active sensor devices per communication round

As seen on Fig. 4, in the network where the direct routing approach is used, the number of active sensor motes is starting to decrease around the 180-th communication round. This can be explained by the fact that the distance between the nodes in the sensor field and the base station is not the same, which will lead to the faster dissipation of the energy in the sensor devices farther away from the base station. The simulations with the network, where the flat routing approach is implemented, provide significantly better results, but nevertheless they show that the sensor devices start to deplete their energy around the 250-th communication round. This can be explained by the multiple retransmissions of the data towards the base station by the devices. Despite the fact that the hierarchical routing approach provides the best results, it also shows one of the largest disadvantages. This is the early dissipation of the energy by some of the devices in the network, and can be observed by the slight slope of the line towards the horizontal line. Actually during this simulation experiment the first module consumes its energy around the 104-th round. This can be explained entirely with the unequal data and processing load of the devices, which is a consequence of the different roles they have in the network.

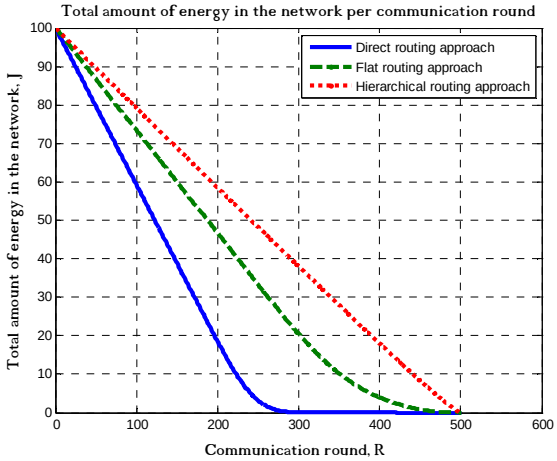


Fig.5. Energy dissipation ratio of the sensor devices per communication round

Fig. 5 presents the ratio of energy dissipation in the networks when the three routing approaches are used. It is easy noticeable that again the approach for direct routing is showing the worst results. This is again because of the much greater distances the data has to be sent, compared to the other two approaches, where the data is being retransmitted either to a closer neighbouring sensor device or to the cluster head.

V. Conclusion

As a conclusion we can state that the approach for hierarchical routing can significantly increase the efficiency of the system. The results, which are demonstrated by this approach, show that it is approximately two times more efficient compared to results from the simulations with the approach for direct routing. When comparing the flat routing approach with the hierarchical routing approach, we can notice a 25% better efficiency in terms of average lifetime of the sensor nodes and in terms of balanced energy consumption. Based on the conducted analysis, we can state, that the hierarchical routing approach is definitely the best approach among the three investigated, but it also suffers from few, but serious disadvantages. An eventual improvement of the mechanisms for effective and balanced dissipation of the energy by the sensor motes can increase significantly the effectiveness of this approach and the lifetime of the hierarchical sensor networks.

Acknowledgement

The work presented in this paper is completed as a partial fulfilment of 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 and Science.

References

[1] G. Smaragdakis, I. Matta, A. Bestavros, SEP: A stable Election Protocol for heterogeneous WSNs, SANPA Workshop, 2004.
[2] Mikhov M., L. Pachemanov, Two-Coordinate Driving System with Wireless Control, Annals of the University of Targu-Jiu, Vol. 10, No. 3, pp. 180-188, Romania, 2010, ISSN 1842-4856.
[3] Georgiev Ts., M. Mikhov, A Sensorless Speed Control System for DC Motor Drives, Scientific Journal of RTU, Vol. 25, No. 4, pp. 155-158, Riga, Latvia, 2009, ISSN 1407-7345
[4] Arnaudov R., G. Stanchev. Pulse Current Controller, Proceedings of ICEST’2002, Nish, Yugoslavia, October 2002.
[5] W. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, Proceedings of HICSS-33, 2000.
[6] P. Zahariev, G. Hristov, T. Iliev, Study on the impact of node density and sink location in WSN, Proceedings of CISSE’09, Bridgeport, USA, 2009.
[7] G. Hristov, P. Zahariev, T. Iliev and M. Iliev, An approach for energy optimization in WSN, 33rd International Convention MiPRO, pp. 203–206, Opatija, Croatia, 2010.
[8] I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, ”A survey on sensor networks”. IEEE CM, 40(8):102–114, 2002.
[9] S. Bandyopadhyay, E. J. Coyle, ”An energy efficient hierarchical clustering algorithm for WSN”, Proceedings INFOCOM’03, April 2003.
[10] A. Chandrakasan, A. Smith, W. Heinzelman, ”An application-specific protocol architecture for wireless microsensor networks”, IEEE T. on WC, pp. 660-670, vol. 1, 2002.
[11] S. Fedor, M. Collier, “An Intra-cluster Architecture to Prolong Wireless Sensor Network Lifetime”, IEEE Conference on Software, Telecommunications and Computer Networks, 2005
[12] S. Bandyopadhyay, E. J. Coyle. ”Minimizing communication costs in hierarchically-clustered networks of wireless sensors”, Computational Networks, 44(1):1–16, January 2004.
[13] C.-Y. Wan, A. T. Campbell, L. Krishnamurthy, ”PSFQ: A reliable transport protocol for WSN”, IEEE ASWN Workshop, pp. 1–11, July 2002.

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

Pavlina Koleva[1], Oleg Asenov[2] and Vladimir Poulkov[3]

***Abstract* – In this paper an Upstream Power Control (UPC) algorithm for Digital Subscriber Lines (DSL) is introduced based on a game theoretic approach of the users in the subscriber loop. A Nash equilibrium power control policy for optimal throughput for the DSL users based on a role game scenario is introduced. The simulation results show that by proper appropriation of the roles of groups of active users in the DSL cable, a maximum of the utility functions for a maximum number of users, could be achieved, keeping the crosstalk under a given limit.**

Keywords – **Digital Subscriber Lines, Game theory, Upstream Power Control.**

I. Introduction

Power control (PC) is an important issue in interference limited multiuser systems, such as the very-high-bit-rate Digital Subscriber Lines (DSL). In these systems, the user's performance depends not only on its own power allocation, but also on the power of all other users and the generated crosstalk interference. The system design involves the estimation of a performance trade-off among the users [1-3].

Power control in DSL systems differs from the PC problem in wireless systems. The DSL transmission environment does not vary over time, but the DSL loops are frequency selective. The DSL systems suffer from a near–far problem which arises when two transmitters located in different distances attempt to communicate with the same Central Office (CO) at the same time. The interference coming from the closer transmitter overwhelms the signal from the transmitter that is farther from the CO. Thus optimal power control schemes need to optimize the total amount of power allocated to each user. For overcoming this problem, algorithms with varying performance and complexity based on power back-off schemes and dynamic spectrum management approaches exist, such as iterative water-filling algorithms, multiuser greedy algorithms, optimal spectrum balancing, etc. In many cases these algorithms are not considering fairness among the users and are in favor of the shorter lines or have a suboptimal performance. In other cases good performance algorithms use high computational complexity [1,3].

Recently game theoretical approaches have appeared studying the power control problem for wireless networks [4-7]. PC algorithms based on the concept of competitive optimality, or strictly competitive games are proposed in [8,9].

In this paper we propose an UPC algorithm based on a role game scenario where different roles to the subscribers are appropriated depending on the distance to the CO, their requested service (transmission speed and bandwidth), crosstalk (activity of other users in the lines), etc.

II. Topology of the DSL plant and environment

DSL technology refers to a family of technologies that provide digital broadband access over the local telephone network. The major problem with this technology is the crosstalk, generated among the lines operating in the same cable bundle. The crosstalk deteriorates the total Signal-to-Noise ratio, thus influencing the overall quality of service performance especially with multimedia services [10].

A very general local loop plant topology is shown in Fig. 1. The CO in a dense populated area can serve thousands of users. The distribution of the subscriber loop is carried in segments of feeder cable. A DSL binder can consist of up to 100 subscriber lines bundled together. Because of their close proximity, the lines create electromagnetic interference into each other, thus causing Near-end crosstalk (NEXT) and Far-end crosstalk (FEXT) noise (Fig. 2).

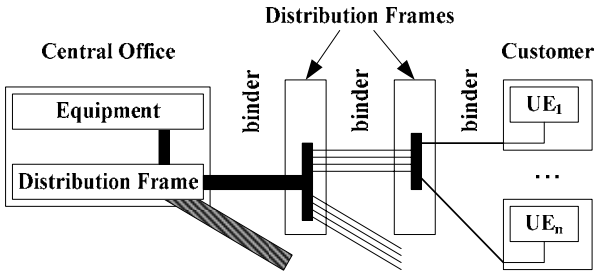


Fig. 1. Topology of a telephone digital local loop plant

The DSL environment is a multiuser environment, because the background noise in the loop is typically small and the system performance is limited by the crosstalk interference. A DSL local loop plant can be modeled as a Gaussian interference channel. In this case this is a multiple transmitter and receiver system, with interference as shown in Fig. 3. The channel from user to user in Fig. 3 is modeled as a frequency-selective channel, whose transfer function in the frequency domain is denoted as $H_{ij}(f)$, where $0 \le f \le Fs$, $Fs = 1/2Ts$ and Ts is the sampling rate.

[1]Pavlina Koleva is with the Faculty of Telecommunications, Technical University of Sofia, Sofia 1756, 8 Kl. Ohridski Blvd., Bulgaria, e-mail: p_koleva@tu-sofia.bg.

[2]Oleg Asenov is with the Faculty of Mathematics and Informatics, St. Cyril and St. Methodius University of Veliko Tarnovo, 5003 Veliko Tarnovo, 2 T. Tarnovski str., Bulgaria, e-mail: olegasenov@abv.bg.

[3]Vladimir Poulkov is with the Faculty of Telecommunications, Technical University of Sofia, Sofia 1756, 8 Kl. Ohridski Blvd., Bulgaria, e-mail: vkp@tu-sofia.bg.

In addition to the crosstalk interference, each receiver also experiences additive background noise, whit Power-Spectral-Density (PSD) σ. The power allocation for each transmitter must satisfy a power constraint:

$$\int_0^{Fs} P_i(f)df \le P_i \quad (1)$$

The achievable data rate for each user is:

$$R_i = \int_0^{Fs} \log_2\left(1+\frac{P_i(f)|H_{ii}(f)|^2}{\Gamma\left(\sigma_i(f)+\sum_{j\neq i}P_i(f)|H_{ji}(f)|^2\right)}\right)df \quad (2)$$

where Γ denotes the SNR-gap.

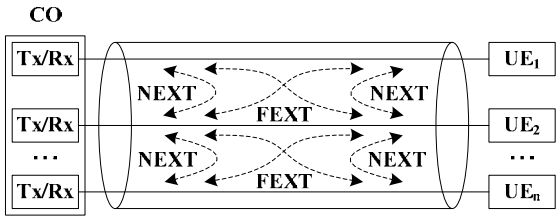


Fig. 2. DSL crosstalk environment

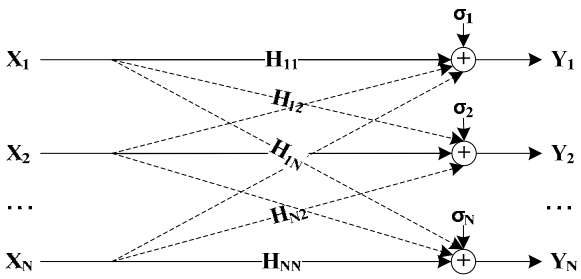


Fig. 3. Gaussian interference network

Objective of the system design is to maximize "jointly" the rates subject to the power constraints (1). For each transmitter, increasing its power at any frequency increases its own data rate, but this also increases its interference into other users. In a local loop topology with different line lengths and the transmitters at the CO transmitting with the same PSD, due to the difference in line attenuation, the FEXT caused by the shorter lines severely affects the upstream transmission and the performance of the long lines. To remedy this power and spectral compatibility problem, the short lines must reduce their upstream PSD. This reduction of the upstream transmit PSD is known as upstream power back-off (UPBO) [11, 12].

Several major UPBO control algorithms have been proposed for VDSL, such as the constant power back-off, the reference length, the multiple reference length, the equalized-FEXT, the reference noise method [12]. All of them require the power or noise spectrum of the short loops to comply with a reference loop or noise. These approaches are relatively simple to implement, because they only require each loop to adjust its power spectrum according to a reference, and do not require any knowledge of the network configuration, activity or QoS of users. With more complex network scenarios more sophisticated power control and allocation methods must be implemented.

III. Role Game Approach for power allocation

In [8] is stated that the majority of crosstalk experienced by a user comes from only a subset of lines within the binder. Such lines are referred as major or dominant crosstalkers and typically correspond to neighboring pairs of a particular line within the binder. In binders whose constituent lines have significantly different lengths near-end users cause significantly more crosstalk than far-end users since the signals of far-end users attenuate before crosstalk coupling occurs. For these reasons large performance gain could be achieved by cancelling crosstalk from dominant crosstalkers.

In [9] the concept of the worst-case interference (WCI) is introduced and the achievable rate of a single so-called "victim" modem in the presence of the WCI from other interfering lines in the same binder group is analyzed. The WCI problem is studied from a game-theoretic viewpoint. The objective is to bound the impact that multiuser interference can have on this victim modem, thereby determining whether service may be guaranteed. Nash equilibrium in this game is interpreted as characterizing a worst-case interference as an optimal response (power-allocation policy) to it.

We use a similar concept to introduce a role game approach for a Nash equilibrium power control policy that could ensure an optimal throughput for the DSL binders, or UEs formed in groups depending on the distance from the CO. The proposed approach is a power control scheme in which groups of DSL lines and their UEs are assigned different roles determined by the distance from the CO, the required service (throughput), crosstalk generated, activity, etc. The different roles are selected so as to achieve the optimum level of user satisfaction. The level of satisfaction is defined by their utility functions. Each role is modeled by the following equation:

$$Role_i(DSL_i) = \left\{ XT\left(\sum_{j=1}^{k} UEA_j\right), UEDis_i, MAX\left[uf_i(T_i, p_i, XTL_n)\right]\right\} \quad (3)$$

where, $XT(\Sigma UEA_j)$ is the crosstalk as a function of users' activity in the bundle (current bundle load), $UEDis_i$ is UE_i distance from the CO, uf_i is the utility function derived by user i, T_i is the throughput, p_i is the upstream power level for UE_i and XTL_n is the crosstalk limit in the bundle.

Based on the UE role model defined in equation (3) the role game approach is applied between the different groups of subscribers. The goal of the role game is to ensure maximum of the utility function for a maximum number of UEs, keeping the crosstalk in the bundle under a given limit. Without loss of generality the value of the Key Performance Indicator (KPI) for a given UE could be represented as the ratio between the throughput that the UE could achieve with his current allowable power level and bandwidth, and the required throughput that is necessary for a given role to have the

necessary quality of performance. In this case maximum of the UE's utility function will be obtained when KPI=1. We assume that the user will be satisfied when he has such a role which ensures his KPI to lie in the interval:

$$1+\Delta \geq KPI \geq 1-\Delta \quad (4)$$

Here Δ is an acceptable deviation in the quality of performance. This deviation is a function of the number of roles. If the KPI is out of this interval for a given UE, the latter will be appropriated a role, away from the maximum of his utility function. So the choice of the value of Δ is a trade-off between the granularity of the roles and performance.

Let's consider a simple game scenario with three groups of users **X**, **A** and **B**, located at a different distance from the CO (Fig.4). They are considered as 3 group players in the game. The number of the UEs in each of the groups is equal and is 20. The goal of the power control mechanism is to ensure the KPIs of all of the UEs in the groups to be in the interval defined above and not to exceed a given limit of the crosstalk. Let's assume that an UE depending on his role generates some corresponding crosstalk to the other DSL lines in the bundle when his KPI=1, defined as $RoleXtalk_k$. The total crosstalk that all the UEs in a group could generate at a given time when they have achieved maximum of their utility functions (KPI=1) is defined through a relative parameter called "*additional XTalk*", dependent on the number of UE and their roles. The PC mechanism is performed in several steps.

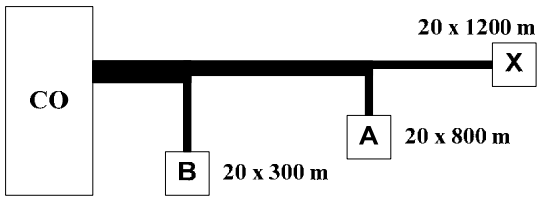
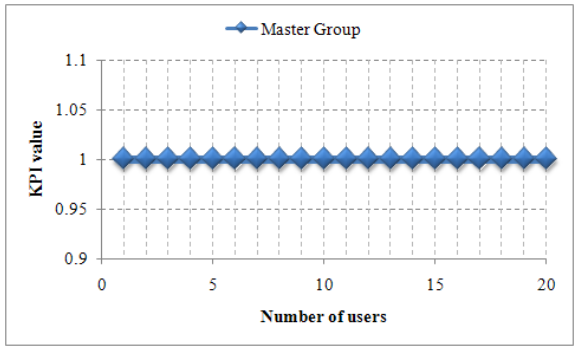


Fig. 4. Game scenario with three groups of users

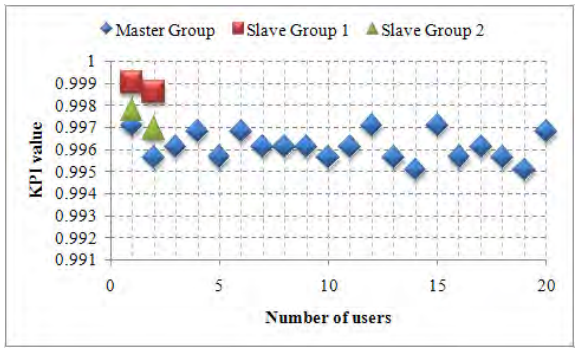
In the beginning, this parameter is estimated for each group, i.e. $AddXTalk(Slot_i,Group_j)=SUM(RoleXTalk_k)$. The group which generates the maximum additional noise is appointed as master group and its estimated $AddXTalk(Slot_i,Group_j)$ is considered as reference. If this reference is under the admissible crosstalk limit all the UEs in the master group will receive the required upstream power. The PC mechanism is applied in such a way, that the UEs in this group will receive the required roles ensuring their KPI to be equal to 1. The other two groups will be considered as slave groups. This means that the number and activity of the UEs and the rank of their corresponding roles in the group contributes the corresponding group to become master.

In the next step, the PC algorithm sets the upstream power for the UEs located in the slave groups in such a way that their KPIs lie within the interval (4). The upstream power of each of the UEs located in these groups is determined as follows. The necessary power to ensure the required throughput for the requested role from the UE in the case of no crosstalk is estimated. Then the PC mechanism allows increase of the power of the UE to a level, which is necessary to compensate the "*additional XTalk*" introduced by the master group. The resulting KPI from the slave group is close but always less than 1, as the "*additional XTalk*" from the UEs in the neighboring slave groups is not compensated. Further, the $AddXTalk(Slot_i,Group_j)$ generated from each group is calculated. If the $AddXTalk(Slot_i,Group_j)$ in one of the slave groups becomes higher than the one in the master, this slave group is appointed master. This means that the number and activity of the UEs and the rank of their corresponding roles in one of the slave groups has become higher than the ones in the current master and a new master group is appointed.

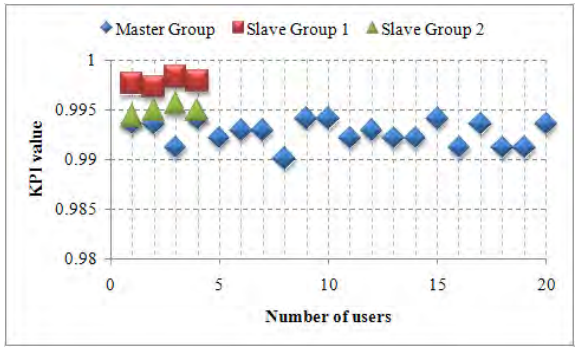
a)



b)



c)



d)

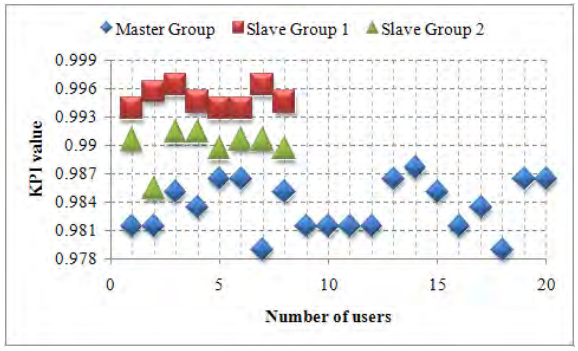


Fig. 5. KPI calculations for the role game algorithm

With the increase of the number and the rank of the requested roles in the slave groups, a tendency for the decrease of the users KPIs in the master group is expected. If the number of the users and/or the rank of their roles are relative low to those in the master group, with high probability the resulting deviation of the KPI of the UEs in the master group will lie within the interval (4). Then the power control mechanism in the CO will not take any action to change the upstream power of the UEs. This could be seen from the results from the simulations of the algorithm shown in Fig. 5. Let's accept that Δ is 0.1, there are 20 active UEs in group **X**, requiring randomly different ranks of roles, and no UEs are active in the other two groups. Thus the parameter $AddXTalk(Slot_i,Group_j)$ is maximum for this group and the latter is appointed as master. All of its UEs receive KPI=1, as shown in Fig. 5a. Following, the KPI of the users is calculated in case when in the slave group a number of 2, 4 and 8 UEs become active – Fig. 5b, c and d. It could be seen that the deviation of the KPIs of the users in the master group are within the interval (4).

In the limited case, of a small difference of the parameter $AddXTalk(Slot_i,Group_j)$ between the current master group and one of the slave groups, it is probable that part of the UE's KPI become less than 0.9 if additional UEs become active.

In such a case the reference crosstalk level of the master group is memorized and the power level of the UEs with least KPI is increased selectively up to the value KPI=1. A new reference $AddXTalk(Slot_i,Group_j)$ in the master group is obtained. The necessary power compensation in the slave groups is calculated for the "*additional crosstalk*" parameter from the master group based on the information from the previous slot, without taking into consideration the increase of the power of some of the UEs in the master group. This will lead to a decrease in the resulting KPI of the UEs in the slave groups. When the KPI of an UE from a slave group becomes less than 0.9, the power control scheme calculates the new reference $AddXTalk(Slot_i,Group_j)$ level and recalculates the individual power of their UEs and respective KPIs.

When changing the reference level of the parameter $AddXTalk(Slot_i,Group_j)$ it is obligatory to check if the limit of the allowable crosstalk level is reached. If so, the system changes into mode "*role decrementing*". This is a case when the master group cannot ensure maximum of the utility functions of all of his users and the PC mechanism will start lowering the rank of the roles of some of the UEs, thus giving them less throughput to keep the crosstalk below the limit. The criteria for role decrementing (applying a role of lower KPI) for part of the UEs, could be different. The system goes out of the mode "*role decrementing*" in two cases: change of the reference group or if the crosstalk falls below 95% of the allowable limit.

IV. Conclusion

This paper introduces an approach for PC for DSL based on defining roles of the subscribers within a subscriber network. The results show that optimal average throughput could be achieved, keeping the crosstalk interference under a given limit. On the other hand applying such a dynamic role appropriation algorithm fairness concerning the upstream power allocation among the subscribers in a DSL environment is achieved.

Acknowledgement

This work was supported in part by the Grant DDVU 02/13/17.12.2010 "Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization" of the Bulgarian Science Fund and by the Technical University of Sofia under Grant 121PD0068-07/2012 "Cognitive Approaches for Power and Interference control in New Generation Telecommunication Networks"

References

[1] T. Starr, J. M. Cioffi, and P. J. Silverman. *Understanding Digital Subscriber Line Technology*. Englewood Cliffs, NJ: Prentice Hall, 1999, ch. 3.

[2] P. Tsiaflakis, M. Diehl, M. Moonen. Distributed Spectrum Management Algorithms for Multiuser DSL Networks. IEEE Transactions on Signal Processing, 56(10):4825–4843, Oct 2008.

[3] K. S. Jacobsen, "Methods of upstream power backoff on very high-speed digital subscriber lines," IEEE Communication Magazine, pp. 210–216, Mar. 2001.

[4] K. B. Song, S. T. Chung, G. Ginis, and J. M. Cioffi, "Dynamic spectrum management for next-generation DSL systems,". IEEE Communications Magazine, vol. 40, no. 10, pp. 101–109, 2002.

[5] K. J. Kerpez, D. L. Waring, S. Galli, J. Dixon, and P. Madon, "Advanced DSL management," IEEE Communications Magazine, vol. 41, no. 9, pp. 116–123, 2003.

[6] T. Starr, M. Sorbara, J. M. Cioffi, and P. J. Silverman, DSL Advances, Prentice-Hall PTR, Upper Saddle River, NJ, USA, 2003.

[7] S. T. Chung, S. J. Kim, J. Lee, and J. M. Cioffi, "A game theoretic approach to power allocation in frequency-selective Gaussian interference channels," in Proceedings of IEEE International Symposium on Information Theory (ISIT '03), pp.316–316, Pacifico Yokohama, Kanagawa, Japan, June–July 2003.

[8] G. Ginis and J. M. Cioffi. Vectored transmission for digital subscriber line systems. IEEE J. Sel. Area. Comm., 20(5):1085–1104, Jun. 2002.

[9] M. Brady and J. Cioffi. The Worst-Case Interference in DSL Systems Employing Dynamic Spectrum Management Hindawi Publishing Corporation EURASIP Journal on Applied Signal Processing Volume 2006, Article ID 78524, Pages 1–11 DOI 10.1155/ASP/2006/78524.

[10] E. Pencheva, I. Atanasov, D. Marinska, "Cross layer design of Application-level Resource Management interfaces", in Proceedings of Second International Workshop on Cross Layer Design, pp. 1-5, Palma de Mallorca, 11-12 June 2009.

[11] W. Yu, G. Ginis, and J. Cioffi. Distributed Multiuser Power Control for Digital Subscriber Lines. IEEE Journal on Selected Areas on Communications, Vol. 20, No. 5, June 2002.

[12] K. S. Jacobsen, "Methods of upstream power backoff on very high-speed digital subscriber lines," IEEE Communication Magazine, pp. 210–216, Mar. 2001.

A possibility for Edge detection using LabVIEW graphical programming environment

Liljana Docheva[1]

Abstract – **In this paper implementation of different edge detectors using LabVIEW graphical programming environment is given. When a system is implemented with LabVIEW, one can produce highly parallel data flow. This will lead to simplification and to speed increasing of the whole system.**

Keywords – **Edge detection, LabVIEW.**

I. INTRODUCTION

Edge detection is a tool, used in image processing, computer vision and machine vision. Feature detection and feature extraction can't be performed without this tool. From the edges of an image can be extracted very important features and thus the amount of data that will be processed is reduced and the analysis of the image is simplified. Edge detection can be used from higher-level computer vision algorithms (object recognition, motion etc.).

II. REVIEW OF SOME OF THE EDGE DETECTORS

Edges are area of the image with significant changes of the intensity, therefore it mast to decide whether a pixel with particular intensity is edge point or no. There are different detectors that can be applied.

A. The Roberts edge operator

The Roberts edge operator is a gradient based operator.

Approximation of the gradient by the finite differences can be written as (Eq. 1 and Eq. 2) [3]:

$$\frac{\partial f}{\partial x} = f(j-1,k) - f(j,k-1) \tag{1}$$

$$\frac{\partial f}{\partial y} = f(j,k) - f(j-1,k-1) \tag{2}$$

This equations corresponds to the following masks (Eq. 3 and Eq. 4):

$$H_1 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{3}$$

$$H_2 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4}$$

[1] Liljana Docheva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: docheva@tu-sofia.bg.

It can be made the following assumption (Eq. 5 and Eq. 6):

$$G_1(j,k) = \frac{\partial f}{\partial x} \tag{5}$$

$$G_2(j,k) = \frac{\partial f}{\partial y} \tag{6}$$

The equation for the magnitude is.

$$A(j,k) = \left|\sqrt{G_1^{\,2}(j,k) + G_2^{\,2}(j,k)}\right| \tag{7}$$

Edge orientation can be calculated by equation 8.

$$\alpha = arctg\left(\frac{G_2}{G_1}\right) + \frac{\pi}{4} \tag{8}$$

B. The Prewitt edge operator

For edge detection with the Prewitt operator the following masks are used (Eq. 9 and Eq. 10)

$$H_1 = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \tag{9}$$

$$H_2 = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{10}$$

The magnitude can be calqulated by Eq. 7. Equation 11 is calqulation of the Edge orientation.

$$\alpha = arctg\left(\frac{G_2}{G_1}\right) \tag{11}$$

A. The Sobel edge operator

For edge detection with the Sobel operator the following masks are used (Eq. 12 and Eq. 13)

$$H_1 = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{12}$$

$$H_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{13}$$

The magnitude can be calqulated by Eq. 7 and the edge orientation by Eq. 11.

A. The Laplasian

This is edge detector that uses the second derivative (Eq. 14) [2,3]. In this case an edge point can be detected by finding the zero-crossing of the second derivative.

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (14)$$

When equation 15 for the Laplasian or equation 7 for the others three edge operators are computed it can be made decision (Eq. 16) whether a pixel belong to the edge or not.

$$A(j,k) = f(j,k) ** H \quad (15)$$

$$A(j,k) \geq \Delta \quad (16)$$

The Laplasian can be implemented using the following mask:

$$H = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (18)$$

The implementation of these four edge detectors using LabVIEW graphical programming environment without any additional tools is given in the paper.

III. Implementation of Edge Detectors Using LabVIEW Graphical Programming Environment

LabVIEW is a graphical programming environment that uses icons to create applications [1]. One is flexible and easy to solve different kind of problems. Highly parallel data flow that one can produce, make this programming environment suitable for image processing. In this paper is shown a way to implement edge detection using LabVIEW without any additional tools.

Implementation of Gradient based operator like Roberts, Prewitt and Sobel is shown on fig.1. There are differences in the masks (eq. 9 and 10 and eq. 12 and 13 accordingly) and in the determining the direction of the edge (eq. 11).

The first step for the edge detection realization is to read the image file. Then the cluster of image data mast to converts into a 2D array. This allows to apply the masks and compute the magnitude ***A*** (Eq. 7). Next it can be determined the direction of the edge (α). The last step is to make decision whether a pixel belong to the edge or not.

Figure 2 presents the masks for the Roberts edge operator.

Implementation of the Prewitt and Sobel edge operators is similar. The differences are in the masks (Fig. 3 and Fig. 4 accordingly) and in the determining the direction of the edge (eq. 11).

As it is mention above the Laplasian uses the second derivative. An implementation of this edge operator with LabVIEW has small differences from the one of Gradient

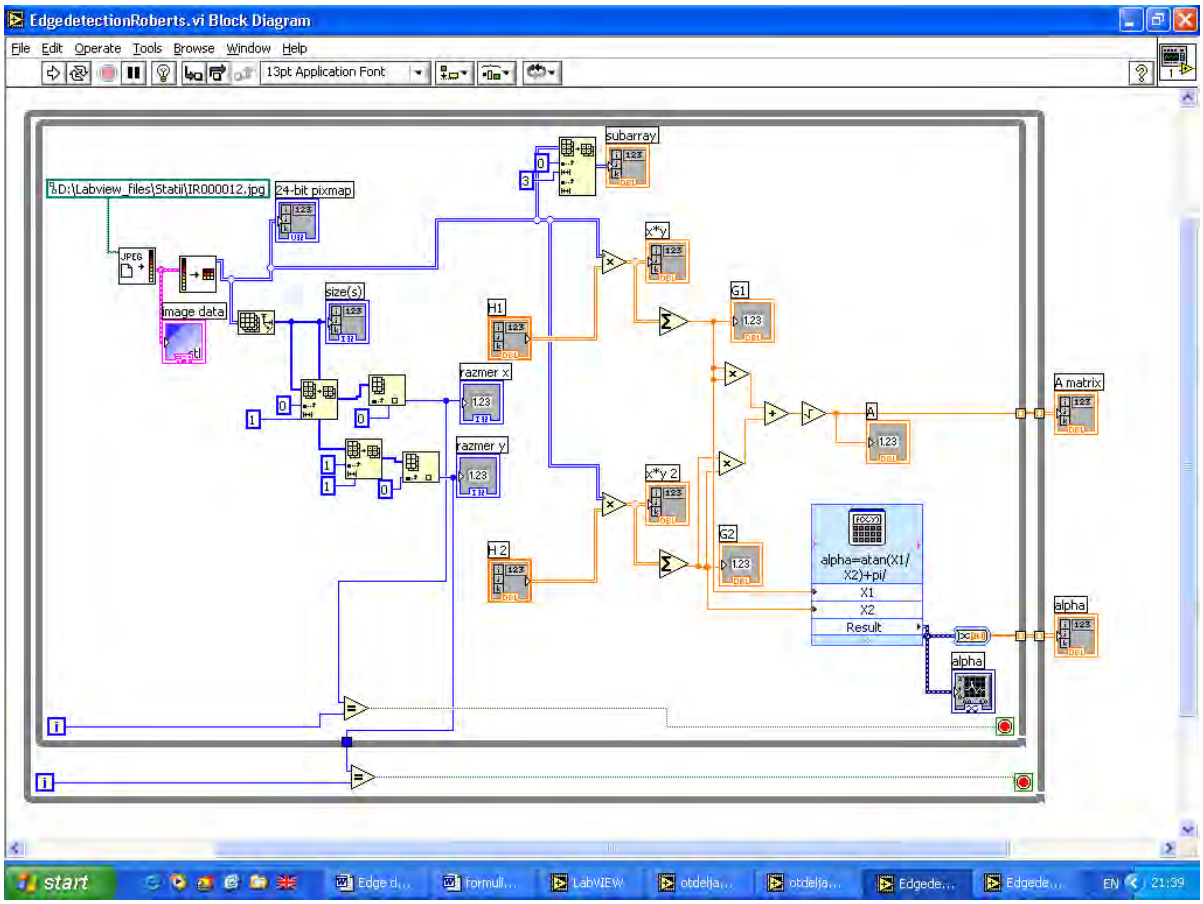


Fig. 1. Implementation of Gradient based operator with LabVIEW.

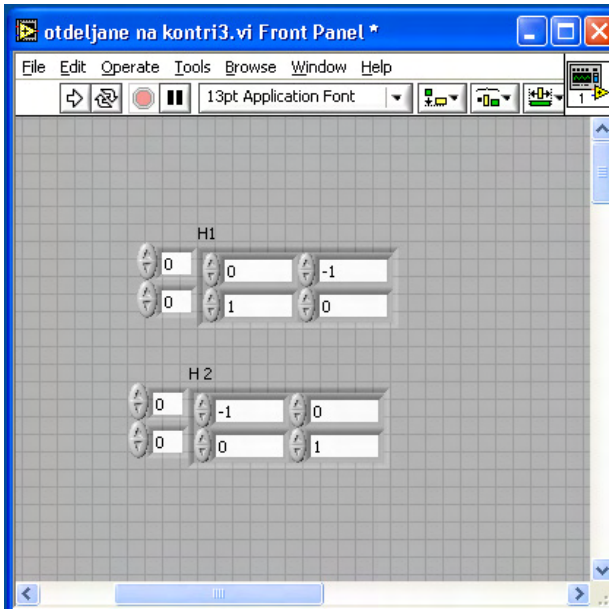


Fig. 2. The Roberts edge operator masks on LabVIEW Front Panel.

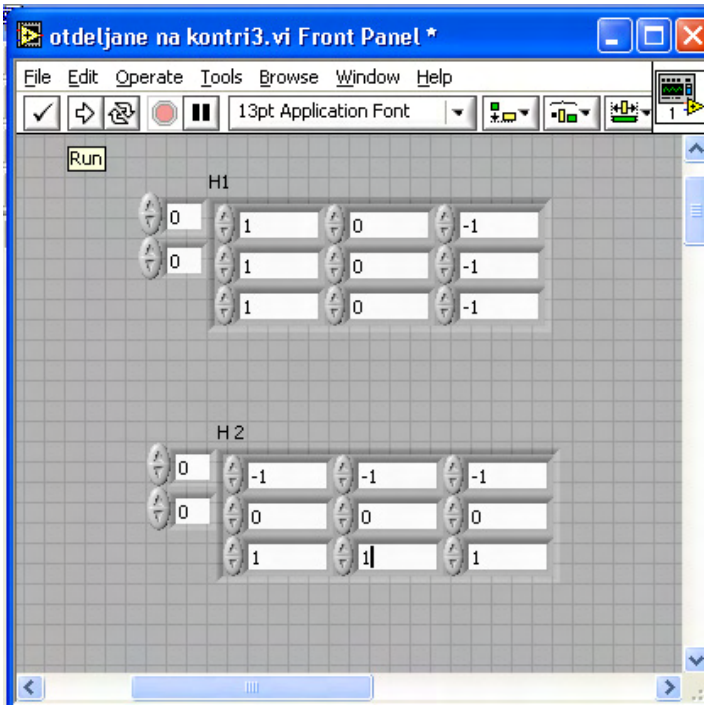


Fig. 3. The Prewitt edge operator masks on LabVIEW Front Panel.

based operator. The Laplasian needs from one mask (Fig. 5) only that simplifies the implementation, but one doesn't provide information about edge direction. This reflects to the LabVIEW block diagram of Laplasian (Fig. 6). It can be seen that one isn't need from computing of magnitude.

It is seen from the LabVIEW block diagram of Laplasian that the decision can be made after applying the mask directly.

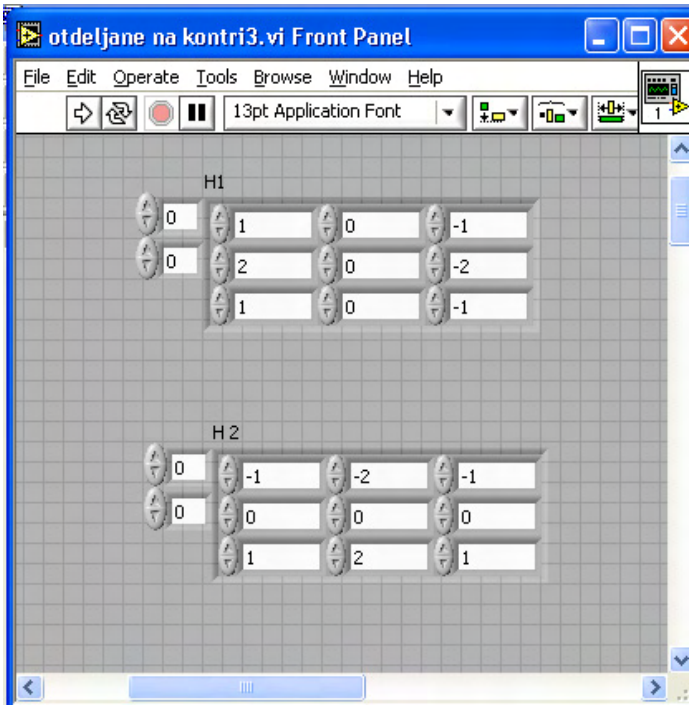


Fig. 4. The Sobel edge operator masks on LabVIEW Front Panel.

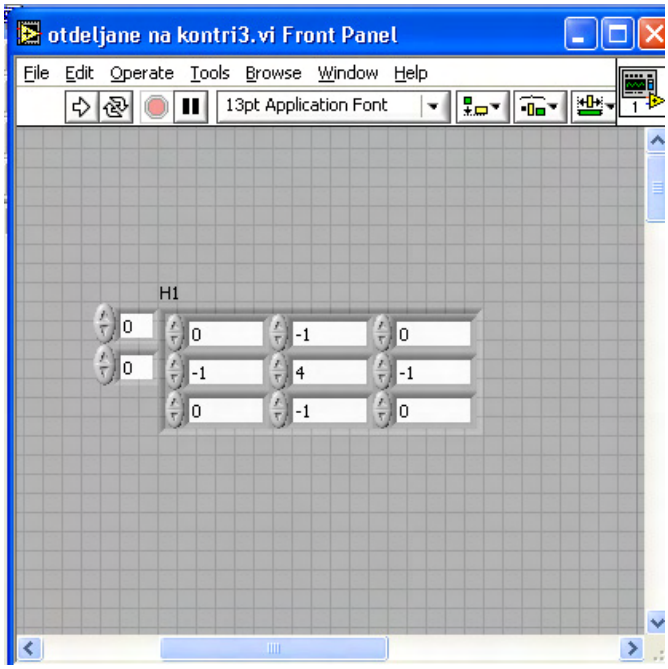


Fig. 5. The Laplasian edge operator masks on LabVIEW Front Panel.

IV. Conclusion

In this paper is shown a way to implement four edge detectors using LabVIEW graphical programming environment without any additional tools. The differences in the implementation can be seen. The solution is simple and flexible.

References

[1] National Instruments "LabVIEW7 Express. User Manual", National Instruments Corporation 2003.

[2] E. Nadernejad, "Edge Detection Techniques: Evaluations and Comparisons", Applied Mathematical Sciences, vol.2, pp.1507-1520, no.31, 2008.

[3] http://www.cse.uunr.edu/~bebis/CS791E/Notes/EdgeDetection.pdf

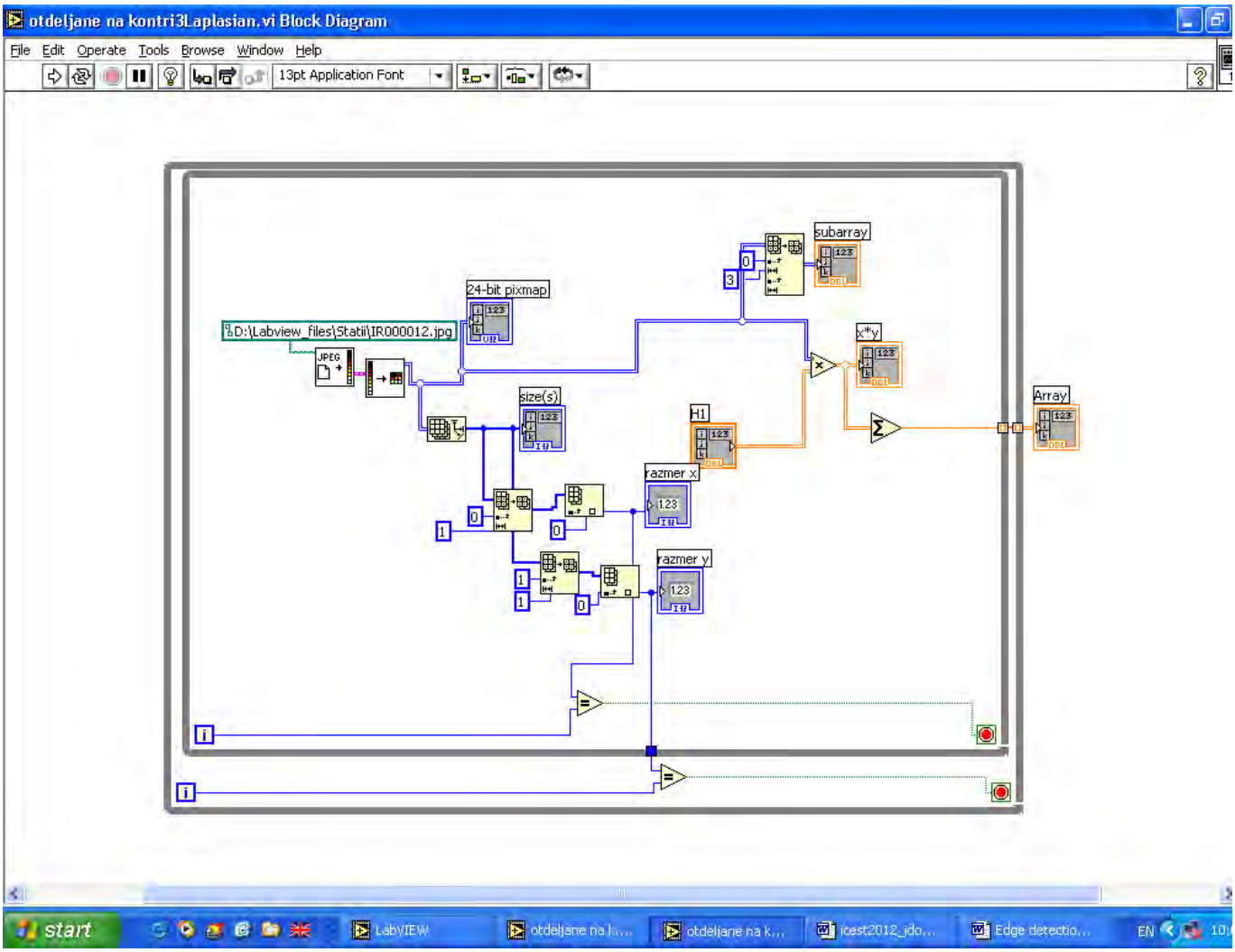


Fig. 6. Implementation of Laplasian based operator with LabVIEW.

Implementation of Analog Neural Networks with Labview

Liljana Docheva [1] and Aleksander Bekiarski [2]

Abstract – **The ability of analog neural networks to process different types of information is well known, but for the real application such as image processing or recognition it is necessary to use modelling and analyzing of the methods chosen for a concrete image processing or recognition. The modelling and analysis included also the specific parameters of the electronic elements of the analog neural networks.**

In this article it is proposed to model and analyze a multi-layered analog neural network with LABVIEW, applying Back Propagation algorithm in learning process. These types of neural network are capable to perform just about any linear or non-linear computation, and can approximate any reasonable function arbitrarily well. LABVIEW graphical programming environment and capability in building analog artificial neural networks are discussed.

The result of the modelling and simulations of the proposed multi-layered analog neural network are shown and in the conclusion are made some remarks of the advantages and the proposition for the future investigations in this direction.

Keywords – Analog neural networks, LABVIEW.

I. INTRODUCTION

Analog neural networks have its wide application cause their high speed, low power consumption and compact implementation. But variation in the size of discrete transistors and the local mobility will cause random parameter variation. An increase in the precision of any component will lead to increase of its area.

Limited accuracy and nonlinear behavior, typical of analog neural networks, don't reduce their application because utilization of appropriate training algorithm (back propagation for example) decreases their influence to a certain degree.

Simulations made in LABVIEW will help to design and test analog neural network with aim to investigate influence of some analog neural network parameters onto its recognition ability.

LABVIEW is graphical programming software, used usually for data acquisition and control. In some articles have been successfully used LABVIEW for building neural networks [1,2] because one has the ability to develop data flow that are highly parallel in structure. Therefore LABVIEW seems to be a very effective approach for building and investigating neural networks.

In addition is very easy during building process to make that the parameters of real components to take part in neural networks. On this way the components of an analog neural network will be close to real ones.

In this article multi-layer feed-forward analog artificial neural network is investigated. The training algorithm applied to this network is back propagation. The aim is to discuss the LABVIEW graphical programming environment, its features and capability in building analog artificial neural networks.

[1]Liljana Docheva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: docheva@tu-sofia.bg.

[2]Aleksander Bekiarski is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria,aabbv@tu-sofia.b.

II. ANALOG NEURAL NETWORK

Many investigations in the implementation of analog neural networks field are known [3, 4, 5, 6]. In this article a VLSI implementation of an analog neural network, depicted in [5] is chosen for building with LABVIEW. The synapse chip consists of a number of MOS resistive circuit multiplier. On the base of the neuron and synapse equations described in [5], equations about analog parameter variation over analog neural network behavior have been worked out [7,8]. These equations will take place in LABVIEW implementation discussed in this article. On this way we have investigated parameter variation influence over behavior of a multi-layer feed-forward analog artificial neural network implemented with LABVIEW. The training algorithm applied to this network is back propagation. The aim isn't to decide some complex problem. We chose a simple one (AND problem) in the beginning and more difficult on the later stage of our research. On this way it will be shown how parameter variation and complexity of the task affects over behavior of an analog neural network.

III. APPLICATION OF LABVIEW FOR AN ANALOG NEURAL NETWORK IMPLEMENTATION

A. Block scheme

Graphical programming environment LABVIEW allows construction and investigations of complex systems analysis including neural networks. There are different ways to realize neural network with LABVIEW. We have clung to architecture of the back propagation network. This implementation with LABVIEW is easy to explain with the block diagram depicted on Fig. 1. Feed forward signal is accomplished from 1 to 5 numbered blocks. The blocks with numbers from 6 to 9 serve as purpose to back propagation error signal. Function of every one of them is explained bellow.

Block 1 is a data base. One includes input patterns that are applied to neural network. In our case the input pattern is a binary vector.

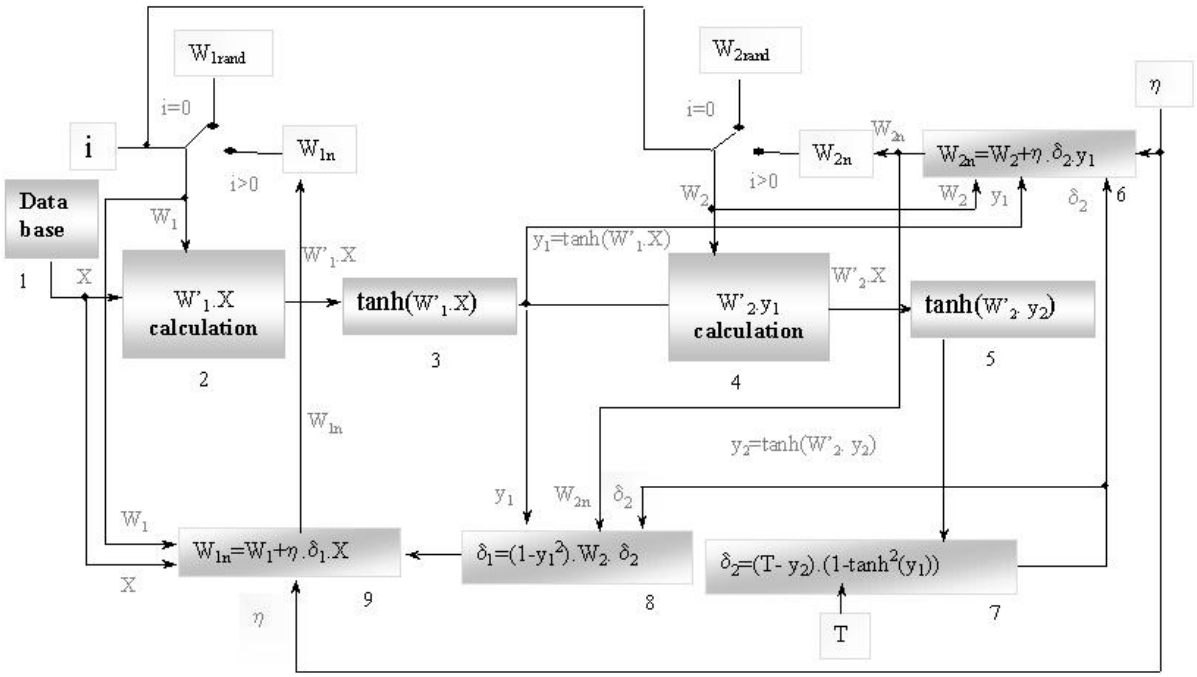


Fig. 1. The analog neural network LABVIEW implementation block diagram

Blocks 2 and 4: The aim of these blocks is calculating the weighted sum for the first layer and for the second layer correspondingly. First at all it mast to decide whether input vector will be multiplied by weight matrix of random numbers (first initialization - $\mathbf{W_{1rand}}$ or $\mathbf{W_{2rand}}$), or matrix of updated weights ($\mathbf{W_{1n}}$ or $\mathbf{W_{2n}}$). For this aim we use the number of iteration **i**. When **i** is equal to **0** the weight mast to initialize for first time with random numbers. For the all next iterations (**i>0**) input vector will be multiplied by the updated weight matrix.

The weighted sum takes part in output voltage of neuron calculating (blocks 3 and 5). Ones calculate activation function (tanh(.) in our case). On the outputs of these blocks are obtained $\mathbf{y_1}$ and $\mathbf{y_2}$. In this paper is considered analog neural network therefore we will denote **w** as $\mathbf{u_w}$ and **y** as $\mathbf{u_y}$ (Eq. 1). For the output layer it can be written [8]:

$$u_{yk}^{2}(t)=\left(\chi_k^2+\Delta\chi_k^2\right)\tanh\left(\left(\xi_k^2+\Delta\xi_k^2\right)\sum_{j=1}^{N}U_{w_{kj}}u_{yj}^{1}(t)\right) \quad (1)$$

where

u_{yk}^{l} is the neuron output voltage;

k – number of neuron,

l - neural network layer number,

χ and ξ - parameters of analog components constructing neuron.

Δ- symbol of variable variation.

After signal is fed forward the error back propagation follows. The firs step is to calculate the equivalent error of the k[th] neuron of the 2[th] layer δ^2_k (Eq. 2) and is accomplished of block 7:

$$\delta_k^2(t)=\left[d_k-\left(\chi_k^2+\Delta\chi_k^2\right)\tanh\left(\left(\xi_k^2+\Delta\xi_k^2\right)\sum_{j=1}^{N}U_{w_{kj}}u_{yj}^{1}(t)\right)\right]. \quad (2)$$
$$.\left(\chi_k^2+\Delta\chi_k^2\right)\left(\xi_k^2+\Delta\xi_k^2\right)\left(1-\tanh^2\left(\left(\xi_k^2+\Delta\xi_k^2\right)\sum_{j=1}^{N}U_{w_{kj}}u_{yj}^{1}(t)\right)\right).$$

Then in block 6 the weight update for the 2th layer is calculated (Eq. 3).

$$U_{wyk}^{2}(t+1)=U_{wyk}^{2}(t)+(\eta.\delta_k^2(t).u_{yj}^{1}(t)) \quad (3)$$

Calculating the equivalent error of the 1[th] layer (block 8) is computed (Eq. 4): on the base of equivalent error of the 2[th] layer:

$$\delta_j^1(t)=\left(\chi_j^1+\Delta\chi_j^1\right)\left(\xi_j^1+\Delta\xi_j^1\right)\left(1-\tanh^2\left(\left(\xi_j^1+\Delta\xi_j^1\right)\sum_{j=1}^{M}U_{w_{ij}}(t)u_{yz}(t)\right)\right)\sum_{j=1}^{K}U_{w_{kj}}(t)\delta_k^2(t), \quad (4)$$

The aim of block 9 is the weight update for the 1th layer (Eq. 5).

$$U_{wyk}^{1}(t+1)=U_{wyk}^{1}(t)+(\eta.\delta_k^1(t).u_{yz}(t)) \quad (5)$$

C. Implementation of an analog neural network with graphical programming software LABVIEW

It isn't difficult to implement the bocks mentioned above with *g*raphical programming software LABVIEW. Before blocks 2 and 4 implementation, decision whether input vector will be multiplied by weight matrix of random numbers mast be made. This is presented on Fig. 2 for first layer. The design for second layer is similar. When the number of iteration **i** is equal to **0** the weight mast to initialize for first time with random numbers. For the all next iterations (**i>0**) input vector will be multiplied by the updated weight matrix.

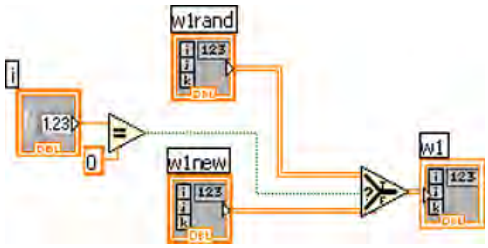


Fig. 2. The choice of weight matrix for the weighted sum of first layer.

Implementation of block 2 is depicted on Fig. 3.

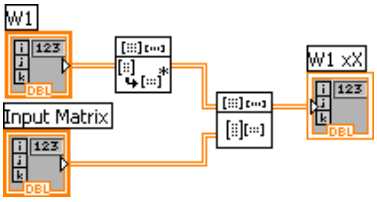


Fig. 3. Implementation of block 2.

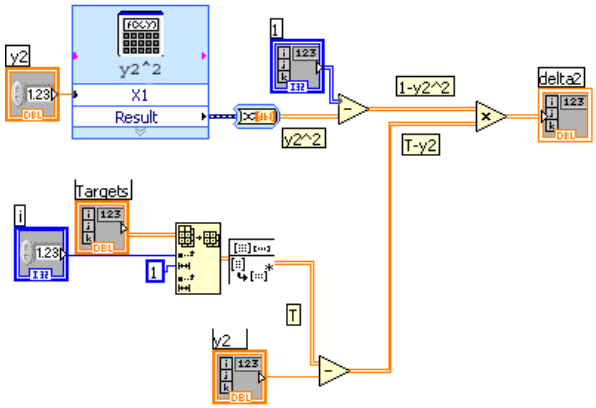


Fig. 4. Implementation of the equivalent error calculation for the second layer

For this aim are used two virtual instruments: **Transposes Input Matrix** and **AxB**. After weight matrix transposing the multiplication of input vector (pattern) and weight matrix is performed with virtual instrument **AxB**. The block 4 is

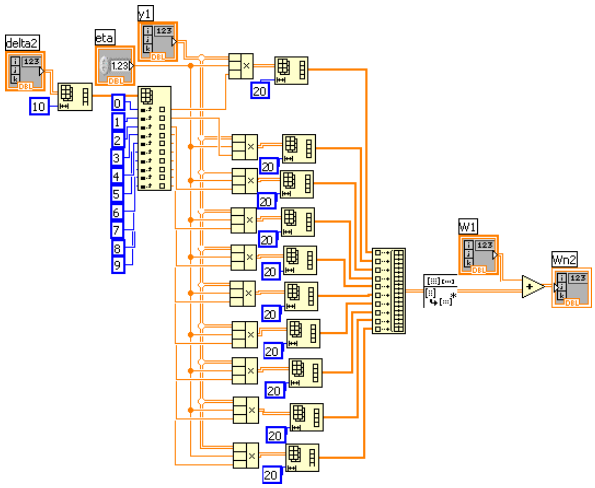


Fig. 5. Weight update for the 2th layer.

implemented on same way.

As it can be seen from Eq. 1 in calculation for output neuron voltage parameters of the real elements of analog neural network take part. The activation function (blocks 3 and 5) is realized by **Formula** virtual instrument.

Figure 4 depicts a block 7 implementation. One is done following Eq. 2.

The block 6 - the weight update for the 2th layer (Eq. 3) is given in Fig 5. After delta for second layer is calculated new weights can be computed. To obtain this operation mast to be used the follow virtual instruments: **Reshape Array** - Changes the dimension of an array, **Index Array** - Returns

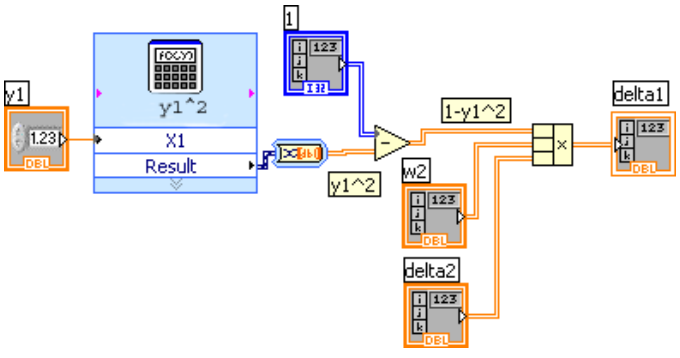


Fig. 6. Implementation of the equivalent error calculation for the first layer

the element of 1-dimension array, **Build Array** - Concatenates multiple vectors to an 2-dimensional array.

After operations of Eq. 3 are accomplished in an output of he block the updated weights are obtained.

Figure 6 depicts a block 8 implementation that follows Eq. 4 expression. Virtual instrument calculate square of neuron output but in addition voltage parameters of the real elements of analog neural network are included in accordance with Eq. 4.

The block 9 - the weight update for the 1th layer (Eq. 5) - is implemented on similar way to the block 6 implementation.

IV. Results

In order to be sure that analog neural network implemented with *g*raphical programming software LABVIEW works correctly we chose a simple task - AND problem. Fig.7 shows

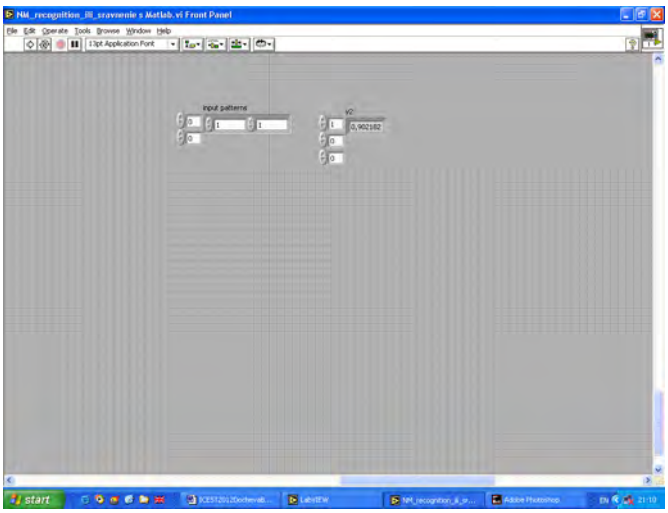


Fig. 7. LABVIEW front panel for one of the applied input patterns

LABVIEW front panel for one of the applied input patterns.

The results for all patterns are given in Table 1. Ones present that network work correctly despite of influence of the analog neural network parameters onto its recognition ability.

TABLE I
THE RESULT OF ANALOG NEURAL NETWORK LEARNING FOR AND PROBLEM

X	Y2
11	0.902
10	0.088
01	0.093
00	0.086

V. Conclusion

In this paper simulations made in LABVIEW are depicted. LABVIEW graphical programming environment, its features and capability in building analog artificial neural networks are discussed. The results present that network work correctly despite of influence of the analog neural network parameters onto its recognition ability.

On the later stage of our research we will examine analog neural network behaviour realized with LABVIEW when a task is more complex

References

[1] J. Fernandez de Canete, S. Gonzalez-Perez, P. del Saz-Orozco, “Artificial Neural Networks for Identification and Control of a Lab-Scale Distillation Column using LABVIEW”, World Academy of Science, Engineering and Technology, 2008.

[2] M.A. Panait, T. Tudorache, " A Simple Neural Network Solar Tracker for Optimizing Conversion Efficiency in Off-Grid Solar Generators", International conference on renewable energies and power quality (ICREPQ'08).2008.

[3] Cyril Prasanna Raj P, S.L. Pinjare, "Design and Analog VLSI Implementation of Neural Network Architecture for Signal Processing", European Journal of Scientific Research , Vol.27 No.2 (2009), pp.199-216,2009.

[4] G. Cauwenhberghs, "An Analog VlSI Recurrent Neural Network Learning a Continuous-Time Trajectory", [1]IEEE Transactions on Neural Networks, vol. 7, N:2, March 1996.

[5] [6] T. Lehman, " Hardware Learning in Analog VLSI Neural Networks ", Ph.D. thesis, Technical University of Denmark, 1994.

[6] S. Draghici, " Neural Networks in analog hardware-design and implementation issues ", Int. J. of Neural Systems, 2000, vol.10, no. 1,pp. 19-42.

[7] L. Docheva, A. Bekiarski, I. Dochev, "Analysis of Analog Neural Network Model with CMOS Multipliers", Radioendineering, vol. 16, N:3, september 2007.

[8] L. Docheva, A. Bekiarski, "Investigationof analog Neural Network used for number recognition in images", International Journal of Neural Networks and applications, 2(1) January-June 2009, pp. 15-18.

Noise- Resistance Performance Estimation of a Chaos Shift Keying Signals

Galina Cherneva[1] , Georgi Pavlov[2] and Elena Dimkina[3]

Abstract – **In this paper the performance evaluation of error probability for a chaos shift keying coherent system over an Additive White Gaussian Noise (AWGN) channel is presented.**

Keywords – **Chaos communication, Chaos-shift keying, Error probability**

I. INTRODUCTION

The chaotic signal is a relatively new field in communication systems. Motivation derives from the advantages offered by chaotic signals, such as spread spectrum, robustness in multipath environments and resistance to jamming. Chaotic signals are non-periodic, broadband, and difficult to predict and to reconstruct; they may be generated by mathematical map functions, electronic circuits or laser optics. Their properties coincide with requirements for signals used in communication systems, in particular for secure communication systems.

Chaos Shift Keying (CSK) [1,2] is one of the encoding/modulation methods, proposed in the literature to send digital messages. The CSK method employs the advantages offered by the inherent phase synchronization [2] given by the drive response principle and well understood correlation methods [3].

The receiver in CSK communication system can use coherent or non-coherent detection techniques. In coherent detection, the receiver is required to reproduce the same chaotic signals sent by the transmitter, often through a chaos synchronization process which is unfortunately a fragile process [3]. In non-coherent detection of CSK, however, the receiver does not have to reproduce the chaotic signals. Rather, it makes use of some distinguishable property of the chaotic signals to determine the identity of the digital symbol being transmitted. Different attractors may differ in variance, meaning of the absolute value, dynamic range, and many other statistical properties [3]. The optimal decision level of the threshold detector, however, will depend on the signal-to-noise ratio of the received signal.

One of the main characteristics determining the effectiveness of a radio communication system is the stability against disturbances [3]. It is characterized with the dependency of the fidelity of received communications on the line energy parameters, algorithms used to transmit information and statistical characteristics of disturbances [3]. With discrete systems of connections, the error probability of distinguishing signals is used for fidelity assessment. [3].

The purpose in this paper, is to present the error probability of a coherent CSK digital system under the influence of additive white Gaussian noise (AWGN), assuming ideal synchronization at the receiver. The solution for the error probability has been derived, in terms of the signal-to-noise power ratio.

II. REVIEW OF CSK APPROACH

The idea of CSK Approach is to encode digital symbols with chaotic basis signals. A block diagram of the communication system with CSK is shown in Fig. 1. The transmitter dynamics is dissipative and chaotic and the transmitter state trajectory converges to a strange attractor. A message is transmitted by changing one or more parameters of the transmitter dynamics which results in a change of the attractor dynamics. At the receiver the message is decoded by estimating to which message the received chaotic attractor corresponds.

The transmitter consists of two chaos generators 1 and 2, producing signals $x_1(t)$ and $x_2(t)$, respectively.

Assume that a chaotic signal is generated by the map

$$x[n+1]=f(x[n]), \quad (1)$$

where

$$x[n]=(x_1[n], x_2[n], \ldots\ldots, x_m[n]) \quad (2)$$

is the state, $\boldsymbol{f} = (f_1, f_2, \ldots\ldots f_m)$ is the discrete functional transformation (maps) the state $\mathbf{x}[n]$ to the next state $\mathbf{x}[n+1]$.

With two different initial conditions, we can generate two sets of chaotic sequences which can be used to represent two binary symbols. Let $\{x_{n1}\}$ and $\{x_{n2}\}$ be the two chaotic sequences representing "0" and "1" respectively.

The outputs of the chaotic signal generators, denoted by $x_1(t)$ and $x_2(t)$, are given by

$$x_1(t) = \sum_{n=0}^{\infty} x_{n1} r(t - nT_r) \quad (3)$$

and

[1]Galina Cherneva is with the Faculty of Telecommunications and Electrical Equipment in Transport at T. Kableshkov University of Transport Sofia, 158 Geo Milev Blvd, Sofia 1574, Bulgaria, E-mail: cherneva@vtu.bg.

[2]Georgi Pavlov is with the Faculty of Telecommunications and Electrical Equipment in Transport at T. Kableshkov University of Transport Sofia, 158 Geo Milev Blvd, Sofia 1574, Bulgaria, E-mail: g_pavlov61@abv.bg.

[3]Elenana Dimkina is PhD Student with the Faculty of Telecommunications and Electrical Equipment in Transport at T. Kableshkov University of Transport Sofia, 158 Geo Milev Blvd, Sofia 1574, Bulgaria, E-mail: elena.dimkina@abv.bg.

$$x_2(t) = \sum_{n=0}^{\infty} x_{n2} r(t - nT_r) \ , \tag{4}$$

where $r(t)$ is a rectangular pulse of unit amplitude and width T_r , i.e.,

$$r(t) = \begin{cases} 1, & 0 \le t < T_r \\ 0, & \text{elsewhere} \end{cases} \ . \tag{5}$$

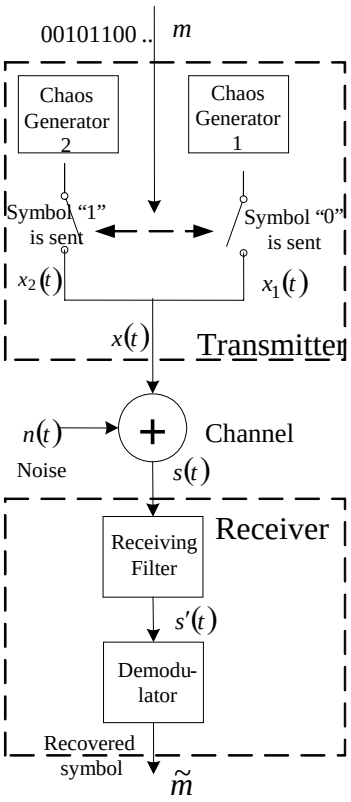


Fig. 1 Binary CSK digital communication system

Assume that the system starts at t=0 and the binary data to be transmitted has a period T_b .

Denote the transmitted data by:

$$\boldsymbol{m} = (m_1, m_2, \ldots\ldots\ldots) \ , \quad m_m \in \{0,1\}$$

If a binary "0" is to be sent during the interval $[(m-1)T_b, mT_b]$, ; $x(t)=x_1$ is transmitted by the communication channel. If the binary symbol "1" is to be sent, $x(t)=x_2$ is transmitted. Here, m is a number of the transmitted symbol.

Let

$$\beta = \frac{T_b}{T_r} \tag{6}$$

be the spreading factor, which is an integer.

Thus

$$u^{(m)}(t) = \sum_{n=0}^{\beta-1} y^{(m)}_{n+(m-1)\beta} r\left[t - \left(nT_r + (m-1)T_b\right)\right] \tag{7}$$

is the transmitted waveform for the m-th bit, where

$$y^{(m)}_{n+(m-1)\beta} = \begin{cases} x_{[n+(m-1)\beta]1} \ , & \text{if } \mathrm{m_m} = 0 \\ x_{[n+(m-1)\beta]2} \ , & \text{if } \mathrm{m_m} = 1 \end{cases} . \tag{8}$$

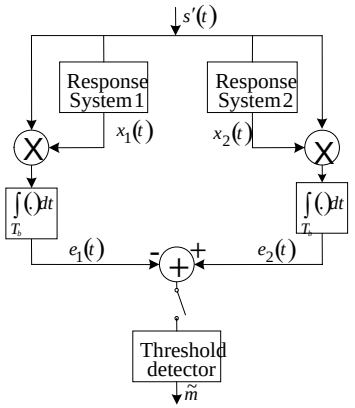


Fig. 2. Block diagram of a coherent CSK receiver

The overall transmitted waveform, $x(t)$, is

$$x(t) = \sum_{m=1}^{\infty} p^{(m)}(t) \ . \tag{9}$$

III. FUNCTIONAL KIND OF THE NOISE-RESISTANT PERFORMANCE OF A COHERENT RECEIVER WITH ADDITIVE WHITE NOISE

For ideal communication system, the channel is anticipated only with additive white Gaussian noise $n(t)$ where it is free from intersymbol interference.This is usually a good starting point for understanding basic performance relationships. In the following analysis, $n(t)$ is replaced by an equivalent noise source $n'(t)$ given by

$$n'(t) = \sum_{n=0}^{\infty} \zeta_n r(t - nT_r), \tag{10}$$

where the coefficients $\{\zeta_n\}$ are independent Gaussian random variables with zero mean and variance

$$\sigma^2_{n'} = \frac{N_0}{2T_r} \ , \tag{11}$$

where N_0 is the white Gaussian noise spectral power density.

The block diagram of a coherent CSK receiver as shown in Fig. 2. Assume that the transmitted signal by the receiving filter is

$$s'(t)=x(t)+n'(t)=\sum_{m=1}^{\infty}u^{(m)}(t)+\sum_{n=0}^{\infty}\zeta_n r(t-nT_r)=$$

$$=\sum_{m=1}^{\infty}\sum_{n=0}^{\beta-1}y_{n+(m-1)\beta}^{(m)}r\left[t-\left(nT_r+(m-1)T_b\right)\right]+\sum_{n=0}^{\infty}\zeta_n r(t-nT_r). \quad (12)$$

For the *m*-th received symbol, the output of correlator$_{(\text{“0”})}$ at the end of the bit duration equals to

$$e_1=\int_{(m-1)T_b}^{mT_b}s'(t)x_1(t)dt=T_r\sum_{n=(m-1)\beta}^{m\beta-1}y_n^{(m)}x_{n1}+\zeta_n x_{n1}. \quad (13)$$

The output of correlator$_{(\text{“1”})}$ can be shown equal to

$$e_2=\int_{(m-1)T_b}^{mT_b}s'(t)x_2(t)dt=T_r\sum_{n=(m-1)\beta}^{m\beta-1}y_n^{(m)}x_{n2}+\zeta_n x_{n2}. \quad (14)$$

The input to the threshold detector equals to

$$e=e_2-e_1=$$

$$=T_r\sum_{n=(m-1)\beta}^{m\beta-1}y_n^{(m)}x_{n1}+\zeta_n x_{n1}-y_n^{(m)}x_{n2}-\zeta_n x_{n2}. \quad (15)$$

If a “ 1” has been transmitted for the *m*-th symbol, i.e., $m_m=1$ and $y_n^{(m)}=x_{n2}$ for $(m-1)\beta\le n\le m\beta-1$, the input of the detector will be given by

$$e_{\text{“1”}}(mT_b)=T_r\sum_{n=(m-1)\beta}^{m\beta-1}x_{n2}^2+\zeta_n x_{n2}-x_{n2}x_{n1}-\zeta_n x_{n1}. \quad (16)$$

The error probability given a “1” has been transmitted is given by [3]:

$$p(e_{\text{“1”}})=F\left(\frac{\overline{e_{\text{“1”}}(mT_b)}}{\sqrt{\text{var}\left[e_{\text{“1”}}(mT_b)\right]}}\right), \quad (17)$$

where

$$F(z)=\frac{2}{\sqrt{2\pi}}\int_0^z\exp\left(\frac{-t^2}{2}\right)dt \quad (18)$$

is the integral function of Cramp’s distribution [3]; $\overline{e_{\text{“1”}}(mT_b)}$ is the mean value and $\text{var}\left[e_{\text{“1”}}(mT_b)\right]$ is the variance of $e_{\text{“1”}}(mT_b)$.

Likewise, it can be shown that when a “0” has been transmitted, the conditional error probability is given by

$$p(e_{\text{“0”}})=F\left(\frac{\overline{e_{\text{“0”}}(mT_b)}}{\sqrt{\text{var}\left[e_{\text{“0”}}(mT_b)\right]}}\right), \quad (19)$$

where

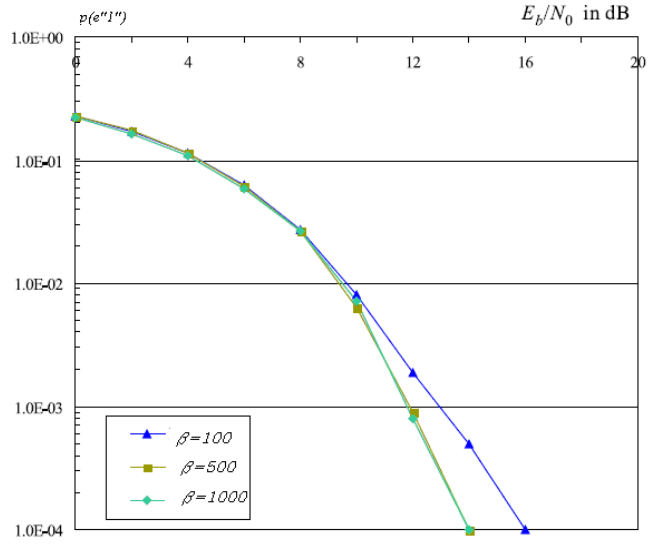
$$e_{\text{“0”}}(mT_b)=T_r\sum_{n=(m-1)\beta}^{m\beta-1}x_{n1}^2+\zeta_n x_{n1}-x_{n2}x_{n1}-\zeta_n x_{n2}. \quad (20)$$

The average bit energy of the system is given by [3]:

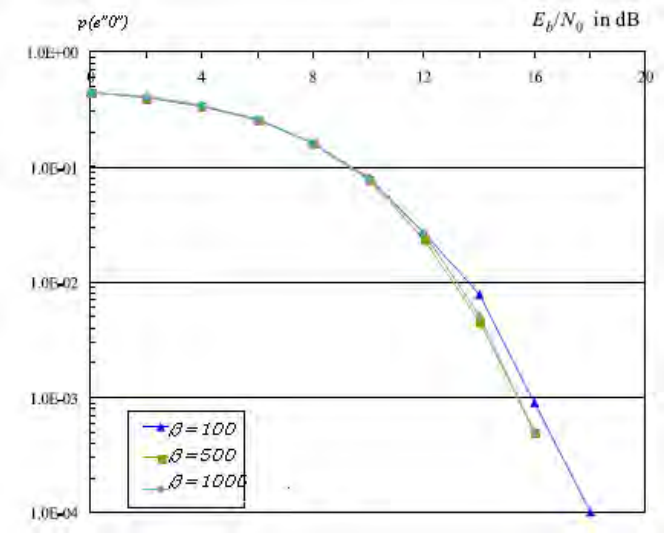
$$\overline{E}_b=T_b\lim_{T\to\infty}\frac{1}{T}\int_0^T x^2(t)dt=$$

$$=T_r\left[\beta\bar{\gamma}(x_{n2},\beta)p(e_{\text{“1”}})+\beta\bar{\gamma}(x_{n1},\beta)p(e_{\text{“0”}})\right], \quad (21)$$

where $\bar{\gamma}(x_{n2},\beta)$ and $\bar{\gamma}(x_{n1},\beta)$ denote the mean-squared values of chaotic sequence samples of length β taken from the chaotic series $\{x_{n1}\}$ and $\{x_{n2}\}$ respectively.



a)



b)

Fig.3. Error probability vs. $\overline{E_b}/N_0$

Since both series $\{x_{n1}\}$ and $\{x_{n2}\}$ are generated from the same map with different initial conditions, we have $\bar{\gamma}(x_{n2},\beta) = \bar{\gamma}(x_{n1},\beta)$ and (21) can be simplified to

$$\bar{E}_b = T_r \beta \bar{\gamma}(x_{n2},\beta)\left[p(e_{"1"}) + p(e_{"0"})\right] = T_r \beta \bar{\gamma}(x_{n2},\beta) \ . \quad (22)$$

The dependence of errors probability $e_{"1"}$ and $e_{"0"}$ for various average-bit-energy-to-noise-power-spectral-density ratios ($\overline{E_b}/N_0$) is shown in Fig. 3.

IV. Conclusion

In this paper we present the the error probability for various average-bit-energy-to-noise-power-spectral-density ratios ($\overline{E_b}/N_0$).The obtained results show that the error probability decreases with increasing $\overline{E_b}/N_0$. Moreover, it can be reduced for the same by using a higher spreading factor.

References

[1] Celikovsk´y, S., and Chen, G. Secure synchronization of a class of chaotic systems from a nonlinear observer approach.IEEE Transactions on Automatic Control11.2005.

[2] G. Kolumbán, Theoretical noise performance of correlator-based chaotic communications schemes. IEEE Trans. Circuits Syst. I, vol. 47. 2000.

[3] J. G. Proakis and M. Salehi, Communications Systems Engineering. Englewood Cliffs, NJ: Prentice-Hall, 1994.

Acknowledgements

This paper was supported by the project BG051PO001-3.3.06-0043 “Increasing, Improving and Extending the Scientific Potential of the University in Transport by Support to Development of PhD Students, Postdocs, Trainees and Young Researchers in the Field of Transport, Power Engineering and ICT in Transport” implemented within the Operational programme . “Human Resources” funded by the European Social Fund.

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

Svetlin Antonov[1] and Snejana Pleshkova-Bekiarska[2]

Abstract – **A research, which combines the measurement of both amplitude and frequency modulation of speech signals and their processing in the processing unit of the cochlear implant, is being proposed. Numeric simulation is used as the basis for a comparison between the usage of the aforementioned combination of both modulations and the usage of only amplitude modulation. Using the proposed algorithm, a comparison between the original and processed signals is drawn.**

Keywords – **cochlear implants, amplitude and frequency modulation, speech processing.**

I. INTRODUCTION

Acoustic characteristics in speech signals allow listeners to derive not only the meaning of the speech but also the speaker's identity and emotion. Previous studies using either naturally produced whispered speech [1] or artificially synthesized speech [2], [3] have isolated and identified several important acoustic cues for speech recognition. For example, computers relying on primarily spectral cues and human cochlear-implant listeners relying on primarily temporal cues can achieve a high level of speech recognition in a quiet environment [4]- [6].

The goal of this study is to verify the relative contributions of spectral and temporal cues to speech recognition in realistic listening situations. A speech signal produced by a male talker is chosen for the purpose. We propose a combination of slowly varying amplitude modulation (AM) and frequency modulation (FM) from a number of frequency bands in speech signals and testing their relative contributions to speech recognition in acoustic and electric hearing. Different from previous studies using relatively ''fast'' FM to track formant changes in speech production [8], [11], or fine structure in speech acoustics [9], [10], the ''slow'' FM used here tracks gradual changes around a fixed frequency in the subband. We evaluate the AM-only, AM plus FM, and the original unprocessed speech signal to compare these 3 situations, and to extract the MSE and the distortion.

II. METHODS

We conducted an experiment to test this hypothesis about the relative contribution of the added frequency modulation in the speech signal processing method in the cochlear implants.

In this experiment the processed stimuli contain either the AM cue alone or both the AM and FM cues. The main parameter is the number of frequency bands varying from 1 to 34.

We use a speech signal produced by a male talker (1,5s.). We conducted an experiment to test this hypothesis. The stimuli used are processed to contain either the AM cue alone or both the AM and FM cues. The main parameter is the number of frequency bands varying from 1 to 34. Different from previous studies, this experiment found that four AM bands were not enough to support good speech performance.

Thirty-four bands were used to match the number of auditory filters estimated psychophysically over the 80- to 8,800-Hz bandwidth [12].

Fig. 1 shows the block diagram for stimulus processing. To produce the AM-only and AM plus FM stimuli, a stimulus was first filtered into a number of frequency analysis bands ranging from 1 to 34. The distribution of the cutoff frequencies of the bandpass filters was approximately logarithmic according to the Greenwood map [13]. The band-limited signal was then decomposed by the Hilbert transform into a slowly varying temporal envelope and a relatively fast-varying fine structure [12], [14], [15]. The slowly varying FM component was derived by removing the center frequency from the instantaneous frequency of the Hilbert fine structure and additionally by limiting the FM rate to 400 Hz and the FM depth to 500 Hz, or the filter's bandwidth, whichever was less [16].

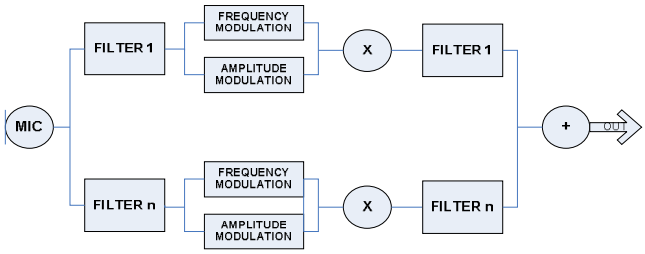


Fig. 1. Signal processing block diagram. The input signal is first filtered into a number of bands, and the band-limited AM and FM cues are then extracted. In the AM-only condition, the AM is modulated by either a noise or a sinusoid whose frequency is the bandpass filter's center frequency (not shown). In the AM_FM condition, the FM is smoothed in terms of both rate and depth and then modulated by the AM. In either condition, the same bandpass filter as in the analysis filter is applied before summation to control spectral overlap and resolution.

The AM-only stimuli were obtained by modulating the temporal envelope to the subband's center frequency and then

[1]Svetlin Antonov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: svantonov@yahoo.com.

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

summing the modulated subband signals [2], [7]. The AM plus FM stimuli were obtained by additionally frequency modulating each band's center frequency before amplitude modulation and subband summation. Before the subband summation, both the AM and the AM plus FM processed subbands were subjected to the same bandpass filter as the corresponding analysis bandpass filter to prevent crosstalk between bands and the introduction of additional spectral cues produced by frequency modulation. All stimuli were presented at an average root-mean-square level of 65 dB (A weighted) with the exception of the SRT measure in Exp. 3, in which the noise was presented at 55 dBA and the signal level was varied adaptively.

A signal, $s(t)$, can be approximated by a sum of N band-limited components, $x(t)$, containing both amplitude and frequency modulations

$$s(t) \approx \sum_{k=1}^{N} x_k(t) = \sum_{k=1}^{N} A_k(t)\cos\left[2\pi f_{ck} t + 2\pi \int_0^t g_k(\tau)d\tau + \theta_k\right] \quad (1)$$

Where $A_k(t)$ and $g_k(t)$ are the k-th band's amplitude and frequency modulations, whereas f_{ck} and θ_k are the k-th band's center frequency and initial phase, respectively.

Fig. 2 shows the block diagram for extraction of AM in the k-th subband. The AM is extracted by full-wave rectification of the output of the bandpass filter, followed by a low-pass filter LPF 1. The cutoff frequency of LPF 1 controls the maximal AM rate preserved in the AM signal. Additionally,

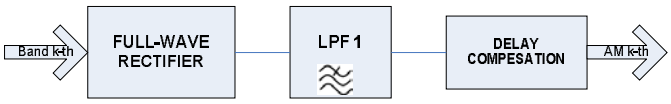


Fig. 2. Amplitude modulation block diagram.

the delay compensation box synchronizes signals between the AM and FM pathways.

Fig. 3 shows the block diagram for FM extraction in the k-th subband. First, the output of the k-th subband, $x_k(t)$, is subjected to a quadrature oscillator with the center frequency. This manipulation is equivalent to shifting the spectrum of $x(k)$ from f_{ck} to zero and $2f_{ck}$ in the frequency domain. The following low-pass filters (LPF 2 and LPF 2') then extract the slowly varying frequency components (a and b) by removing the high frequency component $2f_{ck}$. In signal processing nomenclature, the slowing-varying components and are termed in-phase and out-of-phase signals of the original subband signal $x_k(t)$, respectively.

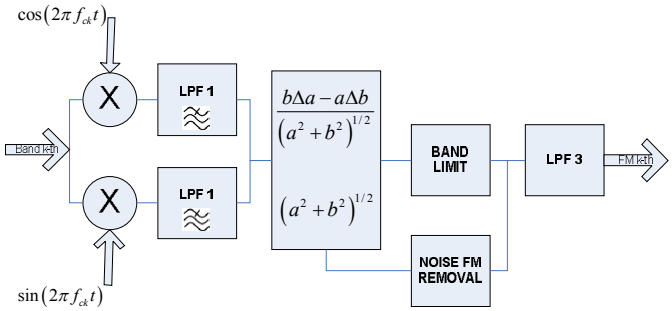


Fig. 3.Frequency modulation block diagram.

Mathematically, if $x_k(t)$ can be described as $x_k(t) = m(t)\cos\left[2\pi f_{ck} t + \varphi(t)\right]$, where $m(t)$ is the amplitude, is the center frequency and is the phase, then the in-phase signal can be derived

$$\begin{aligned} & x_k(t)\times\cos(2\pi f_{ck} t) = \\ & m(t)\cos\left[2\pi f_{ck} t + \varphi(t)\right]\cos(2\pi f_{ck} t) = \\ & \frac{1}{2}m(t)\cos\left[2\pi f_{ck} t + 2\pi f_{ck} t + \varphi(t)\right] + \\ & +\frac{1}{2}m(t)\cos\left[2\pi f_{ck} t + \varphi(t) - 2\pi f_{ck} t\right] = \\ & \frac{1}{2}m(t)\cos\left[2(2\pi f_{ck})t + \varphi(t)\right] + \frac{1}{2}m(t)\cos\varphi(t) \end{aligned} \quad (2)$$

Again, the first term in the above equation can be filtered out

$$b = -\frac{1}{2}m(t)\sin\varphi(t) = \frac{1}{2}m(t)\cos\left[\varphi(t) + \frac{\pi}{2}\right] \quad (3)$$

Dividing b by a will produce

$$\begin{aligned} & \frac{b}{a} = -\tan\varphi(t) \\ & \varphi(t) = \tan^{-1}\left(-\frac{a}{b}\right) \end{aligned} \quad (4)$$

Finally, the instantaneous frequency can be obtained

$$\begin{aligned} & FM = \frac{1}{2\pi}\frac{d\varphi(t)}{dt} = \\ & \frac{d\tan^{-1}\left(-\frac{b}{a}\right)}{2\pi dt} = \\ & \frac{b\left(\frac{da}{dt}\right) - a\left(\frac{db}{dt}\right)}{2\pi\left(a^2 + b^2\right)} \end{aligned} \quad (5)$$

In discrete implementation, differentiation in Eq. (5) can be substituted by calculating the difference in time (Δ) to obtain the slowly varying frequency modulation

$$FM = \frac{b\Delta a - a\Delta b}{2\pi\left(a^2 + b^2\right) \times T_s} \quad (6)$$

where T_s represents sampling period.

III. RESULTS

Fig. 4 shows the spectrograms of the original and processed speech sound: on the top - original test speech signal, on the left - the 1-, 8-, and 32-band amplitude modulation only, whereas on the right- the 1-, 8-, and 32-band amplitude and frequency modulation conditions

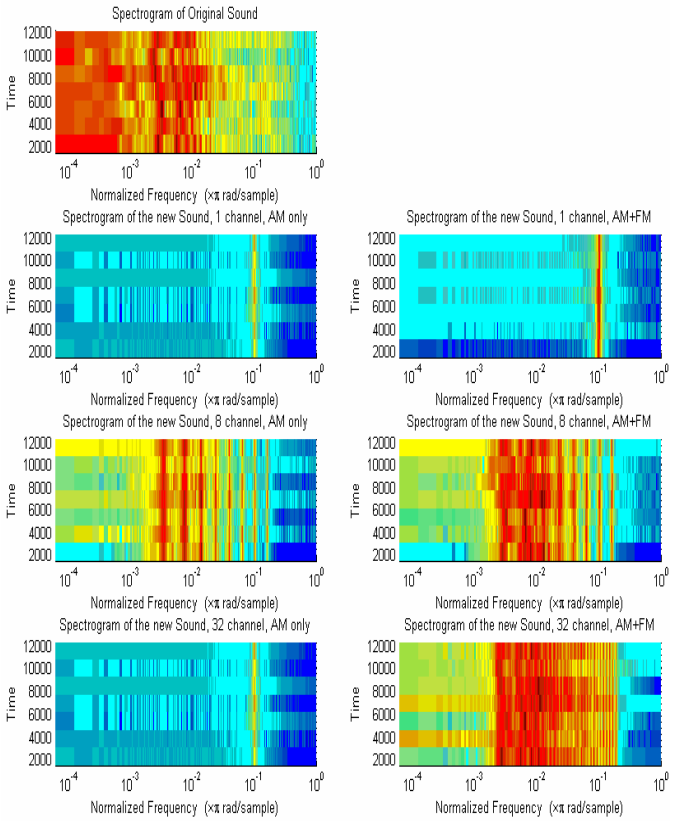


Fig. 4. Spectrograms of the original and processed speech sound. On the top- original speech signal. On left- the 1-, 8-, and 32-band amplitude modulation only. On right- the 1-, 8-, and 32-band amplitude and frequency modulation conditions.

First, we note that the original formant transition is not represented in the AM-only speech with few spectral bands (1 and 8 spectral bands left side), and only crudely represented with 32 bands (left side). In contrast, with as few as 8 bands, the AM plus FM speech (8 spectral bands right side) preserves the original formant transition. Second, we note that the decreasing fundamental frequency in the original speech is represented with even the 1-band AM plus FM speech (1 spectral band right side) but not in any AM-processed speech. The acoustic analysis result indicates that the present slowly varying FM signal preserves dynamic information regarding formant and fundamental frequency movements.

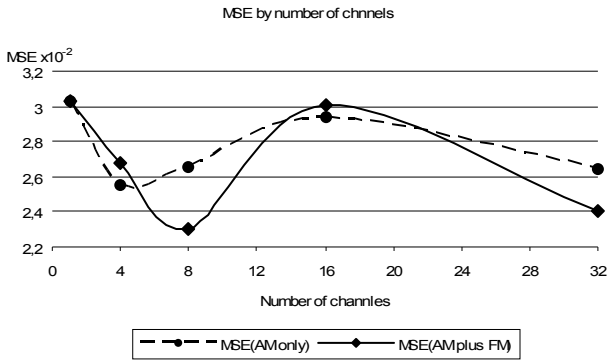


Fig. 5. Diagram showing the results of computing the mean squared error (MSE) depending of the number of the bandpass channels while AM only or AM plus FM condition.

Fig. 5 shows the diagram of the extracted mean squared error (MSE) depending on the number of the channels and the modulation conditions. We can see the lower value of MSE in 8 channels and AM plus FM conditions. We can see again that the MSE is lower in AM plus FM in 32 channels than AM only condition.

TABLE I
MSE DEPENDING ON PROCESSING CONDITIONS

Number of bandpass channels	MSE (AM only condition)	MSE (AM + FM condition)
1	0,03033781	0,03033854
4	0,02550918	0,02673373
8	0,02660029	0,02304768
16	0,02942653	0,03009370
32	0,02648830	0,02400135

Table I shows the exactly the values of MSE which are the base of Fig.5.

MSE is essentially a signal fidelity measure [20]. The goal of a signal fidelity measure is to compare two signals by providing a quantitative score that describes the degree of similarity/fidelity or, conversely, the level of error/distortion between them. Usually, it is assumed that one of the signals is a pristine original, while the other is distorted or contaminated by errors.

Suppose that $x = \{x_i \mid i = 1, 2, ..., N\}$ and $y = \{y_i \mid i = 1, 2, ..., N\}$ are two finite-length, discrete signals, original and processed. The MSE between the signals is given by the following Eq. (7).

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - y_i\right)^2 \quad (7)$$

Where,
N – number of signal samples,
x_i – value of the i^{th} sample in x,

y_i – value of the i^{th} sample in y.

IV. Discussion

Because the FM cue is derived from phase, the present study argues strongly for the importance of phase information in realistic listening situations. We note that for at least two decades phase has been suggested to play a critical role in human perception [17], yet it has received little attention in the auditory field.

The most direct and immediate implication is to improve signal processing in auditory prostheses. Currently, cochlear implants typically have 12–22 physical electrodes, but a much smaller number of functional channels as measured by speech performance in a quiet environment [18]. The results of our research strongly suggest that frequency modulation in addition to amplitude modulation should be extracted and encoded to improve cochlear implant performance. Recent perceptual tests have shown that cochlear implant subjects are capable of detecting these slowly varying frequency modulations by electric stimulation [19].

Acknowledgement

Reserch in the subject has recently commenced through a NIS TU - Sofia funded project, № 121 ПД 0063-07.

References

[1] Tartter, V. C., "Percept. Psychophys" 49, 365–372, 1991
[2] Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K. & Rabinowitz, W. M.,Nature 352, 236–238, 1991
[3] Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. , Science 270, 303–304, 1995
[4] Remez, R. E., Rubin, P. E., Pisoni, D. B. & Carrell, T. D., Science 212, 947–949, 1981
[5] Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C. & Holden, T. A.,Ear Hear. 23, 207–223, 2002
[6] Rabiner, L. (2003) Science 301, 1494–1495.
[7] Dorman, M. F., Loizou, P. C. & Rainey, D., J. Acoust Soc. Am. 102, 2403–2411, 1997
[8] Alexandros, P. & Maragos, P., Speech Commun. 28, 195–209, 1999
[9] Smith, Z. M., Delgutte, B. & Oxenham, A. J., Nature 416, 87–90, 2002
[10] Sheft, S. & Yost, W. A., "Air Force Research Laboratory Progress Report No. 1", Contract SPO700-98-D-4002 (Loyola University, Chicago), 2001
[11] Doupe, A. J. & Kuhl, P. K., Annu. Rev. Neurosci. 22, 567–631, 1999
[12] Moore, B. C. & Glasberg, B. R., Hear. Res. 28, 209–225, 1987
[13] Greenwood, D. D., J. Acoust. Soc. Am. 87, 2592–2605, 1990
[14] Flanagan, J. L. & Golden, R. M., Bell Syst. Tech. J. 45, 1493–1509, 1966
[15] Flanagan, J. L., J. Acoust. Soc. Am. 68, 412–419, 1980
[16] Nie, K., Stickney, G. & Zeng, F. G., IEEE Trans. Biomed. Eng. 52, 64–73, 2005
[17] Oppenheim, A. V. & Lim, J. S., Proc. IEEE 69, 529–541, 1981
[18] Fishman, K. E., Shannon, R. V. & Slattery, W. H., J. Speech Lang. Hear. Res. 40, 1201–1215, 1997
[19] Chen, H. & Zeng, F. G., J. Acoust. Soc. Am. 116, 2269–2277, 2004
[20] G. Casella and E.L. Lehmann, Theory of Point Estimation. New York: Springer-Verlag, 1999

Safe Operating Area Limitations in Class B Amplifiers

Hristo Zhivomirov[1]

Abstract – **The present paper describes analysis of the instantaneous power dissipated by a class B amplifier in the light of the load lines and the Safe Operating Area (SOA) limitations. Both situations of operation with steady state sinusoidal or random signals are treated, as well as working with constant complex impedance and constant resistive component of impedance.**

Keywords – **Class B amplifier, Complex load, Load lines, Safe operating area.**

I. Introduction

The study of the power parameters of the amplifier stages is a major task in the analysis and design of amplifiers. The obtained information concerns the choice of active components and the calculation of their cooling. In this paper the main point is on the SOA limitations of the active components in class B amplifiers, since this is of significant importance for their selection and properly operation but is rarely taken into account by the designers.

II. Analysis

Since the amplifiers in general case operate with complex loads the analysis must be done under these conditions. For the purposes of this paper the impedance Z_L is represented as a sum of active resistance R_L and reactive resistance (reactance) X_L which is composed of capacitance and inductance [1]:

$$\dot{Z}_L = R_L \pm jX_L = \left|\dot{Z}_L\right|.e^{j\varphi} = Z_L \angle\varphi \text{ , } \Omega \qquad (1)$$

$$Z_L = \sqrt{R_L^2 + X_L^2} \text{ , } \Omega; \ \varphi = arctg\left(\frac{X_L}{R_L}\right), \text{rad} \qquad (2)$$

where φ is a phase shift angle between the current i_C and the voltage u_{CE}.

Analysis will be done for both situations:

a) $Z_L = \sqrt{R_L^2 + X_L^2} = const.$

b) $Z_L = \sqrt{R_L^2 + X_L^2}$, at $R_L = const.$

The first case is relatively theoretical in nature, while the second covers most practical situations.

[1]Hristo Zhivomirov is a Ph.D. student with the Department of Communication Engineering and Technologies, Faculty of Electronics, Technical University-Varna, Studentska Street 1, Varna 9010, Bulgaria, E-mail: hristo_car@abv.bg.

It is convenient to express the instantaneous power dissipated by the active devices (e.g. bipolar transistors) in the area of their output characteristics $I_C = f(U_{CE})$. In this occasion the current through the device and the voltage drop across its terminals can be expressed as [2]:

$$u_{CE} = U_{cc} - u_{out} = U_{cc} - U_{out\,m}\sin(\alpha \pm \varphi) \text{ , V} \qquad (3)$$

$$i_C = i_{out} = \frac{U_{out\,m}\sin\alpha}{Z_L} \text{ , A} \qquad (4)$$

Here and forward the starting point ($\alpha = \omega t = 0$) is the point at which the current i_C passes through its zero value.

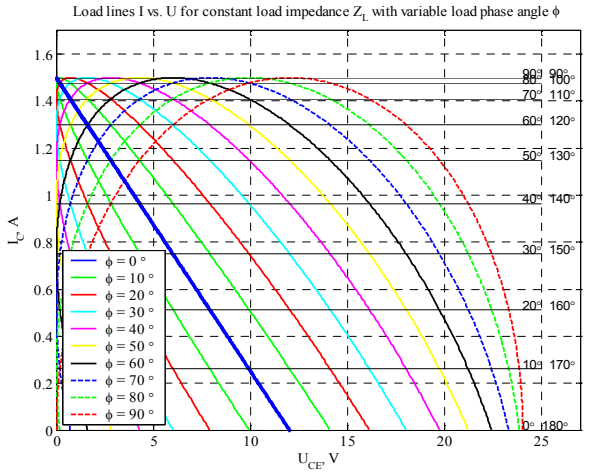


Fig. 1. Load lines at $Z_L = const.$ ($U_{cc} = U_{out\,m} = 12$ V, $Z_L = 8\ \Omega$) when the phase angle α is taken into account

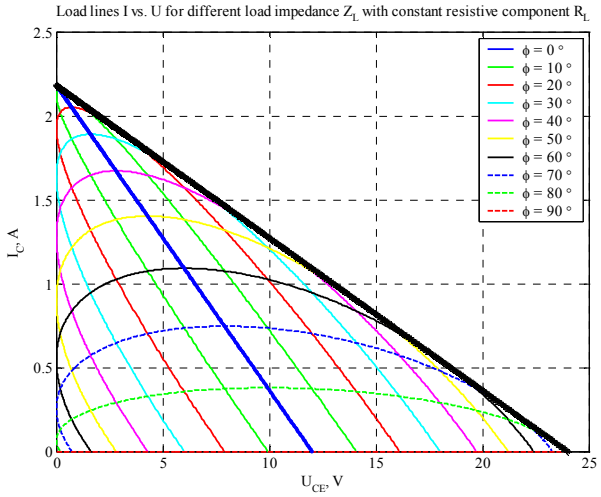


Fig. 2. Load lines at $R_L = const.$ ($U_{cc} = U_{out\,m} = 12$ V, $R_L = 5.5\ \Omega$)

By means of the last two equations a load line can be built in the field of output characteristics for both cases of operation: $Z_L = const.$ (Fig. 1) and $R_L = const.$ (Fig. 2).

The analysis of Fig. 2 shows that load line can be built tangent to the load lines at different phase angles φ, corresponding to the work with an equivalent resistance $R'_L = 2R_L$ and $U_{CEmax} = 2U_{cc}$ [2, 3].

The safe operating area (Fig. 3) is defined as the region of output characteristics of the device in which its failure-free operation is ensured.

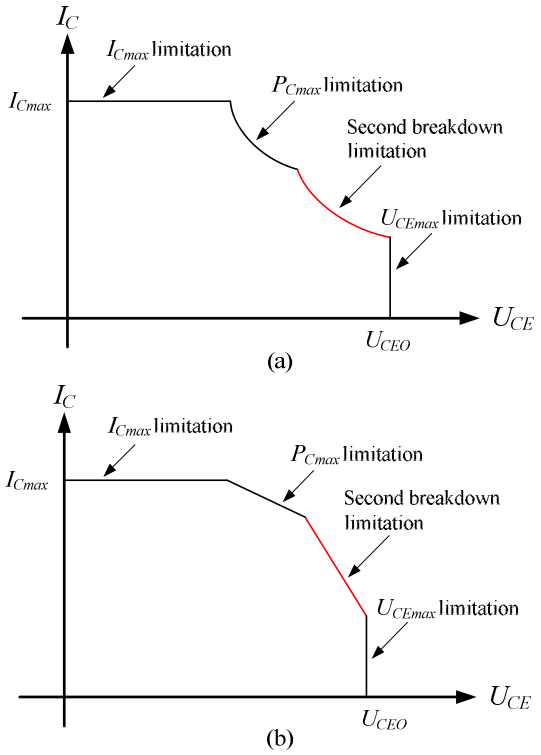


Fig. 3. Safe operating area of bipolar transistor: a) linear scale and b) log scale

The safe operation of the device depends on several limitations: the maximum current through it, the maximum power dissipation and the maximum voltage between its terminals under certain operating conditions.

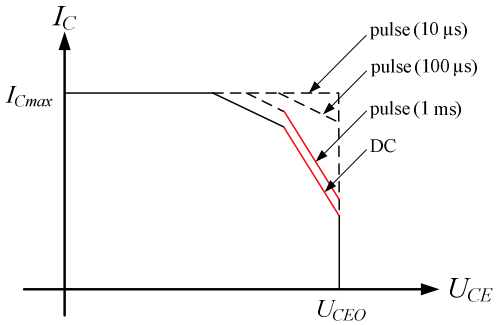


Fig. 4. Extension of SOA, when with pulse signals is worked

Bipolar transistors (BJT) unlike JFETs, MOSFETs, IGBTs, etc. are characterized by the risk of second breakdown [4], as shown in Fig. 3.

It is important to be noted the expansion of SOA when with pulse signals is operated [4] (Fig. 4) and the temperature dependence of the latter must be taken into account.

The theory and practice of signal processing [5] show that the speech and music signals have standard normal (Gaussian) distribution of the amplitudes, the density of which is described by the function [6]:

$$N(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (5)$$

the plot of whose is given in Fig. 5.

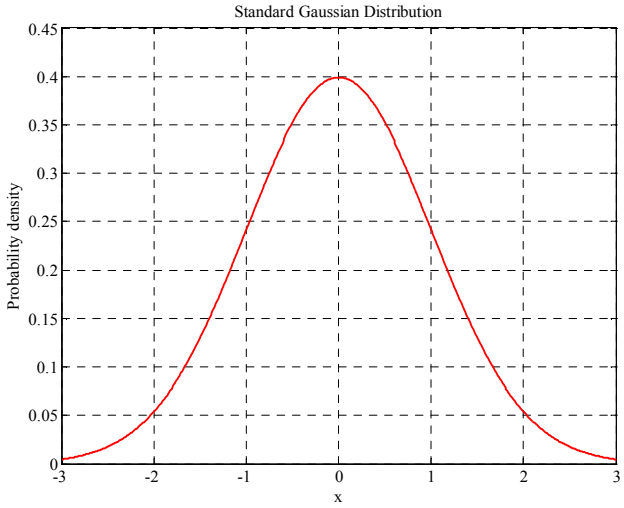


Fig. 5. Probability density of a standard normal (Gaussian) distribution

In order to evaluate the probability the instantaneous value of the random signal $x(t)$ to hit in the interval σ [a, b] it's necessary to integrate the probability density function in this interval [6]:

$$P(a \le x \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \quad (6)$$

In Table 1 the probability of getting into a few different confidential intervals is given.

TABLE I
STANDARD NORMAL CUMULATIVE DISTRIBUTION

Confidence interval [a, b]	Probability P, %
$\pm\sigma$	68.2689
$\pm 2\sigma$	95.4500
$\pm 3\sigma$	99.7300
$\pm 4\sigma$	99.9937
$\pm 5\sigma$	99.9999

It must be noted that the root mean square value U_{rms} of the random signal $x(t)$ coincides with σ [7, 8] (Fig. 6).

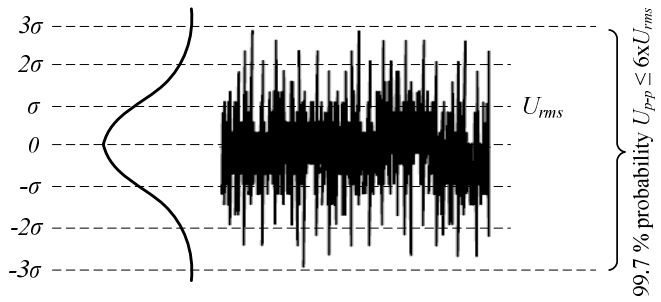


Fig. 6. Gaussian distribution of a random signal

Fig. 7 shows how to evaluate the possibility of the active element to work with a complex load.

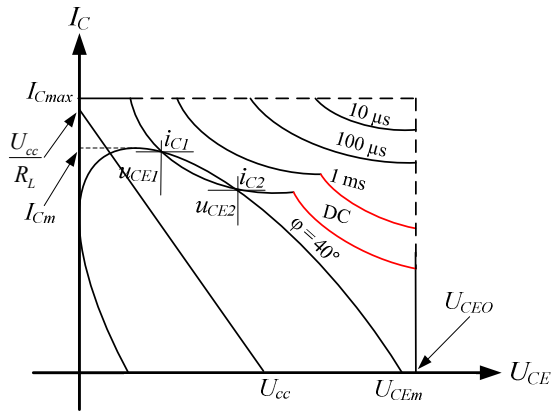


Fig. 7. Evaluation of the ability to operate with a complex load

After the load lines are drawn in the SOA according to Eqs. (3) and (4) (e.g. for values of φ at intervals of 10°), then it is necessary to calculate the residence time of the current t_i in the SOA using [1]:

$$\psi = \arcsin\left(\frac{i_C}{I_{Cm}}\right), ^\circ \tag{7}$$

$$t_i = \frac{|\psi_1^\circ - \psi_2^\circ|}{360 f}, \text{ s}, \tag{8}$$

where f is the frequency of operation, for which the corresponding load line is drawn. When the phase shift φ between the voltage and the current is known, the last can be calculated as follows:

$$f \approx \frac{tg\varphi R_L C_L + 2\sqrt{L_L C_L}}{4\pi L_L C_L}, \text{ Hz}, \tag{9}$$

where: L_L – load inductance, H;
C_L – load capacitance, F.

If the frequency is outside the amplifier band further calculations are not necessary to be made.

In the same manner a t_u can be found (note that $U_{CEm} = U_{cc} + U_{out\,m}$). In this case the failure-free operation is obtained when the load line is entirely within the SOA given for the pulses with duration $t = \max(t_u, t_i)$ and duty cycle [1]:

$$\gamma \ge \gamma_{eq} = \frac{t}{T} = t \cdot f \tag{10}$$

The estimation must be done for all load lines intersect the DC SOA-curve. If the SOA is presented in logarithmic scale then the load lines also must be presented at this way (Fig. 8).

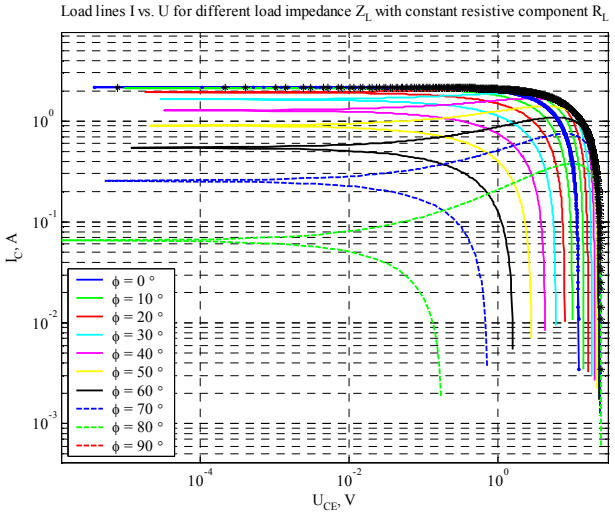


Fig. 8. Load lines at $R_L = const.$ ($U_{cc} = U_{out\,m} = 12$ V, $R_L = 5.5$ Ω) – logarithmic expression

If with random signals is worked then it is necessary to find the probability P (6) of overcoming the DC SOA-curve where:

$$a = 4\left(\frac{k_{out2}}{k_{out\,m}}\right); b = 4\left(\frac{k_{out1}}{k_{out\,m}}\right), \tag{11}$$

here k may represent current or voltage depends on which of two time intervals t_i and t_u is bigger. The coefficient before the brackets shows that confidence interval $\pm 4\sigma$ is used, but it is possible to choose another value. Then about the equivalent duty cycle can be recorded:

$$\gamma_{eq\,rand} = \gamma_{eq} \cdot P \tag{12}$$

In conclusion it is recommended to be used a regime at which the equivalent load line is entirely within the SOA obtained for DC signal.

At the same way estimation can be made when with $Z_L = \sqrt{R_L^2 + X_L^2} = const.$ is operated.

III. Computer Simulations

Fig. 9 demonstrates the SOA of IC LM1876 – class B power amplifier. The circuit parameters are: coefficient of

effective use of supply voltage $\xi = 1$, operating frequency $f =$ 1 kHz, load $Z_L = 5.7 + j5.7\,\Omega$, phase deviation $\varphi = 45°$. The same manner of presentation is used at [9, 10].

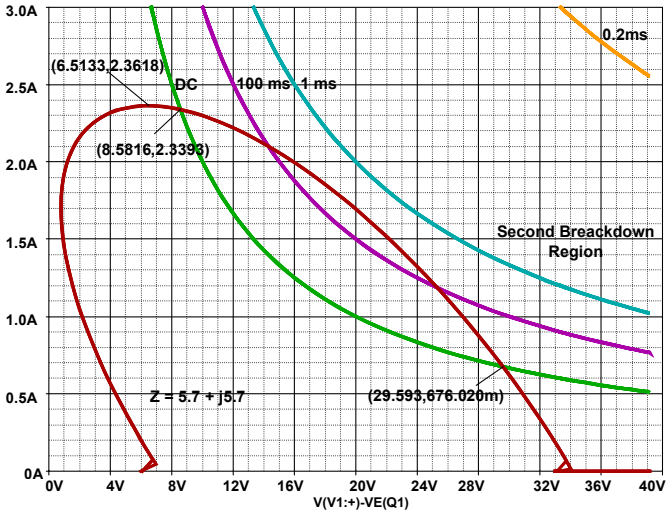


Fig. 9. SOA and load line for LM1876 operating with $R_L = const.$ ($U_{cc} = 20$ V, $R_L = 5.7\,\Omega$, $t_c = 27°$C)

When with the current is worked according to Fig. 7 then can be written:

$$\psi_1 = \arcsin\left(\frac{2.34}{2.36}\right) = 82.5°$$

$$\psi_2 = \arcsin\left(\frac{0.676}{2.36}\right) = 16.6°,$$

hence $t_i = \frac{82.5 - 16.6}{360 \cdot 1000} = 183\,\mu s$. Likewise $t_u = 102\,\mu s$ i.e. t_i dominates and it will be used in following expressions. Consequently the duty cycle of the equivalent pulse train is determined by $\gamma_{eq} = \frac{183 \cdot 10^{-6}}{1 \cdot 10^{-3}} = 0.18$. Under these conditions the amplifier could not be used, since for the utilized IC $\gamma = 0.1$, although the load line is entirely within the 200 µs SOA.

If with random signals is worked, for probability to overcome the DC SOA limitation can be written:

$$P(a \le x \le b) = 0.125,$$

where: $a = 4\left(\frac{0.676}{2.36}\right) = 1.5$; $b = 4\left(\frac{2.34}{2.36}\right) = 3.97$.

Consequently the equivalent duty cycle of the pulses is $\gamma_{eq\,rand} = 0.18 \cdot 0.125 = 0.023$. In this case, dealing with random signals, the amplifier can be used without problem.

IV. Conclusion

The problem of instantaneous power dissipation by class B amplifiers operating with complex load and steady state sinusoidal or random signals is considered. The analysis is realized using load lines and taking into account the SOA limitations.

Differences between the two regimes are marked and conclusion about parameters influencing the selection of the active components is made. Methodology for estimation of the possibility of failure-free operation is provided, based on SOA limitations and instantaneous power dissipation by the amplifier. The simulation results that confirm the theoretical statement are given. The analysis concerns the work of all active components no matter the type – BJTs, FETs, IGBTs etc.

In the statement Fig. 1 is originally presented, and the estimation methodology, Fig. 7 and Eqs. (8), (9), (11), (12) are original contribution.

This work is essential for the design and improvement of amplifier equipment as well as the theoretical analysis of amplifiers.

Consideration of the theme will continue with determination of the average power dissipation by the amplifiers when with random signal is operated.

References

[1] R. L. Boylestad, *Introductory Circuit Analysis*, Upper Saddle River, New Jersey, Prentice Hall, 2010.

[2] B. Cordell, *Designing Audio Power Amplifiers*, New York, McGrawHill, 2011.

[3] D. Self, *Audio Power Amplifier Design Handbook*, Oxford, Focal Press, 2009.

[4] M. Bairanzade, "Understanding Power Transistors Breakdown Parameters", Application Note AN1628/D, www.onsemi.com, 2003.

[5] I. Tashev, A. Acero, "Statistical Modeling of the Speech Signal", IWAENC 2010, August 30 - September 2 2010, Tel Aviv, Israel.

[6] G. Casella, R. Berger, *Statistical Inference*, Pacific Grove, CA, Duxbury, 2002.

[7] B. Carter, R. Mancini, *Op Amps for Everyone*, Oxford, Newnes, 2009.

[8] P. Horowitz, W. Hill, *The Art of Electronics*, Cambridge, Cambridge University Press, 1994.

[9] E. S. Sirakov. "The instantaneous dissipation of a class AB amplifier", 12th International Scientific Conference RADIOELECTRONIKA 2002, May 14 – May 16 2002, Bratislava, Slovak Republic, pp. 457 – 460, 2002.

[10] П. Ангелов, „Обзор на методи за изследване на ефективността при линейни аудио усилватели", Национална конференция с м.у. „Акустика 2007", 25 Октомври 2007, Варна / Списание Акустика, год. IX, бр. 9, стр. 3-8, 2007.

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

Georgy Mihov[1]

Abstract – **The stage of linear segments detection is the most important stage in the subtraction method for removing powerline interference from ECG. A special criterion is used, corresponding to the second difference for the sampled signal. Its essence is interpreted as a mathematical estimate of the acceleration of the signal and also as a non-recursive digital filter. Transfer functions of simple and complex linearity criteria have been synthesized and analysed. The analysis allows a qualitative estimating of the linearity criterion to be done. Theoretical approaches are supported by numerous practical experiments.**

Keywords – **Digital filters, ECG filtering, Mains interference removing.**

I. INTRODUCTION

The subtraction procedure for power-line (PL) interference removal from ECG signals [1-5] has already shown high efficiency. A deep analysis and comparison with other methods for PL interference suppression is done in [6]. Due to its qualities, the subtraction method continues to be subject of complementary investigations [7-9]. Its structure consists of three main stages:

– *Linear segment detection* – every ECG sample is tested whether it belongs to a linear segment;

– *Interference extracting* – if the criterion for linearity is fulfilled the the PL interference is extracted with appropriate digital filter, saved in a temporal FIFO buffer and at the same time is removed from the linear segment;

– *Interference subtracting* – the restored value of the PL interference is subtracted from the original signal in non-linear segments.

The first stage is essential for the accuracy of the subtraction method, i.e. finding linear segments. The error of calculating the interference depends on it. A special criterion for linearity $Cr < M$ is used for this purpose, where M is a practically chosen threshold.

In the practice of the subtraction procedures are used various criteria for linearity [1, 6, 8, 9], but the most proven ones are those using absolute value of 'second difference'. First differences are taken by samples located at distance of one period of the PL frequency, thus eliminating the interference influence

$$Cr = |D_i| = |(X_{i-n} - X_i) - (X_i - X_{i+n})|. \quad (1)$$

[1]Georgy Mihov is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8, St. Kl. Ohridsky Blvd, 1000 Sofia, Bulgaria, E-mail: gsm@tu-sofia.bg.

The parameter n stands for the ratio between the sampling rate Φ and the PL frequency F, i.e. $n = \Phi/F$ represents number of samples within one period of the PL interference. In the case of multiplicity between the sampling rate and the PL frequency, n is an integer.

The procedure error depends on the threshold M. As less is the threshold, as less is the committed error, but the chance to find a linear segment decreases. The involved absolute error by non-zero value of the threshold can be evaluate on $M/(2n)$.

II. LINEARITY CRITERION FOR THE CASE OF NON-MULTIPLICITY

In cases of non-multiplicity between the sampling rate and the PL frequency, the ratio Φ/F is a real number. The non-multiplicity appears also when the PL frequency deviates around its rated value. The round value n^* is used, which is the greatest integer less than or equal to Φ/F. The introduced in [9] linearity criterion for the case of non-multiplicity is performed by summarizing two criteria, normalized with a factor k_n. The basic linearity criteria is done by the equation

$$D_i = (1-k_n)(X_{i-n^*} + X_{i+n^*}) + k_n(X_{i-n^*-1} + X_{i+n^*+1}) - 2X_i, \quad (2)$$

having a transfer coefficient for the PL frequency F

$$D_F = -4(1-k_n)\sin^2\frac{n^*\pi F}{\Phi} - 4k_n\sin^2\frac{(n^*+1)\pi F}{\Phi}. \quad (3)$$

The basic linearity criteria is modified by the auxiliary filter that is summarized by two auxiliary filters, normalized with a factor k_m according the equation

$$A_i = -(X_{i+m^*} + X_{i-m^*})\frac{1-k_m}{4} - (X_{i+m^*+1} + X_{i-m^*-1})\frac{k_m}{4} + \frac{X_i}{2}. \quad (4)$$

Its transfer coefficient for the PL frequency F is done by

$$A_F = -(1-k_m)\sin^2\frac{m^*\pi F}{\Phi} - k_m\sin^2\frac{(m^*+1)\pi F}{\Phi}. \quad (5)$$

The round value m^* is used for represent the number of samples in a semi-period of the PL frequency, which is the greatest integer less than or equal to $\Phi/(2F)$. The modified linearity criterion is

$$Cr = |D_i^*| = \left|D_i + A_i\frac{D_F}{A_F}\right|. \quad (6)$$

Coefficients $k_n = \Phi / F - n^*$ and $k_m = \Phi / F - m^*$ are offered in [9]. More preciously calculation for additional coefficients is done by the formulas

$$k_n = \frac{n^* \cos\frac{\pi n^* F}{\Phi} \sin\frac{\pi n^* F}{\Phi}}{n^* \cos\frac{\pi n^* F}{\Phi} \sin\frac{\pi n^* F}{\Phi} - (n^*+1)\cos\frac{\pi(n^*+1)F}{\Phi} \sin\frac{\pi(n^*+1)F}{\Phi}}, \quad (7)$$

$$k_m = \frac{m^* \cos\frac{\pi m^* F}{\Phi} \sin\frac{\pi m^* F}{\Phi}}{m^* \cos\frac{\pi m^* F}{\Phi} \sin\frac{\pi m^* F}{\Phi} - (m^*+1)\cos\frac{\pi(m^*+1)F}{\Phi} \sin\frac{\pi(m^*+1)F}{\Phi}}. \quad (8)$$

Fig. 1 shows responses of the linearity criterion by Eq. (6), regarding as a digital filter (called D-filter).

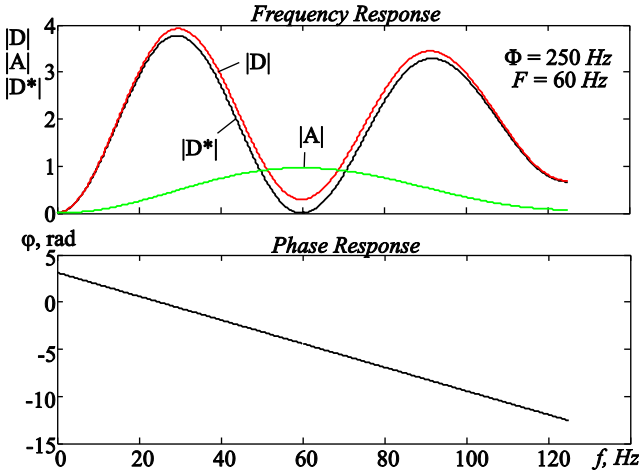


Fig. 1. Linearity criterion for Φ = 250 *Hz*, F = 60 *Hz*.

III. Complex Linearity Criterion for the Case of Non-Multiplicity

The complex criterion for linearity in cases of multiplicity [1] uses n+1 successive first subtractions and the complex second difference is formed as a subtraction of the maximal and minimal first difference

$$Cr = \left|\max(FD_0, FD_1, \ldots, FD_n) - \min(FD_0, FD_1, \ldots, FD_n)\right|. \quad (9)$$

For cases of non-multiplicity two first differences

$$FD0_i = X_i - X_{i+n^*} \text{ and } FD1_i = X_{i-1} - X_{i+n^*+1}, \quad (10)$$

with frequency responses respectively

$$FD0(f) = \sin\frac{n^* \pi f}{\Phi} \text{ and } FD1(f) = \sin\frac{(n^*+2)\pi f}{\Phi}, \quad (11)$$

are summarised with a normalising factor k_d, i.e.

$$FD^*_i = (X_i - X_{i+n^*})(1-k_d) + (X_{i-1} - X_{i+n^*+1})k_d. \quad (12)$$

The factor k_d may be determined by equalising the transfer function of the modified first difference to zero for $f = F$, i.e.

$$FD^*_F = \sin\frac{n^* \pi F}{\Phi}(1-k_d) + \sin\frac{(n^*+2)\pi F}{\Phi} k_d = 0,$$

$$k_d = \frac{\sin\frac{n^* \pi F}{\Phi}}{\sin\frac{n^* \pi F}{\Phi} - \sin\frac{(n^*+2)\pi F}{\Phi}}. \quad (13)$$

Fig. 2 shows responses of first differences represented as digital filters.

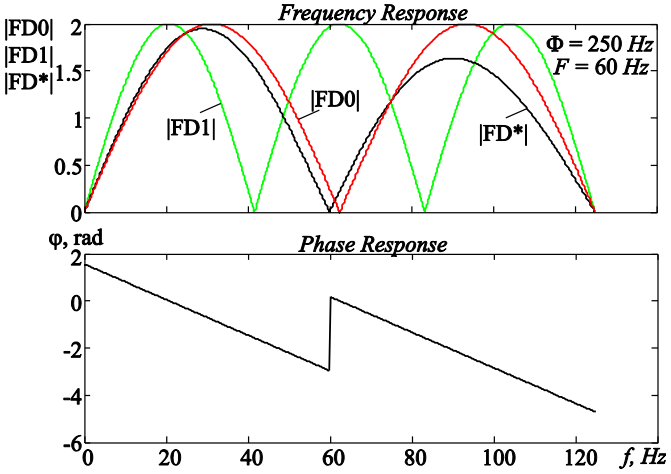


Fig. 2. Linearity criterion for Φ = 250 *Hz*, F = 60 *Hz* (n^* = 4).

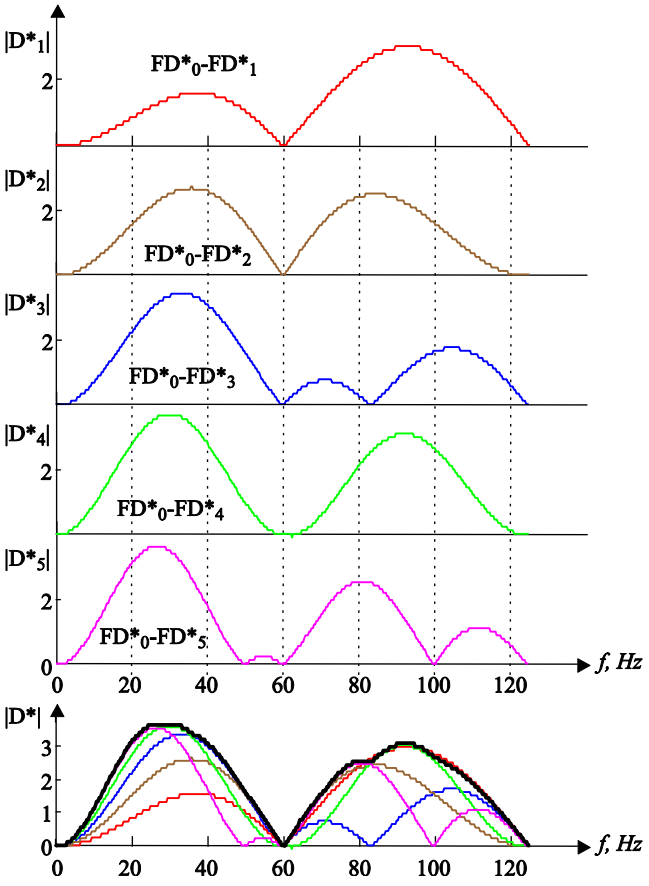


Fig. 3. Linearity criterion for Φ = 250 *Hz*, F = 60 *Hz* (n^* = 4).

In Fig. 4 are shown frequency responces of the family D-filters in case of $\Phi = 250$ *Hz* and $F = 60$ *Hz* ($n^* = 4$). The last graphic shows the total (simultaneous) operation of all linearity criteria. The wrapper of the family of curves forms the transfer function of the filter according to formula

$$Cr = \left|\max\left(FD^*_0,\ldots FD^*_{n^*+1}\right) - \min\left(FD^*_0,\ldots FD^*_{n^*+1}\right)\right| . \quad (13)$$

IV. EXPERIMENTS

The experimental investigation is performed in Matlab environment in the following sequence:

1. An episode from a signal of AHA database AHA_1001d1, which is considered as a conditionally clean from PL frequency (*Original conditionally clean signal*) The testing episode have got a duration of 16 *s* and sampling rate $\Phi = 250$ *Hz*.
2. A synthesized PL interference with amplitude $p = 0{,}5$ *mV* is added to the original signal (*Original signal + interference*).
3. The contaminated signal is treated by the subtraction procedure and the filtered signal is shown on the third subplot (*Processed signal*).
4. The error is calculated as an absolute difference between the processed signal and the originally conditionally clean signal. It is shown on the forth subplot, together the linearity criterion (*Zoomed absolute error & linearity criterion course*).

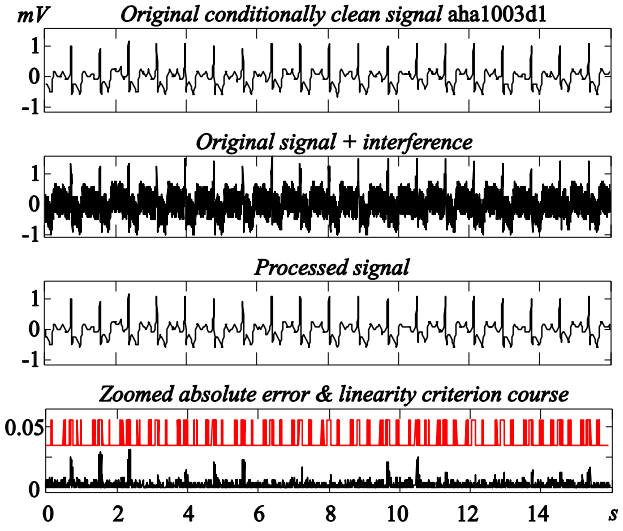


Fig. 4. Experiment with $F = 60$ *Hz* ($n^* = 4$) and a linearity criterion by Eq. (6).

The sequence of the testing is applied in two referent experiments. The first is shown in Fig. 5 with a signal aha1003d1, that is contaminated by a powerline interference with a frequency $F = 60$ *Hz*. The used linearity criterion is performed by Eq. (6) with a threshold $M = 80$ μ*V*. One may observe that the absolute error do not exceed 30 μ*V*.

At the same condition is performed the second referent experiment. The original signal is contaminated by PL interference with a frequency $F = 16{,}7$ *Hz*. The absolute error is higher than in the previous experiment and reaches 60 μ*V*.

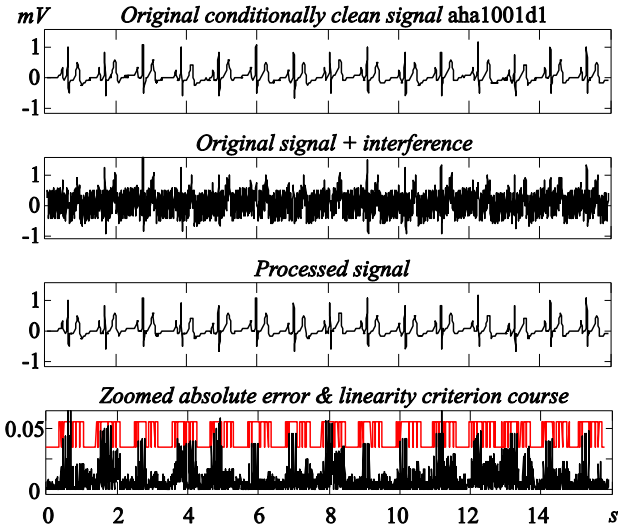


Fig. 5. Experiment with $F = 16{,}7$ *Hz* and linearity criterion Eq. (6).

Next two experiments are performed at the same conditions as previous, but using the offered complex criterion Eq. (9).

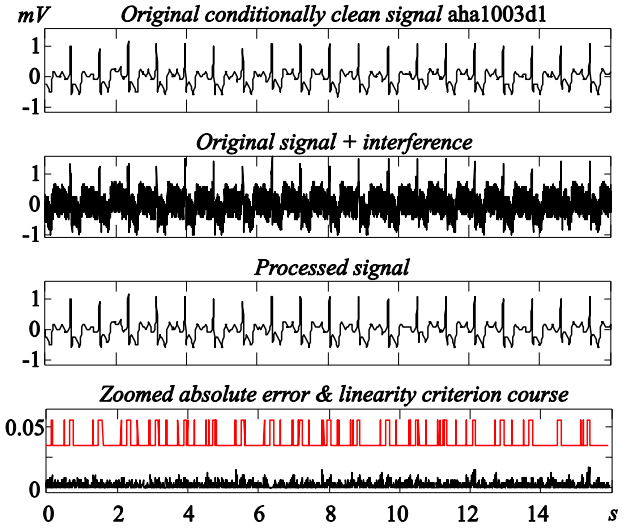


Fig. 6. Experiment with $F = 60$ *Hz* and linearity criterion Eq. (9).

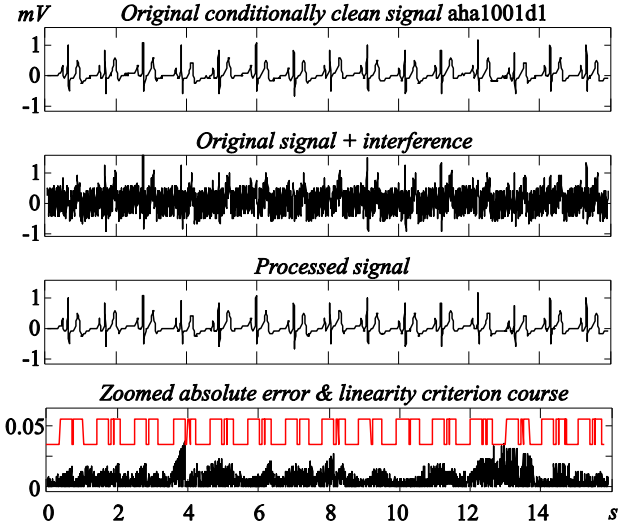


Fig. 7. Experiment with $F = 16{,}7$ *Hz* and linearity criterion Eq. (9).

Next two experiments are performed in condition on PL frequency deviation. An abrupt change in PL frequency F from $+dF$ to $-dF$ is simulated in the middle of the epoch. The fifth subplot (*Frequency deviation & linearity criterion course*) shows the PL frequency deviation (curve a – green), the estimated diversion of the PL frequency (curve b – black) and the criterion for linearity (curve c – red). It is obviously that the error committed, after reaching the stationary value of the PL frequency, is the same as in Figs. 6 and 7.

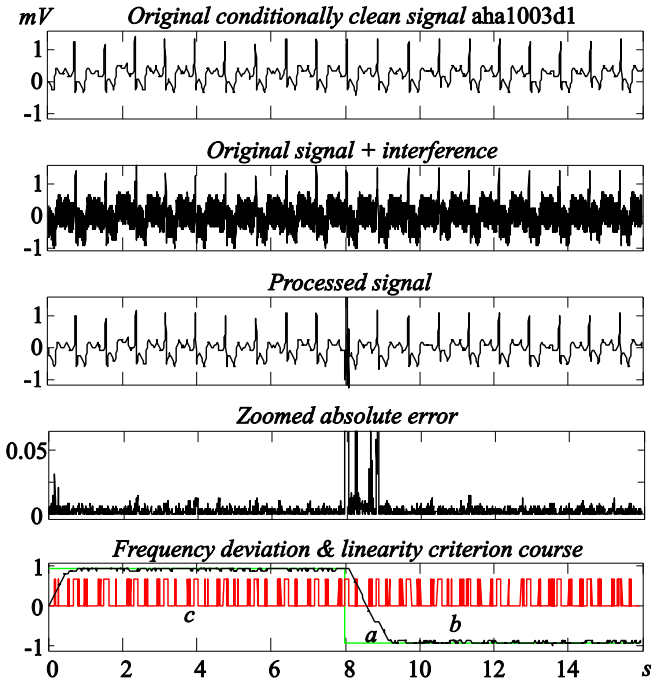


Fig. 8. Experiment with F = 60 *Hz* (n^* = 4) and a linearity criterion by Eq. (9).

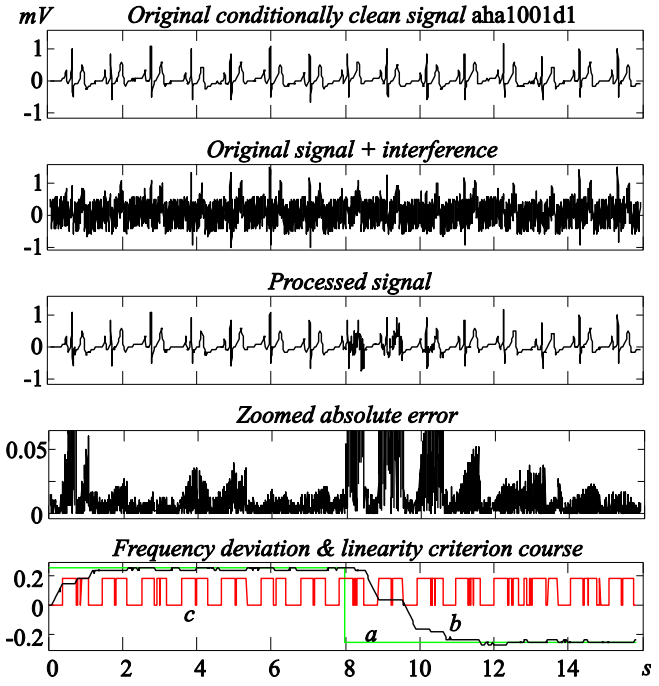


Fig. 9. Experiment with F = 16,7 *Hz* (n^* = 14) and a linearity criterion by Eq. (9).

IV. Conclusion

The article develops the subtraction procedure for removing PL interference from ECG in general case of non-multiple sampling and power line frequency deviation.

The stage of linear segments detection is the most important stage in the subtraction method for removing power-line interference from ECG. A new complex modification of the linear criterion is introduced that retains all needed features for non-multiplied sampling. The offered complex linearity criterion approximately twice reduces the absolute error of PL interference rejection – see experiment shown in Fig. 6 in comparison with Fig. 4 and Fig. 7 in comparison with Fig. 5.

Analogously results are obtained using the offered complex linearity criterion with signals contaminated by PL interference with frequency deviations – see Figs. 8 and 9. Despite good results obtained, one disadvantage has to be pointed: the range of PL frequency deviation, using complex criterion by Eq. (9), is shorten to ±0.5% in comparison to ±3%, using the criterion by Eq. (6). The computational complexity is also higher using complex criterion by Eq. (9), than using the criterion by Eq. (6), despite the Eq. (13) results in a constant that is calculated ones at procedure starting.

The involved absolute error caused by non-zero threshold M = 120μV, is less than 20 μV for the case of Φ = 250 Hz and F = 50 *Hz*. The additional error is due to the own frequency components F in the 'conditionally clean' signal.

References

[1] Ch. Levkov, G. Michov, R. Ivanov, I. Daskalov, "Subtraction of 50 Hz Interference from the Electrocardiogram", Med. & Biol. Eng. & Comput., 22, pp. 371-373, 1984.

[2] I. Christov, I. Dotsinsky, "New approach to the digital elimination of 50 Hz interference from the electrocardiogram", Med. & Biol. Eng. & Comput., 26, pp. 431-434, 1988.

[3] I. Dotsinsky, I. Daskalov, "Accuracy of the 50 Hz Interference Subtraction from the Electrocardiogram", Med. Biol. Eng. Comput., 34, pp. 489-494, 1996.

[4] I. Christov, "Dynamic powerline interference subtraction from biosignals", Journal of Med. Eng. & Tech., 24, 169-172, 2000.

[5] I. Dotsinsky, I. Christov, "Power-line interference subtraction from the electrocardiogram in the presence of electromyogram artifacts", Electrical Engineering and Electronics E+E, No 1-2, pp. 18-21, 2002.

[6] C. Levkov, G. Mihov, R. Ivanov, Ivan K. Daskalov, I. Christov, I. Dotsinsky, "Removal of power-line interference from the ECG: a review of the subtraction procedure", BioMedical Engineering OnLine, 2005, 4:50. (http://www.biomedical-engineering-online.com/content/4/1/50).

[7] I. Christov, "Power-line interference elimination from ECG: dynamic evaluation of the linearity criterion", Electrical Engineering and Electronics E+E, No 7-8, pp. 34-39, 2006.

[8] G. Mihov, I. Dotsinsky, "Removal of Power-line Interference from ECG in Case of Non-multiple Even Sampling", ICEST2007. vol. 2, ISBN 9989-786-06-2, Ohrid, Macedonia, June 24-27, 2007, pp. 633-636.

[9] G. Mihov, R. Ivanov, Ch. Levkov "Subtraction method for removing a powerline interference from ECG: case of powerline frequency deviation and non-multiple sampling", ELECTRONICS'2007, Conference Proceedings, b. 2, pp. 27-32, Sozopol, Bulgaria, 2007.

School promoting tool Multimedia project – Documentary film

Student authors: Ilche Acevski[1], Valentina Acevska[2], Mimoza Jankulovska[3],
Mentor: Igor Nedelkovski

***Abstract* – Advertising and marketing are entering every sphere of our lives. Schools are also adapting their approaches in order to attract more students. This movie is developed for the needs of Technical school in Bitola to increase publicity about school and its activities.**

***Keywords* – Advertising and marketing, School, Camera, Movie.**

I. INTRODUCTION

In a process of movie development digital DV camera, 8mm camera, Vidac Vmagic capture card with automatic digital video recorder, digital image processing, multicamera, animations and special techniques were used. Employed software was Sony Sound Forge, Adobe Premiere Pro, Adobe After Effects, Adobe Photoshop, DVD-Lab Pro 2.

II. TECHNIQUES

A. AD Conversion

After writing the scenario, materials for the film (photos, video, audio) have to be provided first. In many parts of the movie, analog to digital conversion was used (scanning of old photos and convert them into digital format for further processing and making animation, voice recording, and processing of audio from old VHS tapes, transferring video from VHS to digital format, etc.). Therefore will be discussed some techniques for processing signals from the aspect of AD conversion. To create a digital signal, need to convert the continuous data into digital form. This involves sampling and quantization.

Student autors:
[1]Ilche Acevski is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola,Macedonia, E-mail: iacevski74@gmail.com.
[2]Valentina Acevska is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola, Macedonia, E-mail: valle.mk@gmail.com.
[3]Mimoza Anastoska Jankulovska is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola, Macedonia, E-mail: jankmj2@yahoo.com.
Mentor:
Prof. Igor Nedelkovski is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola,Macedonia, E-mail: igor.nedelkovski@gmail.com.

The main idea with sampling and quantization is that a continuous image, f(x,y), need to convert to digital form. Continuous image may be with respect to the x, y coordinates, but also in amplitude. To convert continuous image to digital, have to sample the function in x and y coordinates and also in amplitude. Digitizing the values from the coordinate is sampling. Quantization is digitizing the amplitude values. A detailed discussion is given in reference [1].

B. Digital image processing

In the history section of the film and cooperation with other schools and organizations, scanned and digital images, corrected before doing animation are used. Processing of color images could be divided in two main categories: full-color and pseudo-color processing. In the first area, the images usually are taken with a full-color scanner (full-color sensor). In the second area, the problem is in a color assigning to a special monochrome intensity or variety of intensities. In a wide scope of applications, like internet, publishing, visualization, today are used techniques with full-color image processing. "Although the process followed by the human brain in perceiving and interpreting color is a physiopsychological phenomenon that is not yet fully understood, the physical nature of color can be expressed on a formal basis supported by experimental and theoretical results" [2]. If color photo is given, like in Fig.1, there are multiple ways to create a grayscale photo, given in Fig. 2. Some popular methods are:

By setting a properties in the camera menu to get a grayscale photo. This is exactly a built-in color method in camera filter with de-saturation. With this method the color information is removed from picture taken by the camera.

With the technique of the Channel Mikser often comes to misunderstanding. There is no need the RGB percentage to be equal to 100%. It can be according to the will of the person who processed the photo, and should look as he wants.

Another method is the method of desaturation, but during this, picture remains flat. A pixel can be desaturated by finding the midpoint between the maximum and the minimum of (R, G, B), like in Eq. 1. When using this technique it is necessary to reduce the level of saturation to -100%.

$$(R,G,B) = (Max(R,G,B) + Min(R,G,B))/2 \quad (1)$$

Grayscale filters can be found embedded in many software products as a professional Photoshop. With these elements is easy to create a grayscale photo.

Fig. 1. Color photo

Fig. 2. Grayscale photo

Advantages of using these tools is that can very easily with one click to get a gray image from the color image. Although all may do good job, some filters are better than others.

C. Video

It is well known that the film is a sequence of 24 frames per second. This frame rate per second is determined so that the eye can recognize the images as something that looks like a real motion. So far this is the only way to produce the film, because no one has ever was able to produce moving images in any other way, so if the eye and brain can not recognize the movement in the rapidly changing images, it is a real problem. At lower speed of 24 frames per second can be seen flickering and shaking. Flicker occurs even at this speed, but by increasing the frame rate increases the cost for obtaining the movie. To avoid these, the frame rate could be increased but this would push up film stock costs. With a rotating shutter which blanks the screen once during each frame and one more between frames, the flicker rate could be increased to 48 per second, what is a better way to do that.

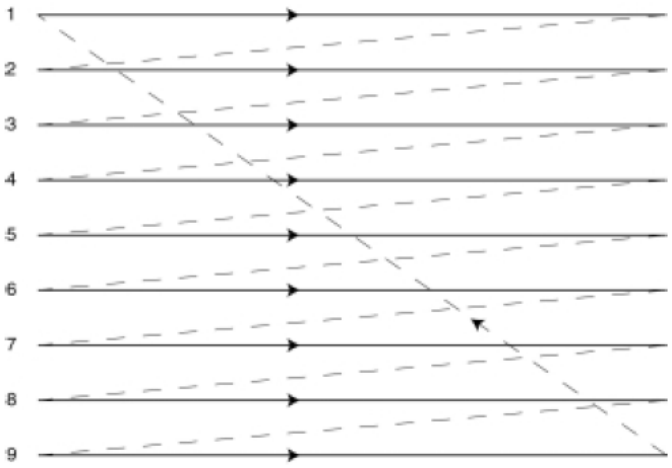


Fig. 3. Progressive scan

The picture looks stable, because the persistence of the eye, when viewing can holds the repeated image long enough, until it come the following image. But when it is a bright images on, at 48 flicker rate per second, could be noticed flickering. Monitors usually have a refresh rate of 75 HZ or more, which is the flicker rate for an image that is easy and nice for watching longer period.

When there is a live action, film is divided into a sequence of still images that can be captured and projected. The television process breaks down the image still further so that it can be sent along a wire, or broadcast. Picture in the camera is scanned by an electron beam that traces out a path which begins at the top of the picture and go down to the end-bottom. The brightness of each segment can be transduced to a voltage and sent to the wire. In Figure 3 can be seen that the electron beam start at the top left and scan the first line of the picture. When it came at the end of the line it goes back quickly to start on the next. When it is done with the entire image, it starts all over again.

"The electron beam in the video monitor or television receiver traces out exactly the same path, controlled by synchronization signals, and reproduces as closely as possible the levels of brightness of the original image on a phosphor-coated screen. The pattern that the electron beam traces out is called a raster" [3].

Like in the film, that takes a number of pictures per second to give the illusion of continuous motion. Because the mains frequency in our country is 50 Hz (60 Hz in North America, Japan etc.), easiest way is to use this as a reference point and use half of the mains frequency, as the frame rate (25 Hz or 30 Hz), rather than the 24 frames per second of film. The flicker rate need to be increased. It must be done in the same way like a shutter in a film projector. At the first place, half of the vertical resolution is scanned, with gap in between adjacent lines, and then going back to the top for filling in the other lines that were missed out, what it means that half of the picture is transmitted at the begining, and then the other half, after the first half. The eye's persistence of vision married everything together and completes the illusion. To throw in some of the technical terms, in the following text, each picture half is called a field, and the complete picture is a frame.

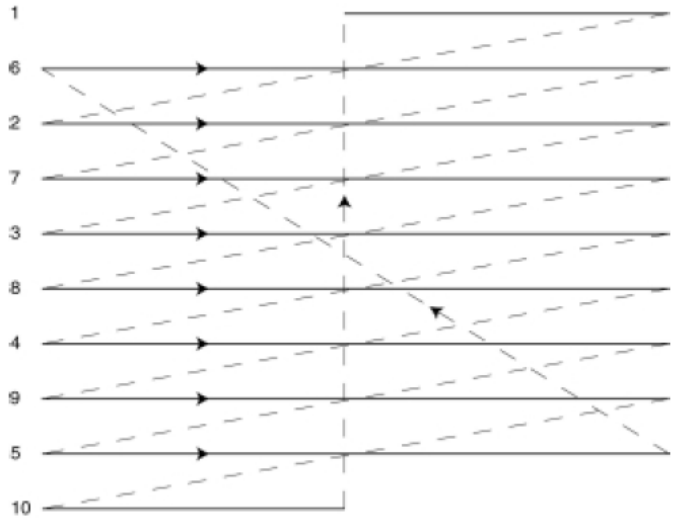


Fig. 4. Interlaced scan

Figure 4 shows an interlaced scan. If higher ratios of interlacing is used, than those with the 2:1 ratio, that is theoretically possible, there will be the noticeable individual lines that start to flicker and also moving vertical lines that will appearing to fragment, spoiling the picture, continuous motion and illusion.

"Interlaced pictures are the way SD cameras have been processing images for many years, and accordingly will get that similar look even on an HD camera. Progressive shots have no fields, as interlaced do; they simply have all the information in one frame and can give a good film look. Additionally, you can also vary the frame rates" [4].

PAL vs NTSC: The colour signals from the picture are transmitted together on the same frequency that is the one of weaknesses of the NTSC system. In this system colour signals are separated only by phase, what means that if there is some phase problem in the transmission path, the total colour bias of the picture will be changed. As the result of this NTSC system's never have the same colour. This problem is not shown in PAL system (developed by Telefunken in Germany). PAL system corrects this problem by alternating colour phase carriers in each line. In PAL system, any errors in phases are average out rather than change the complete colour of the image.

MPEG2 is technique in which a variety of techniques are applied and used to produce a data stream. MPEG2 could be understand by a wide variety of equipment (MPEG, is shortening from Motion Picture Experts Group.) But MPEG2 doesn't specify the equipment itself. Any encoder or decoder designed for video compression, since it can work with an MPEG2 and produce data stream, is good enought. This allows further research, even after the standard is set, and manufacturers can produce better equipment for producers and consumers. This is a contradictory with uncompressed digital video (AVI), where the picture quality is standard defined (best quallity of all given format) and away from the processes like A/D and D/A conversion. MPEG2 works with preferences such as resolution, bit rate, image size, and with the data stream nature of the MPEG2 itself. This means that MPEG2 isn't involved in the equipment for make it to work.

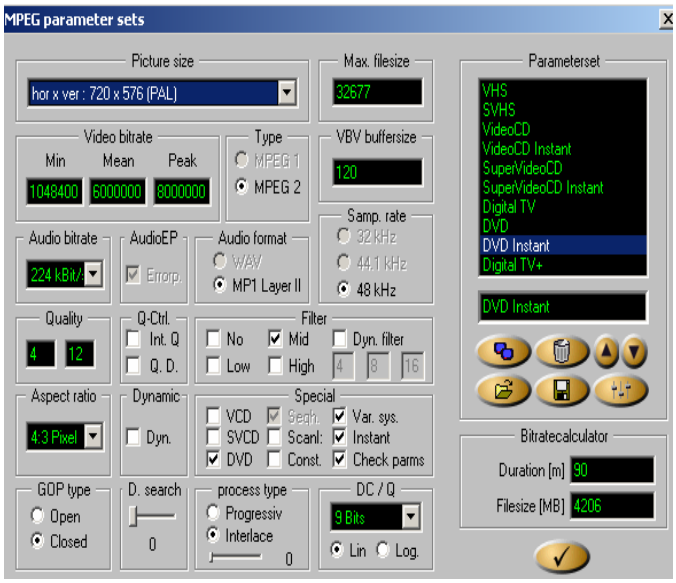


Fig. 5. Motion Picture Experts Group-MPEG parameter sets

Fig. 6. Digitized video in real time to high quality MPEG-2

Vidac Vmagic Capture Card digitize analog video sources internally in PAL and NTSC formats, and compress the digitized video in real time to high quality MPEG-2, Fig. 5 and Fig. 6. VMagic offers the hardware compression of analog SD signals and the coding of standard definition digital and high definition (HD) signals up to a resolution of 1080p30 or 1080i60 including image pre-processing (deinterlacing, filters, mirrors, overlay, etc.), decoding and video output via S-Video, HD-SDI and DVI. A detailed discussion for Vidac Vmagic is given in reference [5].

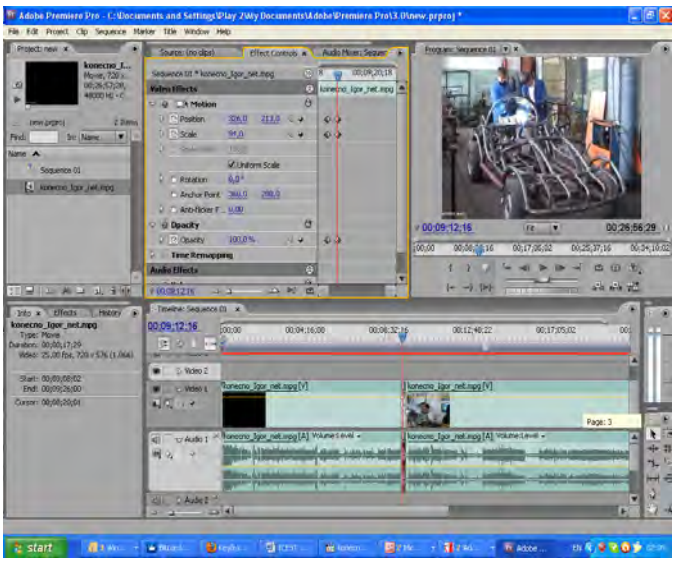


Fig. 7. Keyframes in Adobe Premiere

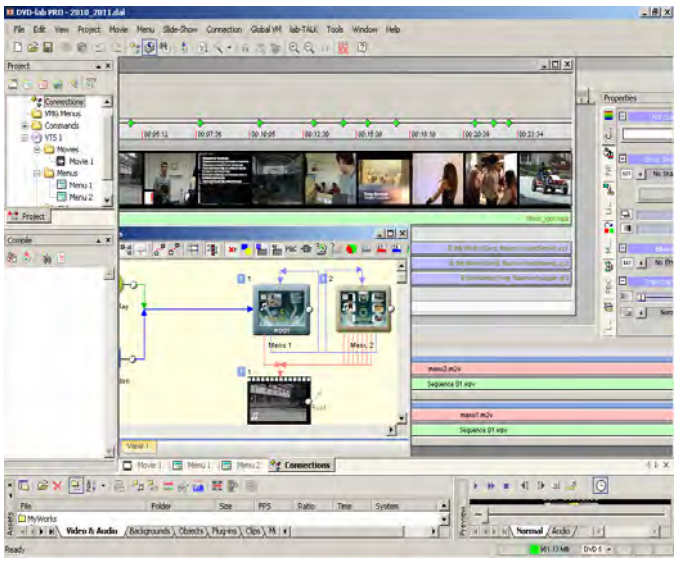


Fig. 8. DVD–Lab Pro 2

D. Audio

Sound Forge software has a complete and familiar set of tools for recording, editing and analyze audio. Sound Forge can produce music loops and samples, restore and digitize old recordings and prepared sound for multimedia and video. Recording and editing multichannel audio files is easy. With Sound Forge editing mono, stereo, and multichannel audio files (up to 32 channels) down to the sample level is nice and easy job. More detailed information is given in reference [6].

Commands like cut, copy, paste, mix, and crossfade audio are used. Perform precise event editing and create fades, crossfades, and mix audio with ease. There are drag and drop function to move audio between channels and can work on one file while processing others in the background. Most of speech was recorded straight into sound card, and then together with processed speech from VHS and DV camera, mixed with chosen background music.

E. Production

For changing the overview of a video or audio effects during time, are used keyframes in Adobe premiere. Keyframes are used for volume, opacity, brightness, scale, speed, position, colour balance, etc. There must be created more than one keyframe, and then set the desired effect values at each keyframe. Premiere will made by itself, a continual change between neighbour keyframes (interpolation). If one keyframe is created where the volume is -30dB and another keyframe 7 seconds later where the volume is +30dB. This will be interpolated with a smooth and continuous 7-second volume increasing. There are two ways in the Adobe Premiere to work with keyframes: In the Effect Controls window and in the Timeline [7]. This is shown in Figure 7.

DVD edition of this film contains English, German and Macedonian subtitles, which increases the range of those who can investigate the movie. For that purposes DVD-Lab Pro 2 [8], was used, Fig. 8.

III. CONCLUSION

During development of this documentary a lot of different techniques were used, and different software through various stages of processing, as explained above.

This movie was presented in several formal occasions and also is uploaded on the official web site of the school http://gorginaumov.edu.mk/?page_id=590. In DVD It serves as a good advertisement of the school and has already attracted bigger number of students and different stakeholders interested in technical area.

ACKNOWLEDGEMENT

Thanks to many students at SOTU “Gjorgi Naumov” - Bitola, who helped to produce this film.

REFERENCES

[1] Hayes, M.H., “Schaum's Outline of Digital Signal Processing”, McGraw-Hill, 1999

[2] Gonzales, R.C., Woods, R.E., “Digital Image Processing, 2ed”, Prentice Hall, 2002

[3] David Mellor, “A Sound Person’s Guide to Video”, Focal Press, ISBN 0 240 51595 1.

[4] Paul Martingell, “Better Location Shooting: Techniques for Video Production”, Focal Press, ISBN 978-0-240-81003-4.

[5] VIDAC, “VMagic (Plus) / VMagic Movie (Plus) Benutzerhandbuch”, VIDAC Electronics GmbH, 2002, Online Available: http://tinyurl.com/VidacSuite

[6] Madison Media Software, “Sound Forge 8”, Madison Media Software, Inc. (Sony Corporation), 2005, Online Available: http://ebookbrowse.com/soundforge-v8-0-pdf-d1868037

[7] Adobe, “Adobe Premiere Pro CS3: User Quide” Adobe Systems Incorporated, San Jose, California, 2008, Online Available: http://tinyurl.com/ad-pr-pro

[8] “Creating custom menus for Rifftrax DVD Reauthors”, Online Available: http://tinyurl.com/DVDLabPro2

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

Student authors: Valentina Acevska[1], Ilche Acevski[2],

Mentor: Igor Nedelkovski

***Abstract* – Animators use 3D animation software to produce the special effects that can be seen in movies, as such as this film of Clock Tower in Bitola. The ultimate goal is to gain animation of the shadow of the Clock Tower, during one day of 24 hours, developed in Maya 2010 and then imported into 3ds Max where the final product was obtained. There will be reviewed and discussed various results when rendering with mental ray.**

***Keywords* – 3d, animation, graphic, rendering, shadow**

I. INTRODUCTION

3D computer animation software is used to produce many of the special effects that can be seen in this movie, as well as in many movies. The architectural model of the Clock Tower represents the shape of the building and its parts that contains some of its important features. For this purpose polygonal modeling is used where points in 3D space, called vertices, are connected by line segments to form a polygonal mesh [3].

The clock tower is surrounded by nature and street lights. Modelling of nature such as trees, plants and flowers is by placing hundreds, even thousands of branches and leaves of the tree, but this approach requires great dedication, effort and time. The techniques of algorithmic and procedural generation of shapes allow 3D graphics to create abstract, but realistic geometric shapes. Fractals and graftals can create convincing plants or trees by copying a few simple geometric shapes. Contrary to the forms of nature that are modeled manually (which can have thousands of geometric shapes), these methods involve a small memory space, and both techniques are visually effective and technically efficient [2].

The texture represents shadowing or some other attribute in computer graphics that adds the area of graphic images to get an impression of physical matter. For example you can achieve a surface look like it was made of metal or glass and have their ability to reflect light. For example, scanned image or photograph of a wall, can be applied to any shape and also to simulate the material from which the wall is made [1].

Student autors:

[1]Valentina Acevska is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola,…Macedonia, E-mail: valle.mk@gmail.com.

[2]Ilche Acevski is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola, Macedonia, E-mail: iacevski74@gmail.com.

Mentor: Prof. Igor Nedelkovski is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola, Macedonia, E-mail: igor.nedelkovski@gmail.com.

Animation is derived from the Latin anima and means the act, process, or result of imparting life, interest, spirit, motion, or activity. Techniques and algorithms directly used to create and manipulate motion are: keyframing (gives the most direct control to the animator who provides necessary data at some moments in time and the computer fills in the rest), procedural animation (involves specially designed, often empirical, mathematical functions and procedures whose output resembles some particular motion), physics-based techniques (solve differential equation of motion) and motion capture (uses special equipment or techniques to record real-world motion and then transfers this motion into that of computer models) [6].

Animations can be divided into: animation of a still object, clear structural animation, dynamic simulation, individual animation and behavior animation. Animation of a still object is the easiest and most present form. It is one of the oldest and most popular forms of computer animation. It can be described as a basic requirement for animation and can be used in any form by any other categories. It can produce animated sequences rendering the scene at different positions, with light from different positions or from different sources, or by moving the point of view (virtual camera) [8].

Light objects in 3D software can be divided into: point/omni light or volume light in Maya (casts rays in all directions from a single source), spot light (casts a beam of light only in one direction and have properties to control the area affected by light), directional light (also called “Infinite” or “Sun” light cast parallel light rays in a single direction, as the sun does at the surface of the earth) and ambient light (provides perfectly even lighting throughout the entire scene, with no shading). Some of the lighting attributes are: diffuse and specular, color, intensity, decay, and attenuation. If shadows are not enabled, light goes through objects. There are depth mapped shadows (use texture maps to store information about which objects are in the shadows, that usually render quite quickly, don't support transparency and they produce are not physically correct, and will not match up with the way soft shadows look in real life) and raytraced shadows (support transparent shadows and correct soft shadows, but they are usually slow when rendering) [4].

Last stage for the final product of each 3D programming package is processing all elements (surfaces, materials, lights and movements) in the 2D image or sequence of images. One of the most important things is to synchronize the time of rendering with the image quality. There are several types of rendering: software, hardware, vector, V-Ray, finalRender, RenderMan and Mental ray [1].

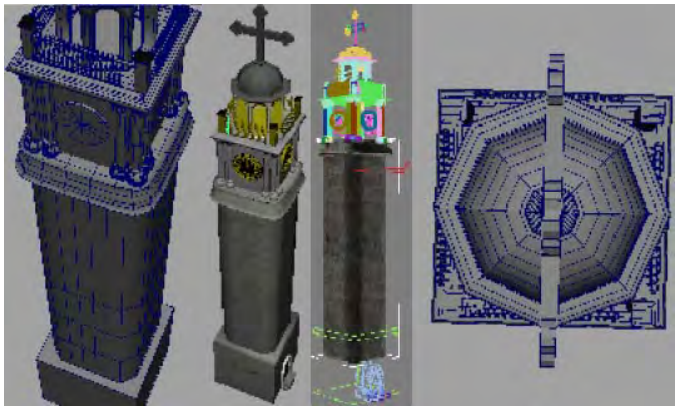


Fig. 1. Model of the Clock Tower in Maya 2010

In this case the ultimate goal is to gain animation of the shadow of the Clock Tower together with all its surroundings, during one day of 24 hours. For this purpose the methodology for the final animation will be briefly discussed. That would cover techniques and software used for modeling, texture, lighting, animation and rendering. Mental ray is mostly used tehnique in this kind of a project. There will be presented results when rendering with Mental Ray using different settings, and they will be compared with the results from the previous researches.

II. TEHNIQUES

A. *Methodology*

Fig. 1 illustrates the process of developing the Clock Tower that was modeled in Maya 2010. Textures for this model shown at Fig. 2 and Fig. 3 is taken from the actual construction in order to be able as close as possible to represent the actual characteristics of the real building blocks used in construction of the object. Some parts of the nature such as trees are taken as finished products from the web. Although the textures are used to add fine detail to the surface, this usually is not enough for modeling rough surfaces that appear on the object, such as the walls, trees or grass. The intensity of illumination of the details to such objects are adjusted independently of the lighting parameters such as direction of the light rays. Better way for modeling the bumps at the surface is applying the function of the noise to a normal surface and then uses the normal vector of rough parts when calculating the brightness of the model. Fig. 3 illustrates this technique that is called bump mapping. Bump mapping gives the illusion of the presence of bumps, holes, carving, scales, and other small surface irregularities. Moreover on a brick wall, a texture map will provide the shape, color, and approximate roughness of the bricks.

Before proceeding further it must be said that a single scene in this project, with most objects and directional light as lighting that gives the sun, modeled in Maya, takes several hours for rendering with Mental Ray. With importing of the same objects in 3ds Max, and replacement of directional light with 3ds Max built-in sunlight system, rendering of the same scene with a few settings is done in a minute.

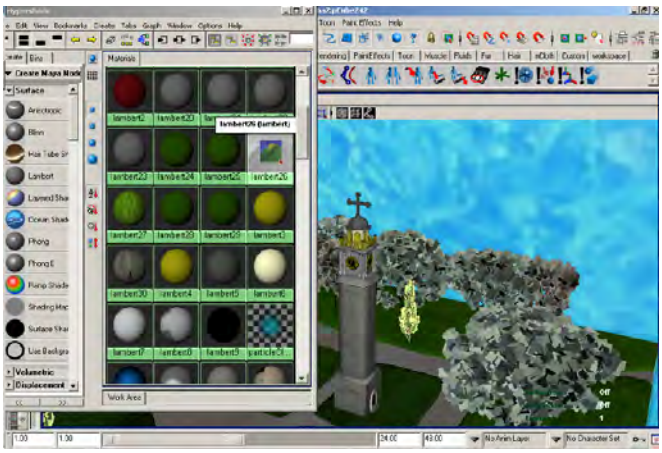


Fig. 2. Textures in Maya 2010

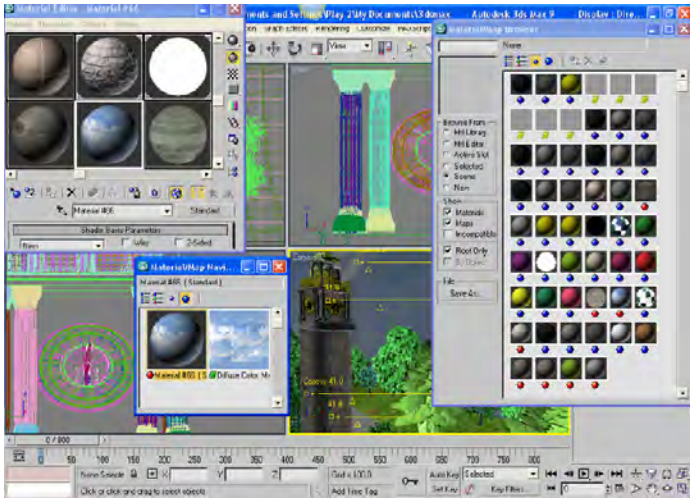


Fig. 3. Textures in 3ds Max

Therefore majority of the further project, part of the texture, animations, and additional street lights (omni light) for nighttime lighting of the park and reflector lights (spot light) for the clocks are prepared in 3ds Max.

Animation of a still object is mostly used in this project. During this, animation of the sun that moves around the planet is prepared, and as a result of that a whole day or 24 hours is obtained. In Maya, that can be done with directional light and animation of that light, while 3ds Max incorporates a special system of functions known as Daylight System. There is also animation of other light sources (street lights and reflectors) that are activated by darkening, or when the sun is on the other side of the planet. There are animations of the arrows of four clocks from the Clock Tower, mostly obtained by rotation on some of the axes, depending on the side of the world. This can be seen at Figs. 4, 5, 6 and 7. The background is also animated as you can see by the position of the sky and clouds on the same Figs. However the biggest impression leaves the camera as animated objects. Some or all external parameters of camera can be animated, where the camera flies through a mostly static environment. As the real camera, the virtual camera makes images based on its position relative to the scene. Which part of the scene will be captured in the image depends on how much the scene is in the scope of view.

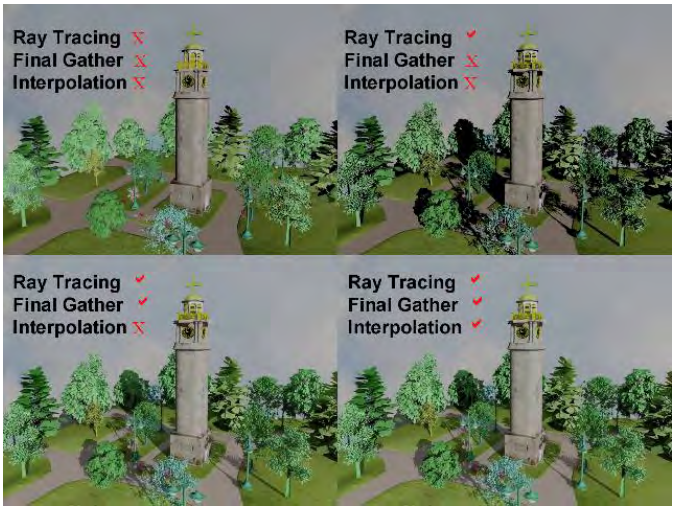


Fig. 4. Results from the 100-th scene (morning)

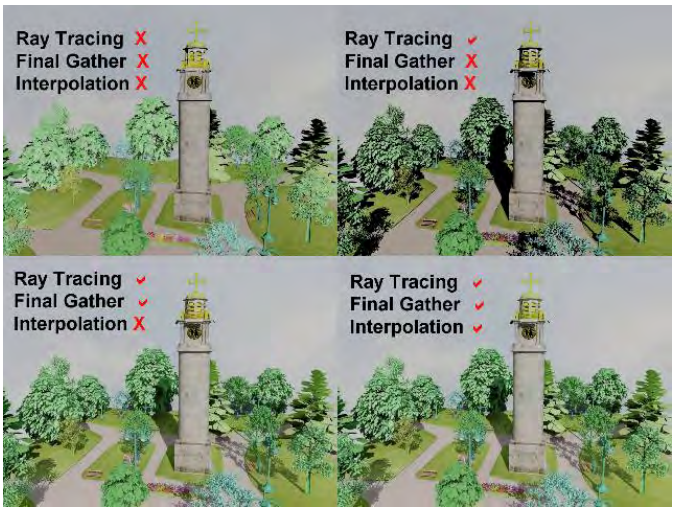


Fig. 5. Results from the 200-th scene (before noon)

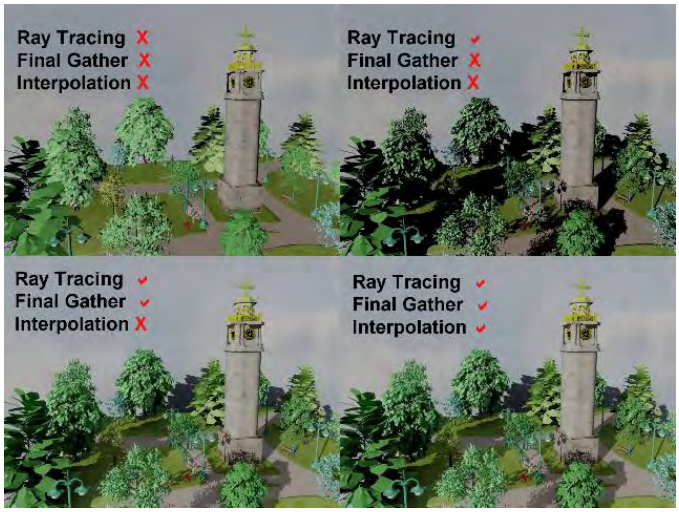


Fig. 6. Results from the 410-th scene (afternoon)

Fig. 7. Results from the7-th scene (night)

By tuning the parameters of the scope of view, the user can simulate many aspects that are encountered in the real camera, including: location (location of the camera in the scene), direction (direction in which the camera is focused), width of the lens or viewing angle (angle in which objects can be captured - depends from the distance between camera and the object), angle or orientation of the camera (rotation angle or the scope of view) and focal distance (distance from the lens in which the object is captured with the highest clarity - objects that are "out of focus" look blurred). The real camera always creates perspective projection represented by a single point of view, objects are decreased with their distancing, and parallel lines that are parallel to the plane converge (gather) to the point of vanishing. When the camera is animated object, it is necessary to control the viewpoint and viewing direction, or two degrees of freedom. The camera must always show the object of interest, which usually can be static, movable, or with movable parts on it, as in this case with the clock arrows.

Mental ray includes many features (Global Illumination, Final Gather, reflection, refraction etc.), most of them used to obtain photorealistic rendered results and is a production quality rendering application that supports ray tracing to generate images.

The software uses acceleration techniques such as scanline for the determination of the primary visible surfaces and binary space partitioning for secondary rays. Also supports simulation of global illumination including photon mapping. Any combination of diffuse, glossy (soft or scattered), and specular reflection and transmission can be simulated. For the final product this tehnique is used. Resolution of 640x480px is used for the final product. Rendering the scenes with very high resolution can be a challenge, not only because the rendering time increases by a quadratic function, but also because large memory is required. Memory problems will become bigger by using Global Illumination and/or Final Gather, because they use extra memory to calculate the indirect effects of lighting. The project will be rendered four times with diferent settings. The project will be rendered four times with diferent features: with Ray Tracing disabled, than with Ray Tracing enabled and Final Gather disabled, then with both enabled but with Interpolation over Num. FG points, and for the end with Radius Interpolation Method instead of Num. FG points.

The hardware configuration used for this project is: AMD Athlon 64 Processor 3000+ 1.81GHz, 1.5 GB DDR2 RAM, NVIDIA GeForce 6600 256MB on Windows XP.

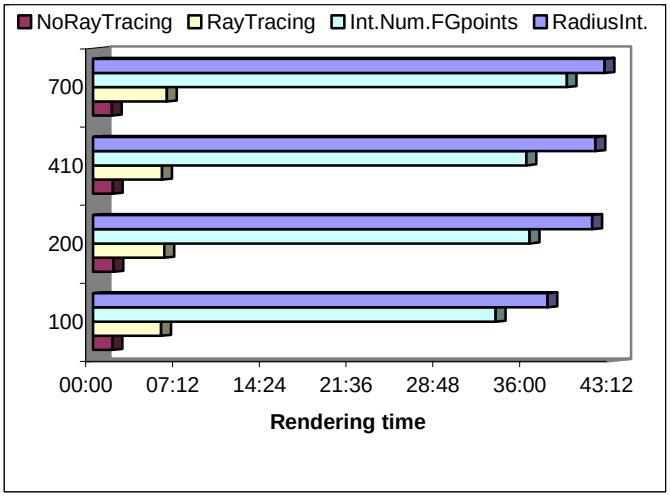


Fig. 8. Results from the rendering

TABLE I
RENDERING TIME

Scene	NoRayTr.	RayTr.	Int.N.FGP.	R.Int.
100	01:36	05:38	33:24	41:42
200	01:41	05:54	36:13	41:25
410	01:38	05:43	35:57	41:40
700	01:33	06:07	39:18	42:26

B. Results

The results of rendering can be seen at the Fig. 4 for the 100-th scene (morning), Fig. 5 for the 200-th scene (before noon), Fig. 6 for the 410-th scene (afternoon) and Fig. 7 for the 700-th scene (night) from a total of 800 scenes. The quality of the scenes with the selected features can be seen from the above mentioned Figs. But another important feature is the duration of rendering. Chart from Fig. 8 and Table I illustrates the time consumed at each rendering on each scene. It also supports caustics and physically correct simulation of global illumination employing photon maps. When rendering some of the scenes with caustics and global illumination enabled, time consuming is 1 hour 22 minutes and 29 seconds. In some cases with some options like Atmosferics enabled Mental Ray will run out of memory followed by error message and with image which contains a black squares in render buckets or 3ds Max might fail.

C. Discusion

From the Fig. 4, 5, 6, 7 and 8 and from Table I can be concluded that if Ray Tracing is disabled, there are no shadow at the scenes, although duration of rendering is about half past one minute. With Ray Tracing enabled there are dark shadows (no so realistic), but good enough for the time consuming of about 6 minutes for a scene. With Final Gather enabled with Interpolation over Num. FG points there are so realistic soft shadows with realistic form, but with time consuming of about 35 minutes for a scene. Almost the same quality is obtained with Final Gather enabled but now with Radius Interpolation Method instead of Num. FG points, and the time consuming is about 42 minutes for a scene. For a total of 800 scenes in the animated movie with the first method, time needed for rendering of all scenes is less than a day. With the second method time needed is about 3 days, with third method is about 20 days, and with last method is about 23 days. Certainly with a better hardware configuration, these times will be improved. The primary feature of mental ray is the achievement of high performance through parallelism on both multiprocessor machines and across render farms There are, also, many other features that can be researched with Mental Ray. Scanline acceleration technique is used for primary visible surface determination and binary space partitioning (BSP) is used for secondary rays. Here is some other features that can be set: Sampling Quality, Rendering Algorithms (Scanline, Raytrace Acceleration, Trace Depth), Camera Effects (Motion Blur, Contours, Camera Shaders, Depth of Field), Shadows&Displacement (Shadow, Shadow Maps, Displacement), Memory options etc.

In [5] and [7] a variety of techniques are discussed for ray tracing dynamic scenes, and some ideas and experimental results a given, but there are no real implementation yet [5], or there are no realistic shadow from the lights [7].

III. CONCLUSION

Finally can be concluded that with current hardware configuration the best way is to complete the project only with the Ray Tracing enabled, which takes 3 days. Complet project can be seen at http://gorginaumov.edu.mk/?page_id=1713. Of course that still remains a big space for further working with other features, with better hardware, and also with other algorithms and new render techniques such as iRay.

REFERENCES

[1] ALAN WATT, 3D Computer Graphic-Third Edition, Skopje, 2010 ISBN 978-608-4606-02-4

[2] Anne Morgan Spalter, THE COMPUTER IN THE VISUAL ARTS, Brown University, 1999 ISBN 0-201-38600-3

[3] Dr Dragan Cvetkovic, RACUNARSKA GRAFIKA, Beograd, 2006 ISBN 86-7991-287-5

[4] Jeremy Birn, Digital Lighting and Rendering-Second Edition, New Riders, Berkeley, 2006 ISBN: 0-321-31631-2

[5] Peter Djeu, Warren Hunt, Rui Wang, Ikrima Elhassan, Gordon Stoll, William R. Mark, *Razor: An Architecture for Dynamic Multiresolution Ray Tracing*, January 24, 2007, Available: http://www-csl.csres.utexas.edu/gps/publications/tr07-razor [Accessed 2 Mart 2012]

[6] Peter Shirley, Michael Ashikhmin, Michael Gleicher, Stephen R. Marschner, Erik Reinhard, Kelvin Sung, William B. Thompson, Peter Willemsen AK Peters, FUNDAMENTALS OF COMPUTER GRAPHICS-Second Edition, Wellesley, Massachusetts, 2005 ISBN 1-56881-269-8

[7] Ufuk Tiryaki, *Design and Implementation of Rendering System with Ray Tracing and Ambient Occlusion*, 2009, Available: http://www.cmpe.boun.edu.tr/medialab/VW/ufuktiryaki.pdf [Accessed 2 Mart 2012]

[8] Проф. Др Игор Неделковски, КОМПЈУТЕРСКА ГРАФИКА (3D МОДЕЛИРАЊЕ И АНИМАЦИЈА) Битола, 2008 ISBN 9989786-02-X

Investigation of Mixture of Gaussians Method for Background Subtraction in Traffic Surveillance

Boris Nikolov[1], Nikolay Kostov[2] , Slava Yordanova[3]

***Abstract* – Many background subtraction techniques have been developed in the past years to improve the precision of motion detection in video surveillance systems. Separating the moving objects from the background is a goal in every modern video surveillance system. Mixture of Gaussians (MoG) is one of the most complex method used for motion detection in video sequences.**

This paper further investigates the MoG method. The algorithm is implemented in Matlab and a typical traffic video is estimated. The accuracy of the algorithm is measured as a function of each variable parameter. An optimal set of parameters along with a filter are proposed in order to increase the performance.

***Keywords* – Mixture of Gaussians, motion detection, background subtraction, video surveillance.**

I. INTRODUCTION

The idea of the background subtraction is to separate the foreground moving objects from the background objects in the scene. Mixture of Gaussians (MoG) is a complex method for background subtraction [1]-[10]. There are many parameters to be set. This paper continues our research [2] and investigates the relation between algorithm parameters and the precision of the background subtraction. The results of using median filter after image processing are given.

II. MIXTURE OF GAUSSIANS METHOD

The MoG method describes each pixel in the frame by using multiple Gaussian distributions. Each pixel is represented by a distribution with its associated variance, weight and mean.

The probability of observing the current pixel value *x* at time *t* at a particular pixel location is given by [1]

$$P(x)=\sum_{i=1}^{K}\omega_{i,t}\eta(x;\mu_{i,t},\Sigma_{i,t}), \quad (1)$$

where K is the number of Gaussians distributions representing each pixel, $\omega_{i,t}$ is the weight of the i^{th} Gaussian at time t, η is the Gaussian probability density function with parameters: x is the current pixel, $\mu_{i,t}$ is the mean of the i^{th} distribution at time t, and $\Sigma_{i,t}$ is the covariance of the i^{th} distribution at time t.

The Gaussian probability density function η is given by [1]

$$\eta(x;\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}. \quad (2)$$

The covariance matrix is formed by [1],

$$\sum_{k,t}=\sigma_k^2 I, \quad (3)$$

where I is the image sequence.

The mean and the dispersion are updated by the rule [1]

$$\mu_t=(1-\rho)\mu_{t-1}+\rho X_t, \quad (4)$$

$$\sigma_t^2=(1-\rho)\sigma_{t-1}^2+\rho(X_t-\mu_t)^T(X_t-\mu_t), \quad (5)$$

where ρ is learning rate determined by

$$\rho=\alpha\eta(X_t\mid\mu_k,\sigma_k). \quad (6)$$

A particular value *x* being observed at a pixel location has a high probability if it is close to the mean of Gaussian distribution with a high weight and a low variance. So, this is the Gaussian distributions that best describes each pixel. Every new pixel value is checked again if there is a match with the existing K Gaussian distributions. By default, a match is defined as a pixel value within 2.5 standard deviations of a distribution [1]. In this paper the deviation threshold will be marked as D.

The weight of the i^{th} Gaussian at time *t* is given by [1]

$$\omega_{i,t}=(1-\alpha)\omega_{i,t-1}+\alpha(M_{i,t}), \quad (7)$$

where α is a learning rate, and $M_{i,t}$ is 1 for the model which matched and 0 for the remaining models.

In case there is no match a new Gaussian distribution is created with a mean equal to the current pixel value. The new distribution replaces the distribution with the lowest weight and highest variance. It is assumed that the background is represented by Gaussian distributions with the highest weight

[1]Boris Nikolov is with the Technical University of Varna, Faculty of Electronics, Department of Communication Technologies, Bulgaria, Varna 9010, 1 Studentska Str.
E-mail: boris.nikolov@yahoo.com

[2]Nikolay Kostov is with the Technical University of Varna, Faculty of Electronics, Department of Communication Technologies, Bulgaria, Varna 9010, 1 Studentska Str.
E-mail: n_kostov@mail.bg

[2]Slava Yordanova is with the Technical University of Varna, Faculty of Computing and Automation, Department of Computer Science and Engineering, Bulgaria, Varna 9010, 1 Studentska Str.
E-mail: slava_y@abv.bg

and lowest variance. To estimate the background, the distributions are first sorted in order of decreasing ω/σ. The pixels that belong to the background are the first B distributions and B is given by

$$B = \arg\min_{c}\left(\sum_{k=1}^{c}\omega_k > T\right), \quad (8)$$

where T is threshold which separates the background from the foreground.

III. Experiments and Results

The investigated method is implemented in Matlab. The processed video footage shows a car passing by the street. The shooting camera is stationary. The lighting of the scene is equal for all the time of the scene. The speed of the car is 22 km/h. The frame rate is 15fps and the resolution is 320x240 pixels.

The performance of the MoG algorithm depends on appropriate set of the parameters: the threshold T (Eq. (11)), the learning rate α (Eq. (7)), the deviation threshold D and the number of Gaussians distributions K. The algorithm is executed for different values of the parameters.

The accuracy is represented as function of each adjustable parameter. For quantitative evaluation of the accuracy the *F-measure* is used [3]. The above mentioned *F-measure* is a trade-off between parameters *recall* and *precision* and is defined by

$$F-measure = \frac{2*recall*precision}{recall+precision}, \quad (9)$$

where *recall* and *precision* are given by

$$recall = \frac{TP}{TP+FN}, \quad (10)$$

$$precision = \frac{TP}{TP+FP}. \quad (11)$$

In (10) and (11) TP is the number of the true positives pixels which are correctly classified foreground pixels and FP is the number of the false positives pixels. These pixels are background pixels, wrongly classified as a foreground. FN is the number of the false negative pixels, which are foreground pixels, wrongly segmented as background.

To prevent the noise in the foreground image a 3 by 3 two dimensional median filter is applied. The results of filtering the image are compared to the original estimated foreground image.

A. Threshold

The MoG method is executed for 12 values of the parameter T. The graph in Fig. 1 shows the accuracy parameter *F-measure* as a function of the threshold T. The lower curve represents the result of the experiment without filtering. The upper curve shows the value of the *F-measure* after processing the foreground image with the median filter. In Fig. 2 are shown the original frame, the estimated foreground image and the foreground image after filtering. The figure represents the results for three values of the threshold T, respectively, too low (Fig. 2a), highest *F-measure* (Fig. 2b) and too high (Fig. 2c).

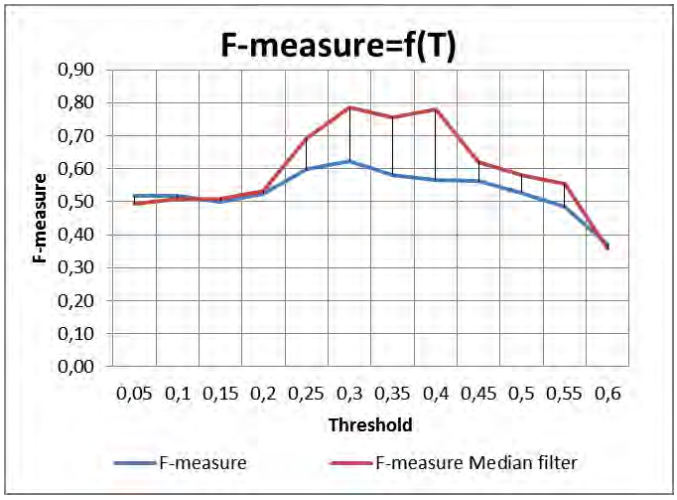


Fig. 1. *F-measure* as a function of the threshold T.

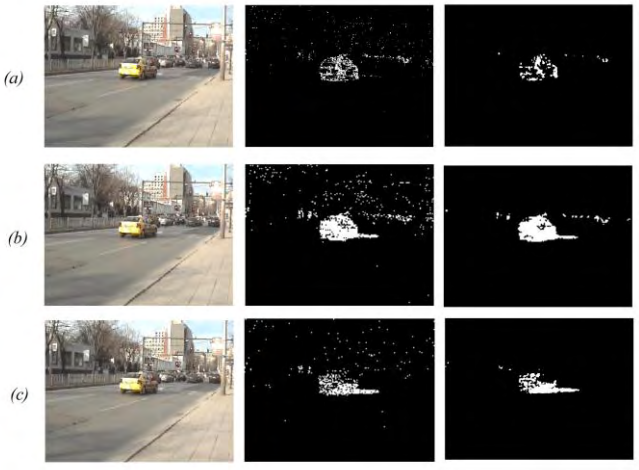


Fig. 2. Original image, foreground image and filtered foreground image executed for three values of T, $T = 0.05$ (a); $T = 0.3$ (b); $T = 0.55$ (c)

B. Deviation threshold

The method accuracy is tested by twelve values of deviation threshold D which determines the match between the new pixel value and the current distribution. In Fig. 3 the experimental results are shown. The highest value of the *F-measure* is 0.69 and is received when $D = 4$, contrary to [1]. The upper graph represents the results of filtering the foreground image with median filter. Obtaining the highest performance after median filtering at $D = 2.5$ is an

interesting result. The *F-measure* value is increased with 16%. The highest difference in accuracy between the original foreground image and filtered foreground image is at $D=1.5$ when an increasing of 31% of the *F-measure* is obtained. In Fig. 4 is shown the estimated foreground image for four different values of the deviation threshold D, respectively original foreground image (second column) and filtered with median filter image (third column).

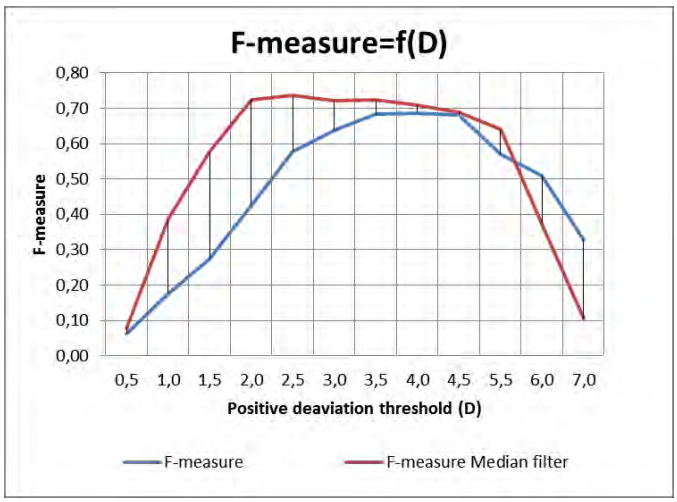


Fig. 3. *F-measure* as a function of the deviation threshold D.

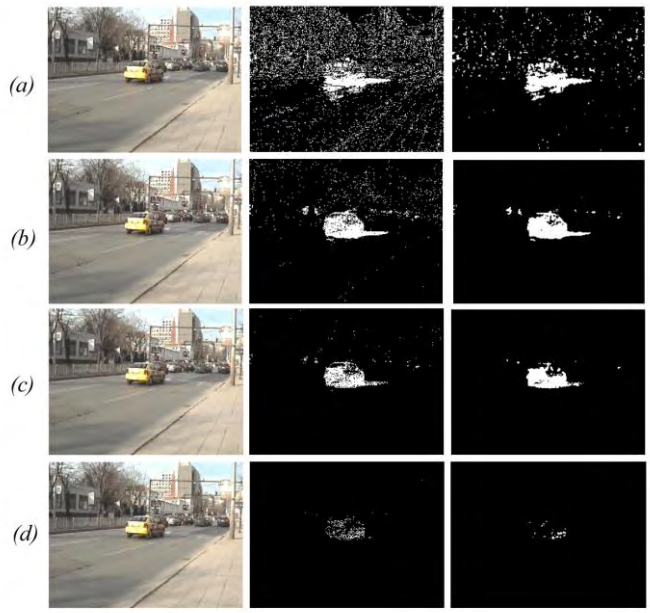


Fig. 4. Original image, foreground image and filtered foreground image executed for four values of D, $D=1$(a); $D=2.5$(b); $D=4$(c); $D=7$(d)

C. Learning rate α

The *F-measure* of the algorithm is tested for 12 different values of the learning rate α. The results are shown in Fig. 5. The accuracy remains constant from $\alpha=0.005$ to $\alpha=0.05$.

The upper graph shows the results of filtering the foreground image with the median filter. If α is below 0.005 to many pixel will have high weight and will be classified as a background and if α is over 0.05 the weight of the unmatched pixels will easily decrease and there will be more foreground objects. In Fig. 6 are shown three cases of running the method: Fig. 6(a) $\alpha=0.0025$, Fig. 6(b) $\alpha=0.01$ and Fig. 6(c) $\alpha=0.5$.

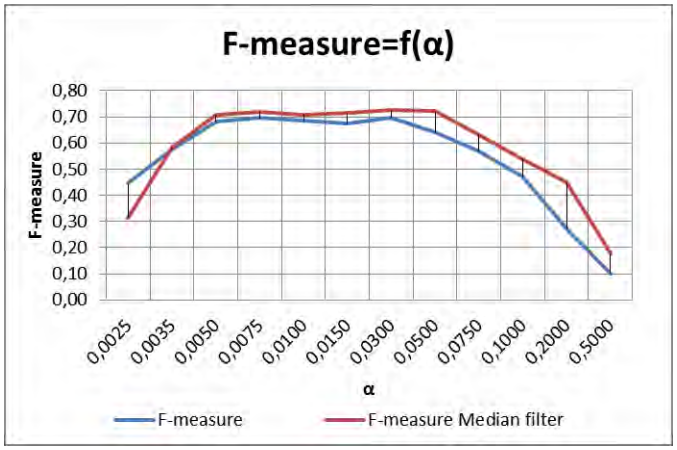


Fig. 5. *F-measure* as a function of the learning rate α.

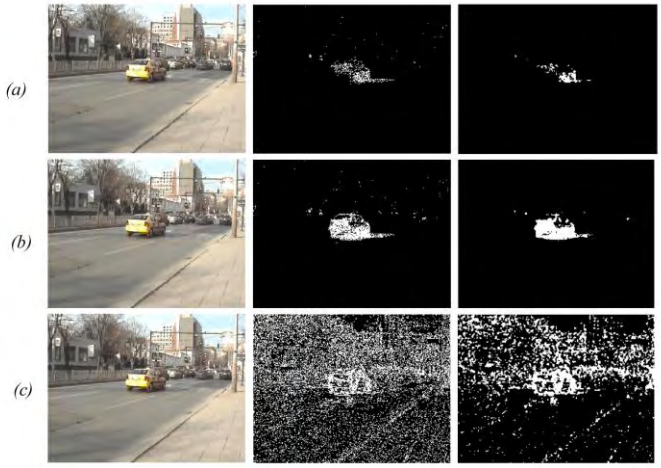


Fig. 6. Original image, foreground image and filtered foreground image executed for three values of α, $\alpha=0.0025$ (a); $\alpha=0.01$ (b); $\alpha=0.5$ (c)

D. Number of Gaussians distributions K

The result of executing the algorithm for six different K Gaussian distributions is shown in Fig. 7. Increasing the number of components in MoG does not help in improving the quality of the extracted foreground image. Highest *F-measure* is obtained at $K=3$ and $K=4$.This is because although more components can model more distributions, simple scenes are often not characterized by complex changes, and updating components of the model causes more noise. In more complex scenes the number of Gaussian distributions should be greater than $K=4$. For instance, with many small moving objects in low contrast and windy scenes, where are

waving tress and not constantly lighting. Another problem of increasing the number of distributions is that the processing time gets longer.

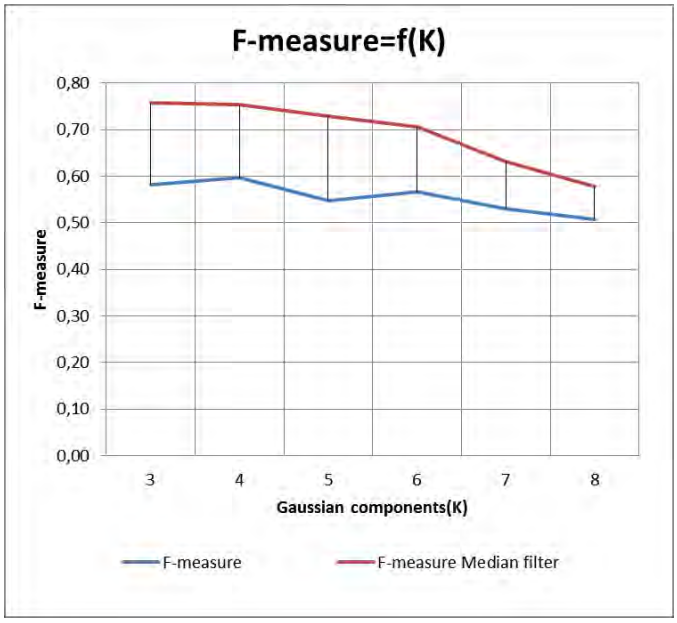


Fig. 7. *F-measure* as a function of number of the Gaussians distributions K.

IV. Conclusion

Background subtraction using MoG method was investigated. The results of running the algorithm and estimating the accuracy as a function of the algorithm variables were shown.

The processed video is a typical traffic surveillance footage. The value of the threshold T which determines the highest *F-Measure* of 0.62 is $T = 3$ and after median filtering the *F-Measure* increases to 0.78. The highest algorithm performance depends on whether or not a median filter is used, when adjusted variable is the deviation threshold. Without filtering the highest value of *F-measure* is reached when $D = 4$ and $F - measure = 0.69$. Applying the filter shifts the best deviation threshold to $D = 2.5$ when $F - measure = 0.74$. Varying the learning rate α from $\alpha = 0.005$ to $\alpha = 0.05$ does not result in a significant change of accuracy. Increasing the number of Gaussian distributions K do not result in high *F-measure* but leads to a long processing time. Filtering the noise in the processed foreground image with a *3 by 3* median filter increases the accuracy of the background subtraction. The average increasing of *F-measure* is 7.8%. Further improvement should be expected with more complex filters.

The MoG background subtraction method is high performance method if the algorithm variables are adjusted correctly. The best MoG parameter set depends on specific conditions of the sequence. The algorithm is very versatile to different applications. Real time implementation of the algorithm via digital signal processor is our aim for feature experiments and research.

Acknowledgement

This paper is funded by National Ministry of Education and Science of Bulgaria under project "Investigation of Communication Systems via Digital Signal Processors", Technical University of Varna, Bulgaria, 2012.

References

[1] C. Stauffer and W.E.L Grimson, "Adaptive background mixture models for real-time tracking*," IEEE International Conference on Computer Vision and Pattern Recognition*, vol.2, pp. 246–252, June,1999.

[2] B. Nikolov, N. Kostov "Precision of Some Motion Detection Methods Using Background Subtraction in Traffic Surveillance Video", *In Proceedings of ICEST Conference 2011*, Nis, Serbia, pp. 19-23, June 2011.

[3] Shivani Agarwal, Aatif Awan, and Dan Roth, " Learning to detect objects in images via a sparse, part-based representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(11), pp. 1475–1490, 2004.

[4] Atev S, Masoud O, Papanikolopoulos N. Practical mixtures of gaussians with brightness monitoring. *IEEE Conf on Intt Transportation Systems, Proceedings (ITS 2004)*, pp. 423-428, 2004.

[5] Dickinson P, Hunter A. Scene modeling using an adaptive mixture of Gaussians in color and space. *IEEE Conf on Advanced Video and Signal based Surveillance (AVSS 2005)*, Como, Italy, pp. 64-69, September 2005.

[6] Zang Q, Klette R. "Parameter analysis for Mixture of Gaussians", CITR Technical Report 188, Auckland University, 2006.

[7] Shimada A, Arita D, Taniguchi R. "Dynamic control of adaptive mixture-of-gaussians background model", *AVSS 2006*, Sydney, Australia, pp. 5 – 5, November 2006.

[8] Utasi A, Czúni L. "Reducing the foreground aperture problem in mixture of Gaussians based motion detection", *6th EURASIP Conf Focused on Speech and Image Processing, Multimedia Communications and Services EC-SIPMCS 2007*, Maribor, Slovenia, pp. 157 – 160, 2007.

[9] El Baf F, Bouwmans T, Vachon B. "Type-2 fuzzy mixture of Gaussians model: Application to background modeling", *International Symposium on Visual Computing (ISVC 2008)*, Las Vegas, USA, pp. 772-781, December 2008.

[10] Cuevas C, Salgado L, Garcia N. "A new strategy based on adaptive mixture of Gaussians for real-time moving objects segmentation", *In Proc. of SPIE-IS&T Electronic Imaging, SPIE*, Vol. 6811, pp. 681111, 2008.

Applied Aspects In Static Images Processing

Gergana Markova[1]

Abstract – **The content extraction from static images is an actual investigation task, because in the modern information networks the amount of information, which is stored, shared and processed under this form, is vastly increasing**

***Keywords* – Image metadata, Contextual metadata, Image Preprocessing.**

I. INTRODUCTION

The metadata (MD) could be examined as one of the ways for formal description of the information content, presented by one or more digital storage devices. The scope and structure of the metadata is determined by the nature of information processing and expected output result. In this sense we can talk about structure of the metadata, oriented to the nature of information presenting to the place, mechanism and characteristics of introduction presenting as well as to the opportunities for further processing. When setting a task for image recognition it is often skipped the stage at which there may be put limits on the input data.The inclusion of this intermediate stage is directly dependent on the fact that each of the algorithms are defined sets of constraints that determine the effectiveness of the algorithm. The MD from an image can participate in the preprocessing stage and through research and definition of certain algorithms, non-structural sets of images can be dynamically classified or limited on the base of MD. In this sense we can define a specific research task, which aims to analyze the necessity of adding one functional layer in the process of imaging called preprocessing, in which the MD of the image will be included. This is shown in Fig. 1.

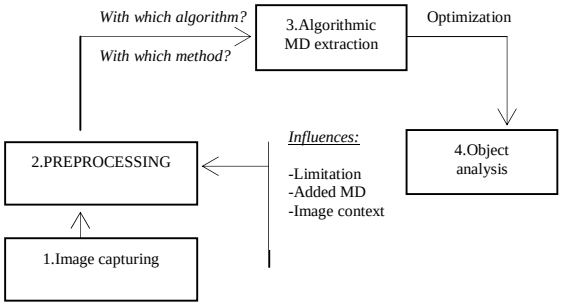


Fig. 1. Stages Image Processing with intermediate layer Preprocessing.

II. EXPLANATION OF THE STAGE PREPROCESSING

1. Image processing by the human visual system

In order to be accepted, processed and classified the environment or only object from this environment, it is projected into human's mind since the first weeks of his birth. Nixon and Aguado [1] described the human visual system as sequence of three models: 1. The eye as a **physical model**; 2. The neural system as **experimental model**; 3. The image processing by the human brain as **psychological model**. The Human vision can cope well with the relative distance, but the problem is the absolute distance while in the direction Computer vision is on the contrary. These three steps, as models and as stages determine this configuration as a preprocessing of the image from the human senses. This information is based on the logic used in basic research and development directions of Image Processing and Computer Vision.

2. Classical approach for image preprocessing

One classical approach in the filed of computer vision is described in [2, 3, 4] and includes:

- ✓ Digital presentation, filtration of the image;
- ✓ Separation of borders and edges for the objects in the image by usage of different operators Robert's, Prewitt's, Sobel's, Huckel's, Canny's, Hough's, Walsh's, Adamar's, Karunen-Loev's and transformation;
- ✓ Methods for discovery of lines in binary images. Segmentation.

In terms of processing of visual information, these authors define the discrete representation of the image and the above described stages for low-level stages. These low-level stages are discussed as *preprocessing***.**

3. Preprocessing, applied in identification of moving object

In the behavioral biometrics one of the sustainable methods of passive and automated extraction of distinguishing features of gait, is usage of low-resolution video material [5].

The main concept of the preprocessing is shown in Fig. 2

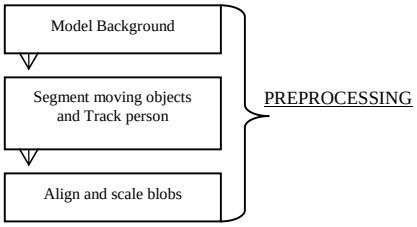


Fig. 2. Image Preprocessing in identification of moving object.

In this case in order the silhouette parameters to be extracted, the noise is decreasing and the signal is increasing. In order the input parameters to be correct for the observed object it is important to be situated far from the camera. In this case the process of segmentation is realized by a technique for separation of the background with great stability in light changing. On the other hand, in the identification it is used the proportions of the human body, based on which is applied

[1]Gergana Markova, Faculty of Mathematics and informatics, 3 arch. G.Kozavov str., bild 3, 5000 Veliko Turnovo, Bulgaria, e-mail: gerganavioletova@abv.bg.

algorithm which would not give result in incorrect image capturing, i.e. using MD from the hardware.

III. CONTEXTUAL INFORMATION AND METADATA GENERATION

1. Categorization of MD according to the context

Modern smart mobile devices generate a lot of context information. This type of MD can be categorized into several categories:

✓ Computing context describes the available resources of the devices. The terminal profile defines the technical features, including: • Type and size of the screen; • Support functions, such as WAP and Java; • Maximum size of memory for WAP decks and Java MIDlets; Support of different content types.

✓ User context contains the user profile like location, close friends, social status, opinion about various subjects and personal interests.

✓ Physical context describes lighting, noise levels, traffic conditions, temperature. The information is used to identify gestures and movement, positioning of the device.

✓ Timed context contains the current day, hour, week, month, year.

✓ Computing context is used to adapt content and transfer it.

Bluetooth and GPS context is generated by mobile phones or small multisensory digital cameras, which accurately locate and retrieve location and events by automatically capturing, at certain intervals [6, 7]. By using sensors are detected changes in the levels of luminance, motion, room temperature. The cameras of the modern mobile phones allow collection of MD in terms of image capturing. Systems are realized, which collect information about the location of the captured image, as well as the use of contextual MD in terms of capturing. The idea is the use of spatial-temporal-social context, in order the location of the captured object to be identified in sharing platform. It is used simplified faceted metadata hierarchy [8, 9]. The data are the object in the image, that we are interested in, its sharing with other users and comparison with already available objects in a database stored on the server [10]. Over the past decade, research has focused on the description of algorithms for extracting content from low-level metadata i.e. from the hardware Fig. 3. According to the data quality from the Onboard light sensor and Motion sensor, functional analysis and classification for further user processing are structured.

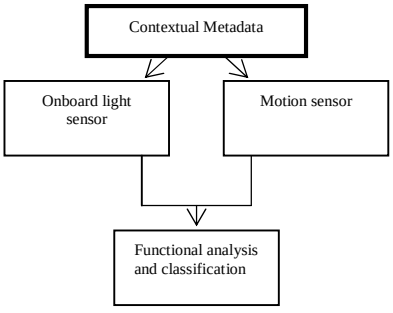


Fig. 3. Process of processing of Contextual Metadata.

The producers allow by manually adding of data by user to be shared and identified MD, which include semantic information. Today the objective is the extraction of content to be completely automated process and to exclude the participation of users.

2. Devices as sources for contextual metadata.

Some hardware devices could generate metadata with the help of software. With them we can understand the conditions under which undergo the process of fixing the image. The most information-rich is the standard RAW. The ROW format has its own variations and each producer supports his own decisions: .raf (Fuji), .crw, .cr2 (Canon), .kdc (Kodak), .mrw (Minolta), .nef (Nikon), .orf (Olympus), .dng (Adobe), .ptx, .pef (Pentax), .arw (Sony), .x3f (Sigma). If the input data are précised and additionally processed the images can "candidate" for inclusion in one or another algorithm, which aims extraction of definite context for further analysis.

3. Preprocessing in Medical diagnostics.

Image diagnostic.

In the image diagnostics in visualization of images are generated shapes of objects by discovering of photons or the usage of electromagnetic waves. When forming the image in medical devices, some of the most important features are *noise level* and image resolution [11, 12, 14]. In these static and video images for the purpose of research, forecasting and diagnostics is necessary the search for a definite type of contextual content. The development of techniques for detection and identification of constituent views in each video, extract clinically important frames, and generate time-compressed clinical summaries, solve specific tasks using metadata - generated and added [9].

3.1. **CT** (Computed tomography) have three advantage:

- Compression of 3D volume of the structure of the human body into a 2D image, where the over and underlying tissues and structures overlap, leading to reduced visibility and contrast.
- The entire volume can be reformatted into one of three positions - sagittal, coronal and axial, depending on what projection should be seen the image.
- The high level of sensitivity of CT, depending on the level of scattered photons ensures improving of the effectiveness. In sequential CT is used graphical matrix 512x512 and 16-bit resolution.

3.2. **MRI** (Magnetic resonance imaging) works in the radio-frequency range. As an advantage it is shown the excellent soft tissue contrast, which can be managed as transmitted by the device input signal.

3.3. **SPECT** (Single-photon emission computed tomography) - gamma camera operates in two modes – for a flat image, capturing an image projection of the body or part of the body and SPECT mode by using different angular distances. In this mode, single or multiple detectors revolve around the patient by defined angular steps. At each step, usually is 64x64 or 128x128 in 2-D projection. These projections are reconstructed into 64x64x64 or 128x128x128 topographic volumes in order to be visualized. One serious problem is the

process of creating and restoring of lost resolution and contrast enhancement.

SPECT is a rapidly changing field, and the past several years have produced new developments in both hardware technology and image-processing algorithms [12]. New clinical devices include high-count sensitivity cardiac SPECT systems that do not use conventional collimation and the introduction of diagnostic-quality hybrid SPECT/CT systems. While there has been steady progress with reconstruction algorithms, exciting new processing algorithms have become commercially available that promise to provide substantial reductions in SPECT acquisition time without sacrificing diagnostic quality.

The γ-ray emissions collected by the SPECT systems are not linearly related to the ray sums of activity in the patient because of tissue attenuation. As a consequence, if these attenuated projections are reconstructed, the resulting tomographic slices contain artifacts and will not accurately reflect the true internal distribution. Factors other than attenuation limit the quality of reconstructed images. These include spatial resolution, scattered radiation, and statistical fluctuations (image noise).

3.4. **NMR** (Nuclear magnetic resonance) - in result of the used technology, it is obtained an image with dimensions 256x256, fixed pixel size 1 mm^2. One problem is the high relation signal/noise, as the noise has a different character. The temperature noise caused by heating of tissues, which deforms the incoming data. Such a problem causes the movement of the patient during the session, heartbeat, breathing, peristalsis, fatty deposits can cause pressure or displacement result is the same.

By using each technology image reproduction for medical purposes, the direction in which they are improved is the removal of fixed problems arising from the used devices and the subject of study. One way to improve the output image is additional image processing before the visualization. The assessment of the image as final result is still being analyzed by a specialist, not automatically. There are already thousands of libraries with available images used for training of specialists and for the needs of telemedicine. Problem remains their classification.

3.5. **Some approaches.**

The basis of most tomographies is the idea that the internal structure of the object can be represented by a series of parallel cross sections. The method of obtaining the 2D tomographic image has two stages: first stage of forming the projection data, the second stage - the projection data to recover (reconstruction and restoration) image cross sections (for example using Fourier theorem for cuts). For the Image Preprocessing in the first phase, incoming data depend on the technical characteristics of the device and setup options. They define the structure in the second stage. If they are not correct, the result of the reconstruction phase would not be correct, this applies to CT and NMR. Therefore, Medical Image processing is considered as one of the most complex areas. In PET and SPECT the actual image is the output of an inverse reconstruction algorithm. It is important to keep in mind, that these kind of images have a very high noise level (10% and more). This makes the interpretation of these images very difficult, especially if the image is used as input of a pattern classification algorithm. In medical imaging such an algorithm will normally try to find the type of tissue at each pixel or voxel.There are used approaches for Medical Image Preprocessing that include filters which are applied to remove noise while preserving semantically important structures such as edges - methods based on nonlinear Partial Differential Equations (PDEs) [13]. It is used for medical images such as mammograms, CT and MR images.

4. Preprocessing in Character Recognition.

Text recognition relates to many fields of implementation - scanning of documents and card indexing, reading a text from photos, automatic reading of meta data from documents, recognition of car registration numbers, numbers of wagons, parcels, road signs, land marks. The quality of meta data of the input data in document scanning is related to brightness, contrast, legibility of symbols, size of scanned image and file format, in which the file is saved. If this step of Preprocessing is neglected, further processing for recognition and classification will be with low reliability. Applied algorithms either will not work out, or will register objects less in number. In one of the methods every character set is a set of templates [14]. On Fig 4. are illustrated several defined stages of sub-divided recognition process:

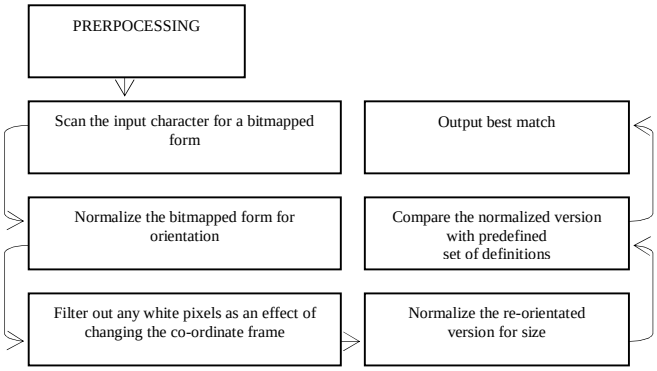


Fig.4. Process of Character recognition in one of the algorithms for recognition.

In the recent past, one software could read, for example, British *license plates* only, other could read plates from Hong Kong only, etc. It was not accidental: the geometry of the plate, as well, as its syntax, were essential parts of the plate reader software. Without the presumption of a fixed plate geometry (character ratios, character distribution, font type, plate's colour, etc.) and a well defined syntax, the algorithm may no even find the plate on the picture. Nowadays, a good algorithm should read all plates from Europe with the same level of quality. There are indeed a wide variety of plate types in Europe:

- black (dark) characters on white (light) plate;
- white characters on black plates;
- one-row plates;
- two-row plates;
- plates with different character-size;
- latin and cyrillic fonts;
- plates with or without region's shield or special mark, etc.

The authors of one such software [17] offer three additional sub-algorithms before the OCR:
-a plate localizing sub-algorithm, responsible for finding the plate on the picture.
-a contrast/brightness normalization sub-algorithm, responsible for equalizing the plate picture.
-a character segmentation sub-algorithm, responsible for finding and cutting out the individual characters on the plates and pass them to the OCR.

No doubt, the better the quality of the input images are, the better conditions the license plate recognition algorithm has, and thus the higher license plate recognition accuracy can be expected to be achieved. In order to expect reasonable results from a plate recognition algorithm, the processed images should contain a plate: with strictly defined resolution; with good sharpness; with high contrast; under good lighting conditions; in a good position and angle of view [15, 18]. Things described above, and attached examples prove the claim, that Image metadata are of great importance as input data and are of a crucial significance for the performance (or not) of the core License Plate Recognition.

IV. CONCLUSION

In [17], we suggested one approach for increasing the efficiency of algorithms for metadata extraction from static images with aim to expand the volume and content of generated MD and enabling search of content in large libraries of static images. Extraction of contextual content from MD in an image is an important step for its adaptation and transfer for classification.

The preprocessing is an important stage of images processing that direct the correct data input, meeting certain conditions, to correct preprocessing algorithm.

REFERENCES

[1] M. Nixon, A. Aguado, Feature Extraction and Image Processing, First edition 2002
[2] М. Стоева, Ю. Петкова, Обработка на визуална информация, Варна, 2009
[3] Г. Гочев, Компютърно зрение и невронни мрежи, ТУ, София, 1999
[4] D. Marr, E. Hildreth, Theory Of Edge Detection, Psychology Department And Artificial Intelligence Laboratory, Cambridge, 1980
[5] D. Sharmila MCA, Mphil, Dr. E.Kirubakaran, Member, IACSIT, Image and Formula Based Gait Recognition Methods, International Journal of Computer and Electrical Engineering, Vol. 2, No. 2, April, 2010
[6] D. Byrne, B. Lavelle, A. Doherty, Gareth J.F. Jones, A.F. Smeaton, Using Bluetooth & Gps Metadata To Measure Event Similarity In Sensecam Images, CDVP & AIC, Dublin City University, Ireland, 2007
[7] J. Huang, C. Tsai, Omniguider-The Next Generation of GPS Navigation System, International AR Standards Meeting-June 15-16 2011
[8] R. Sarvas, E. Herrarte, A. Wilhelm, M. Davis, Metadata Creation System for Mobile Images, MobiSYS'04, June 6-9, 2004, Boston, Massachusetts, USA, ACM 2004
[9] D. Marc, Editing Out Video Editing, IEEE MultiMedia, 10 (2). 2-12, 2003
[10] D. Marc, S. King, N. Good, R. Sarvas, From Context to Content: Leveraging Context to Infer Media Metadata, University of California at Berkeley School of Information Management and Systems, USA
[11] O. Demirkaya, Image processing with MATLAB: applications in medicine and biologyet, CRC Press Taylor, 2009
[12] M. Madsen, Recent advances in SPECT imaging. / Nucl Med 48: 661-673, 2007
[13] J. Weickert, Ch. Schnцrr, PDE-Based Preprocessing of Medical Images, Department of Mathematics and Computer Science University of Mannheim, Germany, 2000
[14] F. Hussain, J. Cowell, Character Recognition of Arabic and Latin Scripts, Department of Computer & Information Sciences, De Montfort University, 2005
[15] Hsien-Chu Wu, Chwei-Shyong Tsai, Ching-Hao Lai, A License Plate Recognition System In E-Government, National Taichung Institute of Technology, Taiwan, 2006
[16] Е. Витальевна, И. Сафонов, Способ повышения качества цифрового изображения на основе метаданных, описание изобретения к патенту 09.06.2010
[17] G. Markova, O. Asenov, M. Todorova, One approach for increasing the efficiency of algorithms for metadata extraction from static images, ICEST 2011, Nish, Serbia
[18] http://www.platerecognition.info/

Management of Software Project using Genetic Algorithm

Milena Karova[1], Nevena Avramova[2], Ivaylo Penev[3], Yulka Petkova[4]

Abstract –**This paper presents a heuristic method – genetic algorithm to solve the Project Management Problem. The problem is complex, NP complete. The objectives are to minimize the project duration and to minimize the project cost. The constraints are that each task must be performed by at least one person and every person must have a set of knowledge. The algorithm must define the degree of dedication of each employee. The Genetic Algorithm (IGAPM) proposes a binary chromosome encoding, single crossover, two types of selection and flip-bit mutation.**

Keywords – **Project Management Problem, Genetic Algorithm, chromosome, fitness function, project cost, constraints.**

I. INTRODUCTION

The software projects consist of interrelated activities with a set of necessary skills to perform them. The activities should be performed as much as possible at lower costs and less overlap utilizing available resources (staff with skills). Project activities and resources need to be organized in such a way that the project duration and costs are minimized and the project quality is maximized.

The presented algorithm uses as instance an implementation of software project. Project activities steps are: specification (making specifications according to customer requirements), programming (individual characteristics), architecture (defining the system architecture) and interface testing. The knowledge as a resource that is not quantified a number of related activities and employees. Any employee involved in the project implementation has a set of knowledge (skills), enabling its participation in the project. The project cost includes the employee's salaries. These are direct costs. Indirect costs of a project, such as licenses, rent, equipment and more are not included in the project.

II. THE PROJECT SCHEDULING PROBLEM

A. Definition of the Project Problem

The software project consisting of five activities Ak, which are common in projects of this type. The resources to the project are employee with skills. The links between the activities are presented in Fig.1.

The skills (knowledge) necessary for the project was presented as a set: knowledge = {C #, MySQL, PHP, Network, Design, PMR}. PMR (project management roles) is the allocation of roles in project management.

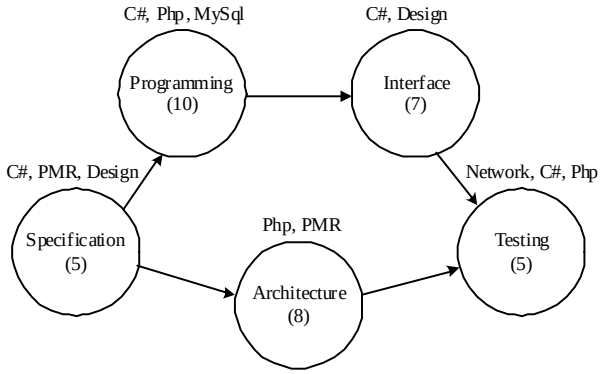


Fig. 1. Software Project Activities

Define the project P_V = {Ak1, Ak2, Ak3, Ak4, Ak5} [Table 1].

TABLE I
PROJECT ACTIVITIES

Activities	Name Ak_i^{title}	Duration (days) Ak_i^{t}	Activities precedents Ak_i^{pred}	Skills Ak_i^{know}
Ak1	Specification	5	no	C#, PMR, Design
Ak2	Programming	10	Specification	C#, Php, MySql
Ak3	Architecture	8	Specification	Php, PMR
Ak4	Interface	7	Programming	C#, Design
Ak5	Testing	5	Interface, Architecture	Network, C#, Php

There is a set of people working on the project P_V = {E1, E2, E3, E4, E5} [Table 2]

The duration of activity tj in the project P_V depends on Akj dedication-part-time percent of employees (which is dependent on knowledge) [1]. The duration and the other activities of the project are: $t_1 = 3$, $t_3 = 4$, $t_4 = 3$ and $t_5 = 2$. The activities supplying on critical path are determined by the

[1]Milena Karova is with the Department of Computer Science and Technologies at Technical University of Varna, 1 Studentska str, Varna 9010, Bulgaria, E-mail: mkarova@ieee.bg.

[2]Nevena Avramova is a master student with the Department of Computer Science and Technologies at Technical University of Varna, 1 Studentska str, Varna 9010, Bulgaria, E-mail: nevenka@abv.bg

[3]Ivaylo Penev is with the Department of Computer Science and Technologies at Technical University of Varna, Email: ivailopenev@yahoo.com

[4] Yulka Petkova is with the Department of Computer Science and Technologies at Technical University of Varna Email: jppet@abv.bg

critical path method. For P_V project there are: specification, programming and interface testing.

B. Evaluation Functions

The project duration is determined by the critical path. Many of the activities involved in the critical path t_{cr} [Eq.1] denote by CPA (Critical Path Activities) [4].

$$t_{cr} = \sum_{j \in CPA} \frac{Ak_j^t}{\sum_{i=1}^{Z} m_{ij}} \quad (1)$$

TABLE II
PROJECT STAFF

Employee	Name E_i^{name}	Salary (leva) E_i^{sal}	Project Dedication (%) E_i^{comm}	Knowledge E_i^{know}
E_1	Ivan Ivanov	1100	100	C#, MySql, Php
E_2	Georgi Vasilev	900	100	Php, Network
E_3	Gergana Kancheva	1100	50	Php, PMR, Design
E_4	Veselin Georgiev	900	100	Design, PMR, Network
E_5	Emilia Avramova	1000	100	C#, MySql

$$t_{pr} = t_{All} - t_{cr} \quad (2)$$

$$t_{All} = \sum_{j=1}^{A} \frac{Ak_j^t}{\sum_{i=1}^{Z} m_{ij}} \quad (3)$$

$$Oc_M = u_{cr} * t_{cr} + u_{cost} * S_{cost} + u_{pr} * t_{pr} \quad (4)$$

Eq.2 is an overlap time. Eq.3 is the full project duration. Eq.4 is the fitness function. u_{cr}, u_{cost} and u_{pr} are weight parameters and m_{ij} is a employee's part-time of the project activities.

III. THE GENETIC ALGORITHM IMPLEMENTATION

A. Algorithm Description

IGAPM (Implementation Genetic Algorithm for Project Management) is the realization of genetic algorithm for software project management. The algorithm is a part of a group planning algorithms and it is defined as a tool for planning projects (in particular, and software projects). The application provides the ability to manage projects so that project will be completed at - short term and / or minimal cost and with minimal overlapping activities.

IGAPM provides: an interface enabling the user to implementation of projects and their saving in format XML; GA management through its basic and additional parameters to obtain the optimal solution for each run of the GA (the decision to submit via charts) and removal of most - good fitness function.

IGAPM application is realized by programming language C# development environment and Visual Studio. Net 2008. For plotting graphs is using MS Chart VisualStudioAddOn.exe. The implementation of the algorithm realizes the critical path method based on the literature [4].

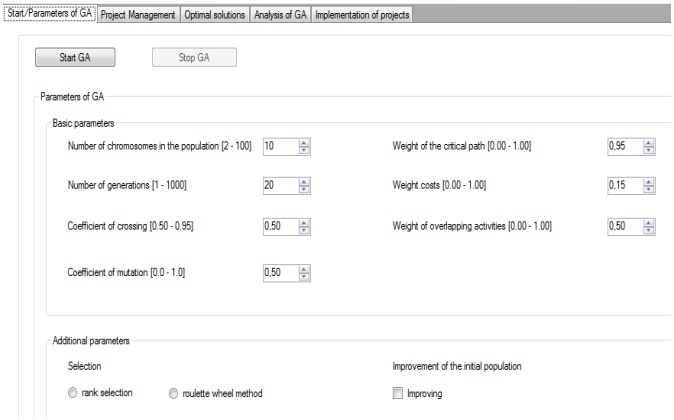


Fig. 2. IGAPM Interface

B Algorithm Pseudo code

Start
size = size_generation, i = 0
Create a population *P (i)* based on input data: *Knowledge Employees* P_V.
Evaluation of the population *P (i),* using parameters u_{cr}, u_{cost} and u_{pr}.
Repeat while (*i <size_generation*) or (interrupted by the user)
Selection (*sel, P (i)*); result *parent_1* and *parent_2*.
Crossover (*parent_1, parent_2*); result *child_1* and *child_2*
rnd = random number in the range 0 to 100.
if (*rnd > Pm*) **Mutation** (*child_1, mut probability*) end if
Evaluation *child_1 child_2*, using parameters u_{cr}, u_{cost} and u_{pr}.
Addition *child_1 and child_2* in *P (i).*
i = i + 1;
Repeat
End
Output at - optimal solution R

C. Chromosome encoding (matrix presentation)

IGAPM proposes matrix presentation (M) of chromosome. The value of m_{ij} gene is represented by one of eight values [1], approximately evenly distributed in the interval [0,1]. So each gene is described by three bits. The meanings of the bits are examined from left to right. The first bit in the unit ie 100 presented positive value 0.56 in the second bit in the unit ie 010 is positive 0.28, third in a bit that has value 001 is rounded to 0.16. If one bit is zero, therefore the value that represents is zero. The binary gene 000 is set to 0 and a gene with a record 111 is set to 1. Each gene is able to adopt the following actual values: 0, 0.16, 0.28, 0.44, 0.56, 0.72, 0.84 and 1. These values are approximate percentage of employee's project part-time (performance of Ei). Akj activity and project managers are responsible for employee's supervising in the organization. On Fig. 3 it is shown a sample format of chromosome M.

TABLE III
CHROMOSOME EXAMPLE

M	Specification	Programming	Architecture	Interface	Testing
I.Ivanov	0,44	0,84	0,28	0,84	0,84
G.Vasilev	0	0,16	0,84	0	1
G.Kanaheva	0,16	0,28	0,44	0	0,16
V.Georgiev	0,56	0	0,56	1	0,56
E.Avramova	0,28	1	0	0,16	0,28

The size of the chromosome can be represented by size | Ak | * | EW | * 3. IGAPM used chromosome composed of genes with a record length of size three. Binary record length 3 provides sufficient values to describe the percentage of employee's dedication. The binary format allows diverse describe of employee's project part-time. The bit increasing will significantly increase the solving time.

D. Initial Population, Selection, Crossover and Mutation

Creating the initial population:
Start
Size_population = population size
Establish population *P(0)* by the number of chromosomes in the population size_population (set by the user interface)
while (*size_population!* = 0) repeat
Create new chromosome corresponding to any restrictions
Normalization of chromosome
Calculation of fitness function for chromosome Oc_M
if (one or more activities have assigned staff)
$Oc_M = Oc_M * 100$
end if
The new chromosome is added to *P(0)*
end repeat
End
Output: population *P(0)*

The main purpose of selection is to obtain suitable parents for crossover, to get children with a good fitness function. IGAPM implementations are realized two methods of selection, determined by the user interface. Both methods select two parents for crossover. The methods of selection are: rank selection and method of "roulette."

For crossover IGAPM uses (two parents and two children) one - positional crossover (k = 1) that supports the rules for chromosome's composing. The probability of crossover is determined by the user interface.

IGAPM uses a mutation in Invert bit (Flip Bit). In random it turns one bit of the gene in the chromosome. The mutation is applied in random order on the chromosome and is applied for each gene from the choose line. The mutation probability (Pm) is determined by the user interface.

The last step of each generation is performing update of the population. IGAPM uses method for removing the poorest chromosomes. To the next population of children O(I) only good individuals are transferred. The chromosomes with poor fitness function are removed [3].

The program is completed by two ways [2]. One is when the algorithm reaches the maximum number of generations, set by the user and the second way is forcibly stopped by the user.

IV. TESTING AND RESULTS

There are a variety of methods for evaluating and experimenting with the Genetic Algorithms. There is no uniform methodology for testing and evaluation. Most studies include subjective evaluations based mainly on tasks that they solve. The standard approach exists. It analyzes and evaluates the Genetic Algorithm, changing the following parameters: number of generations, number of chromosomes in the population, crossover and mutation probability, type of selection and improvement of the initial population.

The behavior of the GA is defined on Fig. 5 and Fig.6. For finite number starts with a constant configuration of genetic parameters and genetic operators, the variations of the values of fitness function are relatively constant.

The analysis is based on the different number of GA generations. Experiments were performed with common parameters: number of chromosomes in the population = 10, crossover rate Pc = 0.65, mutation rate Pm = 0.55, weight of the critical path $u_{cr} = 0.95$, weight cost $u_{cost} = 0.15$ and weight of overlapping activities upr = 0.5. Additional parameter is the type of selection - by rank.

TABLE IV
RESULTS- NUMBER GENERATIONS/ FITNESS FUNCTION

Generations	Best fitness	Worst fitness	Average fitness	Project cost
20	17,35	23,75	19,157	1058
150	17,39	26,1	17,801	1065
400	**17,11**	27,57	17,262	1018
600	**17,51**	23,52	19,03	**1009**
1000	17,25	25,05	17,44	**1041**

The user interface allows for changing the number of generations in the interval [1, 1000]. The number of generations is one of the important factors defining the duration of algorithm.

Table IV shows that the increasing the number of generation has limit. The great number of generation couldn't improve a good value of fitness function. It depends on the specificity of current problem (activities' number, number of employees and etc.)

Fig. 3 shows the actual duration as result of IGAPM for 20 generations. The project duration is less than the initial (given duration). The project cost is 1058.

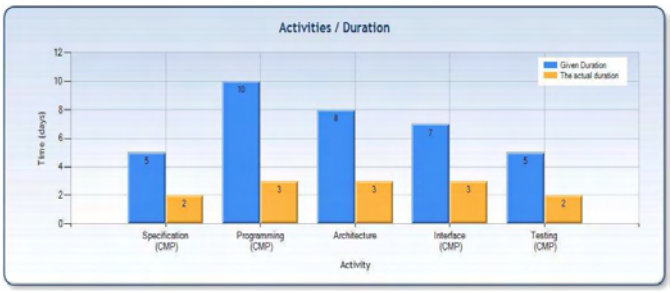


Fig. 3. Given Duration/ The Actual Duration (20 generations)

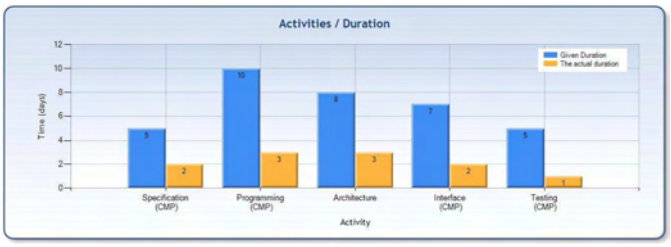


Fig. 4. Given Duration/ The Actual Duration (600 generations)

Fig. 4 is the graphical view of the best result of Tabl. 4. The project cost is 1009.

The evolution process of genetic algorithm is shown on charts Fig.5 and Fig. 6. Genetic algorithm does not allow increasing of the fitness function. It eliminates bad chromosomes with higher fitness function. This ensures a reduction of the fitness function over time.

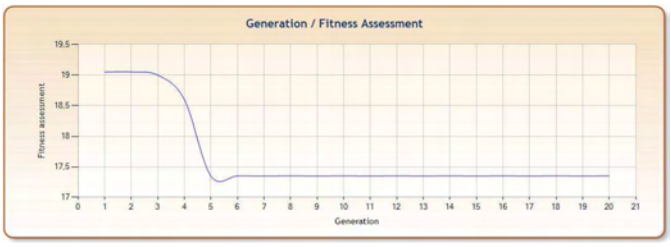


Fig. 5. Generations/ Fitness function (20 generations)

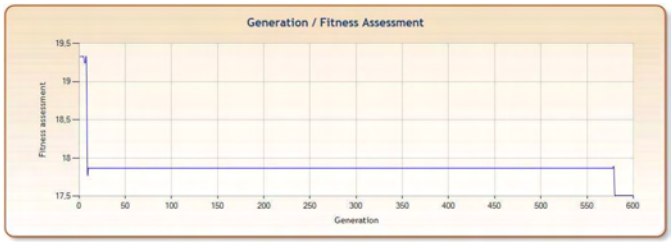


Fig. 6. Generations/ Fitness function (600 generations)

The Figures show that regardless of the number of generations the function finds its optimal solution for a small number of generations and their further increase is not necessary. The role of mutation or other factors is important in order to escape the algorithm from local minimum [Fig. 6].

Standard approach to research and evaluate the genetic algorithm does not exist. There is no exact number of algorithm executions. The optimal result depends on the specific task, on use of genetic operators and genetic parameters and on the number of generations.

The complexity of the algorithm can be calculated by considering the worst case |Ak| = |Ew|. If n = |Ak| therefore chromosome has n^2 genes. Any operation performed on the chromosome was assumed to be performed per unit time. If you refer to s_0 population size and number of generations is *i*. Complexity is Eq.5.

$$O\left(\sum_i s_i * n^2\right) \quad (5)$$

V. Conclusion

Using Genetic Algorithm (GA) for resolving Project Management Problem is one of the advanced applications of heuristic algorithms for automation. The GA is a good decision for resolving the conflicts (resources and costs, project duration and employee's knowledge).

The transformation of the actual requirements into chromosomes is a hard phase and the important factor for GA success. It has observed that the optimal results can be achieved if the GA is trained in further.

The future work involves a testing of the algorithm with various instances to improve the dominance of the project resources on genetic algorithm evolution process.

References

[1] E. Alba, J. Francisco Chicano, Software project management with Gas, Information Sciences 177, An International Journal, pp. 2380-2401, 2007

[2] S. Barthelemy, Principe general des algorithmes genetiques, www.barth.netliberte.org, 2000

[3] H. Sonke, A Self-Adapting Genetic Algorithm for Project Scheduling under Resource Constraints, Naval Research Logistics, John Wiley&Sons Inc, pp49:433-448, 2002

[4] http://www.pmipr.org/html/Presentations/ The Critical Path Method.pdf

Railway Infrastructure Maintenance Efficiency Improvement by Using Tablet PCs

Slobodan Mitrović[1], Svetlana Čičević[2], Slađana Janković[3], Norbert Pavlović[4], Slaviša Aćimović[5], Snežana Mladenović[6], Sanjin Milinković[7]

Abstract – **State of infrastructure of railway facilities affects safety and efficiency of railway traffic. Bearing this in mind, the usage of Tablet PC in wireless outdoor environment, as a support of infrastructure investigation activities of maintenance stuff is considered in this paper. The result of the investigation is recorded through data update process in the corresponding portal that is a part of the model for integration of traffic information systems in a cloud computing technological environment of Serbian Railways.**

Keywords – **Tablet PC, cloud computing, wireless, railways maintenance.**

I. Introduction

During railroad infrastructure inspection and/or investigation process Serbian railways personnel often spend a significant amount of time gathering and distributing information on paper documents.

Further, station stuff was about to perform tasks in several steps in order to convert paper-based data entry to electronic forms. Hence the necessity for technological improvement of this process arises.

Bearing in mind that the tablet PC provides a platform that combines data entry and retrieval with advanced communication and collaboration capabilities, in this paper the usage of such kind of device is considered. In this case the proposed solution is data updating process within the corresponding portal that is a part of the model for integration of traffic information systems in a cloud computing technological environment of Serbian Railways. This possibility is tested with a web interface that provides quick and easy way to access and update data, in order to provide the maximum in new mobile computing functionality that users can learn quickly with a minimum of training. Besides, the usage of such class of devices and personnel education for system and information management is not cost nor time consuming comparing to benefits.

[1]Slobodan Mitrović is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: s.mitrovic@sf.bg.ac.rs.

[2]Svetlana Čičević is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: s.cicevic@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]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.

[6]Snežana Mladenović is with the Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia, Email: s.mladenovic@sf.bg.ac.rs.

[7]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. Proposed Infrastructure Model

Use of tablet PCs as a standard tool in railway infrastructure investigation activities requires presence of wireless infrastructure within railway stations and other facilities. Thus, standard WiFi outdoor engineering principles should be applied for facilities which lack this kind of equipment. In those environmental conditions the centrally controlled wireless mesh solution based on Lightweight Access Point Protocol (LWAPP) [1] is proposed (Fig.1). Stationary infrastructure could be created using Wireless LAN controllers and outdoor Root/Mesh Access Points (RAPs/MAPs). Position of RAP (WLC AP) and MAPs relies on the railway station/facility configuration and must be considered evenly, because a site survey reveals issues such as interference, Fresnel zone, or logistics problems.

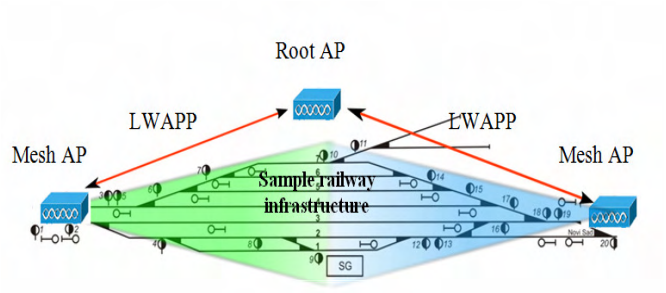


Fig. 1. Example of LWAPP based AP structure on a railway yard

The typical rail classification yard characteristics are [2]:

- 250 + acres of area covered
- 2+ linear kilometers end to end length
- Metropolitan and rural area environments
- Limited road access within the yard
- Very high levels of Radio Frequency Interference

Rail yard WiFi implementations often fail due to [2]:

1. Co-channel interference
2. Excessive Attenuation
3. Contention loss (too many clients converging on a single AP).
4. Inaccurate Signal strength mapping.

So, our experiences are in accordance with the findings of other researchers [2]:

1. Successful WLAN design must incorporate RF engineering considerations such as attenuation loss in long antenna cable runs.
2. Co-channel interference will often occur due to the presence of multiple and rogue Access Points competing for bandwidth.
3. Existing tools for signal strength mapping are difficult to apply to large outdoor environments.
4. Security and encryption are essential requirements for most applications.

Security and encryption requirements are crucial for this proposed model. It has to be taken into account that different models of tablet PC with different platforms (Windows, Android, IOS, etc) could be the choice of one Railway Company. Thus, we need security and encryption method that will be platform independent as much as possible. In this particular case, the choice is on IEEE 802.1x standard [3], with use of RADIUS server (IETF RFC2865) [4], AAA and LDAP protocol [5] (Fig.2).

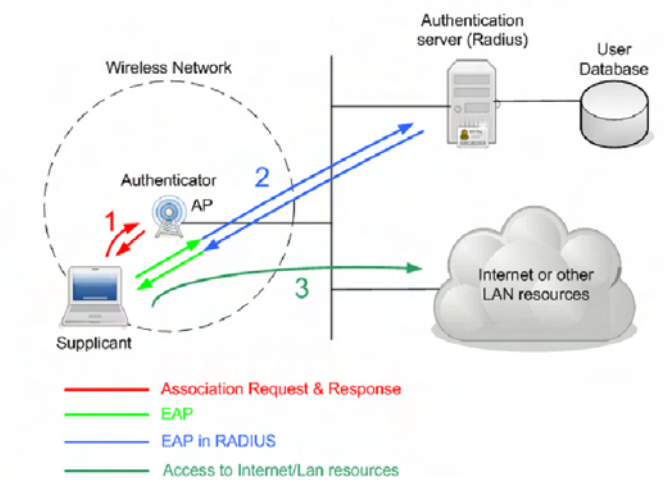


Fig. 2. Use of IEEE 802.1x standard

During the authentication process, only the client EAP messages could reach the NAS (e.g. WLC AP), which are forwarded to the appropriate RADIUS server. If the authentication is completed successfully, the user is allowed to access other network resources (such as DHCP, Internet).

User credentials protection is accomplished by using EAP-TTLS (Tunneled Transport Layer Security). These protocols are based on PKI (Public Key Infrastructure), which allows the client to authenticate communicating server, and also to send credentials through the protected tunnel. After connecting, client is ready to use Cloud resources.

III. USING THE CLOUD RESOURCES

The proposed integration model of the Serbian railways traffic Information System [6] involves the Community cloud delivery model, which means that only those computers which belong to Serbian Railways IP range should have right of access to a hosted service. This model involves the development of the classic n-tier application with the following layers:

- Database layer – SQL Azure database,
- Data access layer – Windows Azure Hosted Service, ADO.NET Entity Framework,
- Presentation layer – ASP.NET application.

In our case study Windows Azure application is implemented as an ASP.NET *Web role* application RAILROAD STATION INFRASTRUCTURE (RSI). *Web role* supports representation of user interface through IIS (Internet Information Services). RSI uses the ADO.NET Entity Framework - ADO.NET in a set of technologies that support the development of application-oriented data. *EntityDataSource* control enabled us to connect objects on a Web page with the data in our Entity Data Model. RSI Web client application allows users to view and/or update the contents of Azure SQL database Serbian railroad infrastructure. The main menu consists of following menus: stations, switches, lines and objects (Fig 3).

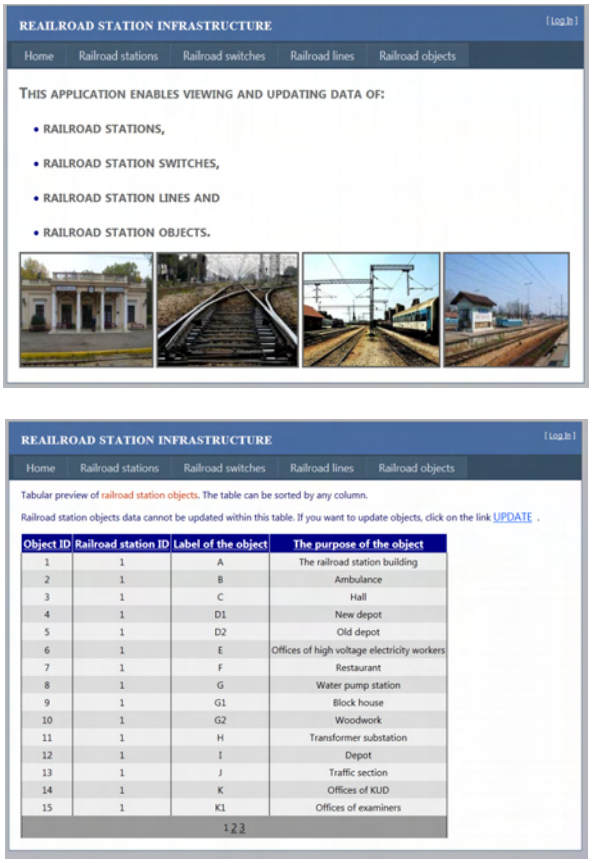


Fig. 3. RSI Portal, main menu and objects data table

IV. TESTING PHASE

The usage of tablet PC has been tested in real conditions. For testing purposes, a secured 802.1x environment has been created for access to the Cloud IS. The 8.4" tablet PC Prestigio MultiPad PMP3384B has been chosen, because the size of its display was estimated as to be suitable for optimal operation from the ergonomics point of view. This tablet works in the Android operating system, which has been successfully configured to access the Cloud IS (Fig. 4). Tablets that work in MS Windows operating system were not selected for this test because, according to previous experience, we assumed that they will work without problems. On the other hand, various tablet PC models based on the Android OS are available at prices that are affordable for purchasing.

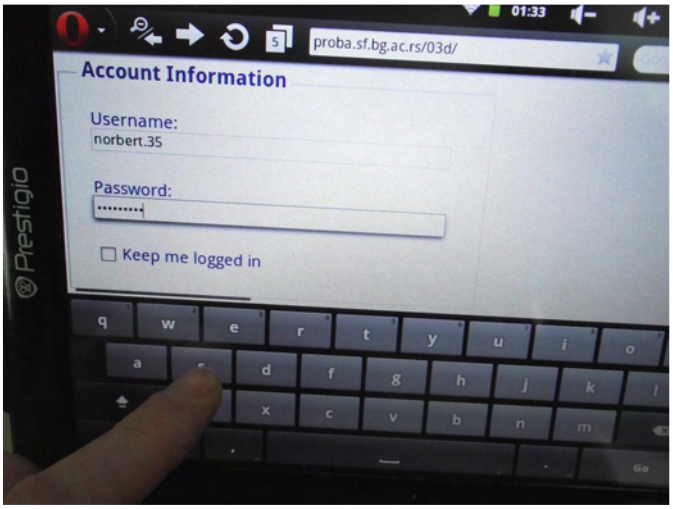


Fig 4. Cloud IS login sequence

Test web application was set up in the way that it contained two frames, one for railway station infrastructure schematic diagrams and another for tabular data (loaded from the Cloud) or satellite maps. The purpose of those satellite maps was to help railway personnel in finding objects from diagrams more easily (e.g. to perform the inspection tasks on site more successfully) (Fig. 5).

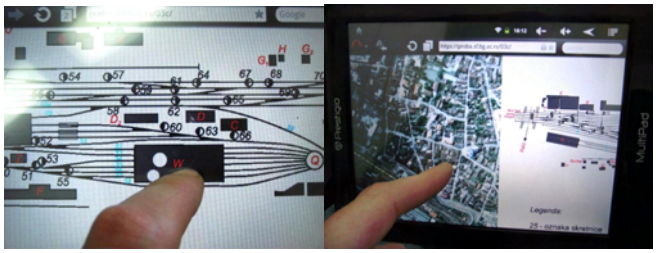


Fig 5. Usage of schematic diagram frame and satellite maps

On the other hand, the idea was to show whether the proposed web interface architecture was suitable for operation within certain type of web browser.

On the very beginning of test, it was encountered that the default Android web client (html reader), did not met required criteria. Therefore, The Opera mini browser was additionally installed, providing the satisfied level of functionality (Fig. 6).

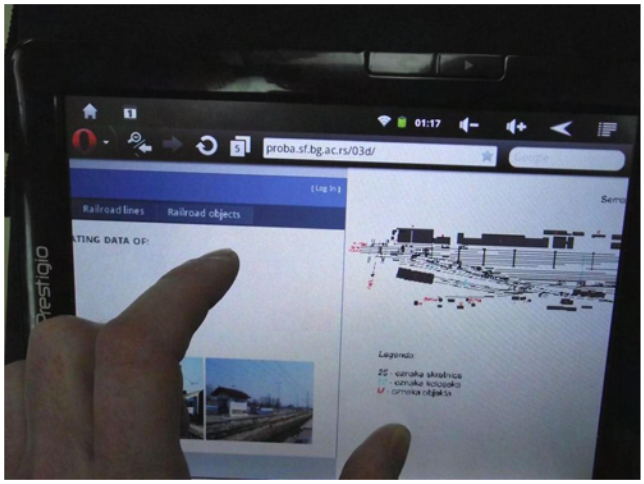


Fig 6. Frame structure of web application

Upon completion of inspection tasks the participants updated tabular data which were previously investigated (Fig. 7).

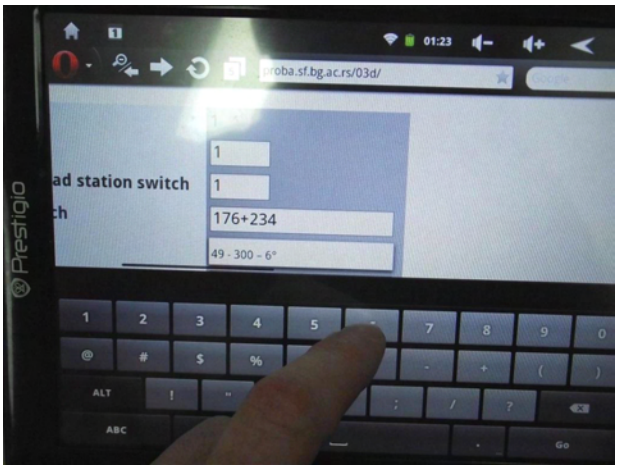


Fig 7. Updating tabular data

In addition the parallel testing with use of traditional forms of documentation (on paper) has been performed, but those are is beyond the scope of this paper.

The results of this experiment showed that the performances of inspection staff (represented by the time it took to perform tasks) were satisfactory. Furthermore, the number of errors that has been made was insignificant.

It was observed that the time needed for initial training in work with the Tablet PC device was very short, averaging less than 10 minutes. The participants were very satisfied working with this type of device as well as with the application.

V. CONCLUSIONS

The usage of Tablet PCs with the scope of railways infrastructure investigation process brings many advantages, comparing with use of traditional (paper) form of documentation. However, there are several types of difficulties in implementation such model of technological improvement. First, there are many issues that have to be considered during WLAN engineering process. Secondly, proper security measures have to be implemented in order to protect data that has to reviewed and updated.

Forasmuch as, there are several tasks that have to be accomplished in our future work.

- Railway infrastructure could be characterized as industrial environment and a part of our future work will be related to the development of general WLAN implementation model that could be particularly applied on Serbian Railways facilities.
- Some security measures have already been implemented. For example, RADIUS and LDAP server, as well as WLC APs are positioned to IP locations, defined by Serbian Railways Intranet (SRI) IP plan, so they are protected with firewall system from external attacks. Network segmentation to different VLANs, are also followed with standard security measures, according to facility position and importance within SRI. However, development of some security procedures related to usage of Tablet PCs and personal access to Cloud database, as well as accuracy of data entry and/or updating is highly recommended.
- As part of wider research, which is however, beyond the scope of this paper, a series of experiments have been conducted testing the possibility of applying tablet PCs with different types of touch screens that would be most suitable for usage in the above mentioned environment.

On the other hand, working with this kind of tablet PCs is helpful in making the state of information system more accurate. This leads to proper decision process, gaining the safety and efficiency of railway traffic, especially when the whole traffic picture is distributed to all important management elements through the Cloud environment.

ACKNOWLEDGEMENT

This work is partially supported by the Ministry of the Science and Technological development of the Republic of Serbia under No. 036012, No. 036022, and No. 036006.

.

REFERENCES

[1] Calhoun, et al. „Lightweight Access Point Protocol", RFC 5412, 2010.

[2] WHITE PAPER: Lessons learned: Making WiFi work in outdoor industrial environments, EF&I Services Corp, retrieved from http://www.eficorp.com/docs/whitepapers/OutdoorWiFi.pdf, 15.01.2012.

[3] IEEE Standard For Local And Metropolitan Area Networks-Port-Based Network Access Control, IEEE 802.1X-2010, The Institute of Electrical and Electronics Engineers, Inc. 3 Park Avenue, New York, NY 10016-5997, USA, 2010.

[4] Rigney, et al. „Remote Authentication Dial In User Service (RADIUS) ", IETF RFC 2865, 2000.

[5] J. Hodges, R. Morgan: „Lightweight Directory Access Protocol (v3): Technical Specification", IETF RFC 3377, 2002.

[6] S. Janković, et al, „A Model for Integration of Railway Information Systems Based on Cloud Computing Technology", ICEST 2011, Conference Proceedings, Volume III, pp.833-836, Niš, 2011.

Intelligent learning system for High education

Aleksandar Kotevski[1] and Gjorgi Mikarovski[2]

Abstract – **The purpose of this paper is proposing model for intelligent system for learning in high education, by giving suggestions to users. The main goal is to be found and present the most useful and practice text content to student from a large set of data, which are stored in database. In other words, this model is going to present adaptive learning system, so throughout the interaction with the students it's going to adapt to their needs.**

The most appropriate content that will be presented to students will be found in using phrase searching, under the following criteria: the most visited page from student, student's keyword in his user profile, pattern relations and the words and phrases that are often used from users while searching content.

Student interaction is the most important part for this model. The user's responses are the most crucial factor that affecting the increasing intelligence of the learning system.

Keywords – ***Data mining, intelligent, learning system, patterns, phrases***

I. WHAT IS DATA MINING

Data mining is an important data analysis methodology that has been successfully employed in many domains, and which has become especially popular after the World Wide Web made large volumes of data on many topics widely available. Data mining is an important paradigm for educational assessment. The usual assumption is that mining is performed after educational activity with that activity having been designed without regard for the mining process [1].

Gartner Inc.'s definition of data mining is the most comprehensive: "…the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories and by using pattern recognition technologies, as well as statistical and mathematical techniques".

It uses a combination of an explicit knowledge base, sophisticated analytical skills, and domain knowledge to uncover hidden trends and patterns. These trends and patterns form the basis of predictive models that enable analysts to produce new observations from existing data. Speaking about intelligent systems we stress that these systems apply techniques from the field of Artificial Intelligence (AI) to provide broader and better support for the users of Web-based educational systems. [2]

From educational aspect, data mining can be used as tools for better allocating resources, content and useful links that are related with student needs, and also improving the effectiveness of learning systems. Because almost every learning system has large volumes of data, data mining can be used to discover hidden patterns and relation that are very helpful when system has to make decision for next content that should be present to the student. By collecting information on a particular student's performance as well as other cognitive and no cognitive variables, the software can make inferences about strengths and weaknesses, and can suggest additional work [3]. Since web-based educational systems are capable of collecting vast amounts of student profile data, data mining and knowledge discovery techniques can be applied to find interesting relationships between attributes of students, assessments, and the solution strategies adopted by students [4].

It means that the main goal of intelligent learning system is to recognize uncovered hidden patterns and the relation between learning content.

II. RELATED WORK

Using system for online learning has more advantages, but also and some disadvantages. Because each learning system has number of courses and learning content, large bodies of text, images and multimedia materials, user will waste the time while they are searching in large set of information and found something useful for them. Some recent works addressed this problem. One approach supports the analysis of large bodies of texts by interaction techniques together with a meaningful visualization of the text annotations. For example Compus [5] supports the process of finding patterns and exceptions in a corpus of historical document by visualizing the XML tag annotations. Another approach is to use data-mining algorithms integrated with visual interfaces so that non-specialists can derive benefit from these algorithms [6]. Data mining is an emerging methodology used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students [7]. Galit has completed a case study that uses student's data to analyze their learning behavior to predict the results and to warn students at risk before their final exams [8]. Han and Kamber explained that k-means is a well-known clustering algorithm tends to uncover relations among variables already presented in dataset. Erdogan and Timor used educational data mining to identify and enhance educational process which can improve their decision-making process [9].

III. PROPOSED MODEL

The goal of this paper is to propose model for intelligent system for learning in high education, by giving suggestions to users. It means that system will learn from students visiting history and will adapt to user's needs. In that content, learning system is going to propose learning content to registered

[1]Aleksandar Kotevski is with the Faculty of Law, Partizanska B.B. 7000, Bitola, Macedonia E-mail: aleksandar.kotevski@uklo.edu.mk.

[2]Gjorgi Mikarovski is with the Faculty of Technical Scientist of Bitola, Ivo Lola Ribar B.B.,7000,Bitola,Macedonia, Email:gjorgi.mikarovski@tfb.edu.mk.

students by following criteria: history visits, keyword in student's profile, searching words, pattern relations and student response.

On the other site, the learning content author manually fills-upthe phrases in field's title, META tags and keyword. Phrases included within the keyword and description META tags are valuable precursors to the contents of the document.

The titles, which are enclosed within the TITLE tag of the Web documents' headers, were identified as another useful source of phrase sets whenever the contents of the TITLE tag are delimited either by commas, semi-colons or vertical bars[10]. Each phrase within the TITLE tag can be considered as a potential phrase that could be used to define the target phrase.

```
<html>
<title>title page</title>
<meta name="keywords" content="word1, word2, word3, word4, …">
<body>
learning content
learning content
learning content
….
learning content
</body>
```

For more efficient searching, the system is going to search for matching phrases in title, META tags and learning content.

A. User history

The learning system will bookmark all pages that the user had been visiting, and also order of visiting. Visiting history may be getting as parameter while proposing next content. To enable this feature, we are going to use database table with following fields: preview-page(preview visit page), active-page (page that user is viewing), next-page (next page that user will view) and repeat(parameter that indicate how many times this order of page visiting was matched).

Let say user is visiting page B. Preview visit page was A. In that situation, system will search in table and find out each record that fulfills the condition: preview-page = A and active-page = B. Result set is two-dimensional array: next-page and indicator repeat. After query execution, user will get list for recommendation page for visiting, ordered by repeat indicator. It means that result with the highest repeat indicator will be to the top of list of suggestions.

```
$prev_page = $_SESSION[‘prev_page’];
$current_page = $_SESSION[‘current_page’];

$result = mysql_query("SELECT next_page FROM tbluser_histiry where preview_page = ‘”.$prev_page.”’ And active_page=”’. $current_page.”’ order by repeat desc");.
```

B. Keyword

Each user has ability to insert keywords which are important for searching content. User can insert keyword while registering to the system, or later in his profile settings page. When user has inserted some keywords, system will search in database and will show all learning content related to the keywords. Keyword repeating in the content will determine the order of proposed list. Let say we have keyword K, and there are few records A,B,C and D.

Keyword K is found in A 5 times, in B 3 times, in C 6 times and in D 10 times. So, proposed list will have following order:D, C, A, B.

```
$result_kewywords = mysql_query("SELECT keywords FROM tbluser_profile where user_id=’”.
$_SESSION[‘user_id’].”’");

while($row_kewywords = mysql_fetch_array($result_kewywords)){$list_keyword[]=$row_kewywords[‘keywords’];}

$result = mysql_query("SELECT * FROM tblcontent where title like ‘%”.$row_kewywords[‘keywords’].”%’ and kywords like ‘%”.$row_kewywords[‘keywords’].”%’ and content like ‘”.$row_kewywords[‘keywords’].”’");
```

C. Searching keywords

The proposed model will have the ability to store searching history of each user. In addition, each word and phrase that user was using in search form will be saved in database. The table has the following fields: word (unique), repeat. Let say table already has record with value intelligent (word field). If user inserts word intelligent in search form, then system will check in database if that word already exists in table. If yes, system will update only repeat value (increment for 1). Otherwise, system will insert new row in the table. While user searching through learning content, system updates list permanently. The list is sorted by the repeat field and is using while system decision for proposing learning content.

```
$result_kewywords = mysql_query("select keywords from tbluser_search_history where user_id =’”.
$_session[‘user_id’].”’");
while($row_kewywords = mysql_fetch_array($result_kewywords)){
$list_keyword[]=$row_kewywords[‘keywords’];
}
$result = mysql_query("select * from tblcontent where title like ‘%”.$row_kewywords[‘keywords’].”%’ and kywords like ‘%”.$row_kewywords[‘keywords’].”%’ and content like ‘”.$row_kewywords[‘keywords’].”’");
```

On the other words, the system will give suggestion to user through searching learning content by words and phrases from the list.

D. Related patterns

Each searching pattern has own rating value. Patterns with the highest rating value are often in use in searching process instead of patterns with lowest rating value. On the other site, each pattern has relation with other pattern. That's why this model proposes using kind of relationship mining. In that content, this model proposes using relationship mining for discovering relationships between patterns. This may take the form of attempting to find out which patterns are most strongly associated with a single pattern of particular interest, or may take the form of attempting to discover which relationships between any two patterns are strongest.

For this reason, the proposed model is going to use table for relation between searching patters plus field with synonyms for selected searching pattern.

E. User response

User response is the most important part of this model. Response will be the best direction for making learning system more intelligent. In that content,we'll require from user to fillquestioner with following questions:

- Does suggested page was adequate for user
- Specify next page that is related with selected page, but it isn't into the suggested list

User interaction is the most important part for this model. The user is factor that influence in system learning for user purposes. It means that after content showing, user has to rate if the content has helped to them. That rating tell to system does selected text pattern is useful for user or not.

Each user has own setting profile, so he can set up what factor has more priority while searching learning content. Let say user has following priority settings:

1) Searching keywords
2) User history
3) Keywords
4) User response

It means that system will recognize following array:

$sort_array = ('searching_kewywords','user_history','keywords','user_response')

Result set will be order by relevance:

```
select * from tblcontent where title like '%".$keyword."%' and kywords like '%".$keyword."%' and content like '".$keyword."'")
order by (
 (
case when '".$sort_array[0]." like '".$keyword."'
then 1
else 0
end
 ) + (
case when '".$sort_array[1]." like '".$keyword."'
then 1
else 0
end
 ) + (
case when '".$sort_array[2]." like '".$keyword."'
then 1
else 0
end
 ) + (
case when '".$sort_array[3]." like '".$keyword."'
then 1
else 0
end
 )
) desc
```

IV. Conclusion

For universities, data mining techniques could help to provide more personalized education, maximize educational system efficiency, and reduce the cost of education processes. It may guide us to increase student's retention rate, increase educational improvement ratio, and increase student's learning outcome.

The data collected from different applications require proper method of extracting knowledge from large repositories for better decision making. Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful information from large collections of data [11].

References

[1] Steven L. Tanimoto, University of Washington, Dept. of Computer Science and Engineering: Improving the Prospects for Educational Data Mining

[2] Adaptive and Intelligent Web-based Educational Systems, Peter Brusilovsky, School of Information Sciences, University of Pittsburgh, 135 North Bellefield Avenue, Pittsburgh, PA 15260, USA, ChristophPeylo, Software LogistikimArtland, Friedrichstr. 30, 49610 Quakenbruck, Germany

[3] Ducational measurement and intelligent systems, Valerie J. Shute, Diego Zapata-Rivera

[4] Data mining for a web-based educational system, BehrouzMinaei-Bidgoli

[5] Telecommunications, 2007. ConTel 2007. 9th International Conference, Vranic, M.; Pintar, D.; Skocir, Z.:The use of data mining in education environment

[6] Mining "Hidden Phrase" Definitions from the Web, Hung. V . Nguyen, P. Velamuru, D. Kolippakkam, H. Davulcu, H. Liu, Department of Computer Science and EngineeringArizona State University, T empe, AZ, 85287, USA

[7] Mining Educational Data to Analyze Student's Performance, Brijesh Kumar Baradwa, Saurabh Pal, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011

[8] Galit.et.al, "Examining online learning processes based on log files analysis: a case study". Research, Reflection and Innovations in Integrating ICT in Education 2007

[9] A data mining application in a student database. Journal of Aeronautic and Space Technologies July 2005 Volume 2 Number 2 (53-57)
[10] Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1999
[11] Church, K.W., and Helfman, J.I., Dotplot: A Program for Exploring Self-Similarity in Millions of Lines of Text and Code, In Proc. of the 24th Symposium on the Interface, Computing Science and Statistics V24, 58-67. 1992
[12] Fekete, J. and Dufournaud, N., Compus: visualization and analysis of structured documents for understanding social life in the 16th century. In Proc. of the Fifth ACM Conference on Digital Libraries, 47-55. 2000
[13] Jing Luan, Chief Planning and Research Officer, Cabrillo College Founder, Knowledge Discovery Laboratories: Data Mining Applications in Higher Education
[14] Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1999

Using Cloud Computing in e-learning

Gjorgi Mikarovski[1] and Aleksandar Kotevski[2]

Abstract – **Cloud computing is Internet-based computing, where data are stored online on one or more servers and provided to the computers and other devices on-demand. It is use to provide computer applications to users without the need for those users to install software on their computers. In other words, users can store files online and access them from anywhere and anytime using a device (computer or mobile device) connected to the internet. On the other site, e-learning has become so popular and has huge influence in educational process. But, it usually requires many software and hardware resources.**

Viewed from several aspects, using cloud computing in e-learning will contribute to lower costs and more efficient use of e-learning.

Keywords – **e-learning, cloud computing, education, remote, internet**

I. INTRODUCTION

The benefits of cloud computing are largely financial, according to panelists: The organization pays according to how much and how often they need services. Software and storage are hosted and supported on the servers of the cloud computing provider, so, educational institutions don't buy software only one person uses, invest in technologies that are quickly outdated, or spend hours and hours on technical support. Cloud computing also offers a wider range of software than would be practical to purchase individually.
Cloud Computing helps academia by:

- In the current tight funding situations, a low cost option to high end computing
- Offering an easy way to scale up and down based on their needs thereby conserving the tax payers money for more fruitful research a several fold increase in efficiency, thereby, helping scientists to get their results fast.

Cloud computing in education gives better choice and flexibility to education IT departments. The platform and applications you use can be on-premises, off-premises, or a combination of both, depending on your academic organization's needs. The advantages that come with cloud computing can help you resolve some of the common challenges you might have while supporting your education institution.

[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 Low, Partizanska bb 7000 Bitola Macedonia,
E-mail: aleksandar.kotevski@uklo.edu.mk

- Cost. You choose a subscription or, in some cases, a pay-as-you-go plan—whichever works best with your organization's business model.
- Flexibility. Scale your infrastructure to maximize investments. Cloud computing allows you to dynamically scale as demands fluctuate.
- Accessibility. Help make data and services publicly available without jeopardizing sensitive information.

II. UNIVERSITIES IMPLEMENT CLOUD COMPUTING

Cloud computing remarkably boosts the learning ability of the students. With this technology, learning approaches and strategies unheard of before, or, at the very least, thought to be undoable, are now being used on a large scale.

The University's normally uses the cloud to effectively implement collaborative learning approaches where the students are able to work alongside students from other locations in order to achieve a common goal.

So, how do universities implement cloud computing? How do they effectively integrate it to their systems? Cloud computing services are categorized to three: infrastructure, platform, and software.

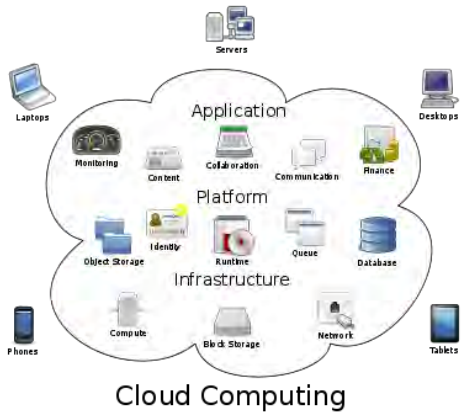


Fig.1 Cloud Computing infrastructure

1. *Infrastructure* - cloud can be used as a digital place where data, even servers, can be stored and protected. Indeed, cloud computing will allow university administrators to actually control much of their resources with more efficiency and with less cost

2. *Platform* - important application of cloud computing in the higher education sector is the PaaS, or Platform as Service. Basically, cloud computing will allow universities to use it as a platform where they will be able to access other services and more advanced and

more dedicated applications. As a matter of fact, PaaS will not merely allow one to access advanced services, it will also allow creation of unique services.

3. *Software as Service* - Basically, cloud computing allows users – in this case, universities and the learners in those universities – to actually utilize a wide range of applications and software online. As we all know, there are literally thousands of these useful applications in the internet, some of which are free while the others are not. At any rate, SaaS will allow users to access all of these.

This refers to a movement to turn computer terminals and notebooks into "client" machines that primarily (or only) execute applications running on servers somewhere out there on the Web. For example, instead of running Word from a notebook's hard drive, you'd run a copy of the program that lives on a remote server…and perhaps even save your documents there.

This approach has advantages:

- Software use is monitored and controlled.
- Software version control is simplified.
- Virus dangers are minimized.
- Source data and resulting files may be stored, managed, and protected centrally, behind server firewalls.
- Less advanced (and expensive) computers can be issued to employees.
- A lost computer is less likely to compromise company or customer data.

Files can be stored online using Cloud computing. Cloud computing is Internet-based computing, where information is stored online on a server and provided to the computers and other devices on-demand.

Some of the major services providing cloud computing are:

- Windows Live SkyDrive
- iCloud from Apple
- Box.net,
- Drop box

One of the best technological advancements to come to eLearning is cloud-computing, which will significantly streamline the learning process and infrastructure, making it easier on students, teachers, and administrators as they strive towards academic excellence.

Cloud-computer essentially allows all members of an academic community to store information in a central cloud location; basically, they don't keep separate files on their hard drives. Instead, the online program keeps them on a server to which all members have access. Users upload versions of the files to the shared server, and that server keeps the information secured based on what specific sharing settings each person uses. Because the information is located on a central computer, it actually creates many more opportunities for others to use that information. Fortunately, this has some great benefits for eLearning practices.

III. E-LEARNING PROTOTYPE APPLICATION

The traditional e-learning platforms consist of the learning management system, learning content management system, assessment and communication modules (especially forum and messaging). The third generation of e-learning platforms provides with advanced services such as online courses, tutorials and webinars. The education process in engineering means theory and practice, individual study, team projects or experimental work and involves laboratory equipment, simulation/emulation software packages and applications.

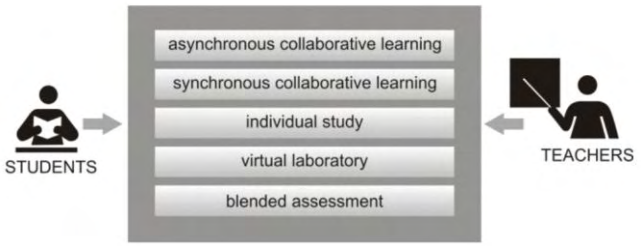


Fig.2 Blended Learning Model for Higher Education in Engineering.

The production of a prototype e-learning application affairs office was necessary in order to examine the functionalities of the services and the students in a real time. The structure of the application is separated into two larger parts introduced as: 1 part that provides the requirements of the students' affairs office and the services for European EKTS, and 2. part that provides the requirements of the teaching staff and e-learning services for the students.

Case 1: hosting prototype application on local servers

Requirements:

1 – Server
1 – Backup Server
1 – License Windows 2008 Server Enterprise with IIS 7
1 – License MSSQL
1 – Antivirus System
1 – Router
1 – Internet Line with static IP address

Main advantages: application is host on local server and response time from our institution is faster than cloud hosted application.

Main disadvantages: application response time from Internet is slower than cloud hosted application, too much costs and too much worries (servers, software license, software upgrade etc.)

Case 2: hosting prototype application on Cloud

Requirements:

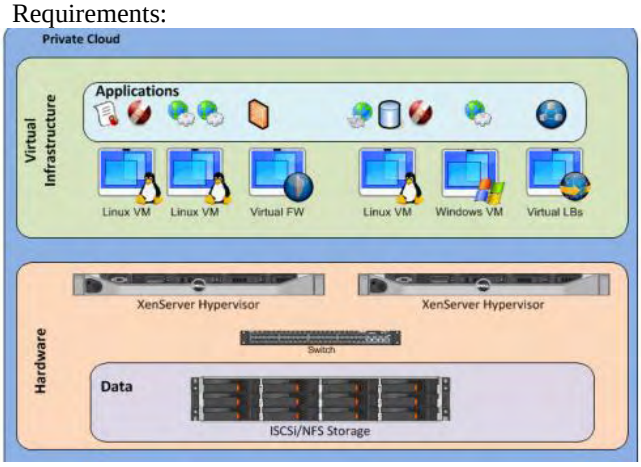
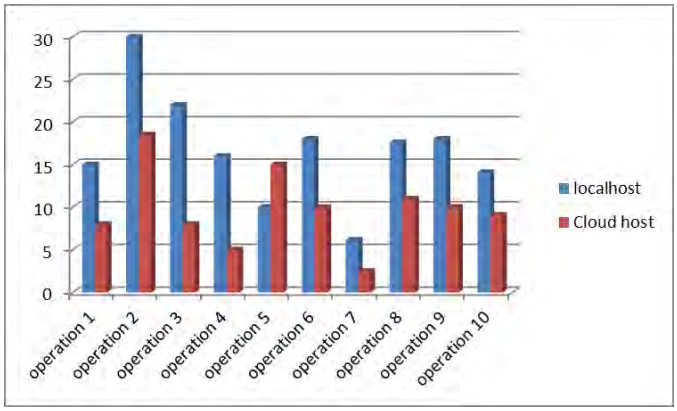


Fig.3 Cloud Setup.

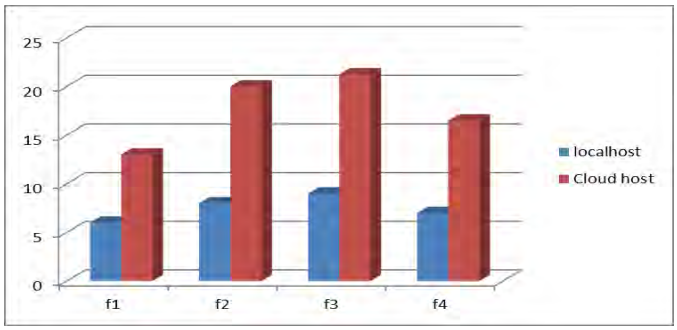
1 – Access on Go Daddy Hosting Cloud Servers

Main advantages: application has great response time from any location, no routers, no servers, no software updates, no software licenses, no viruses, coast benefits etc.



G1. Response time in (s) on same application host on "cloud" and "local host"

Main disadvantages: application has 180% lower response time from local hosted application.



G2. Response time in (s) on same application host on "cloud" and "local host"

Cloud hosting has desirable features including low up-front costs, elasticity of resources, and cost savings that result from economies of scale. Local host provides greater direct control over infrastructure than can be achieved when leasing shared infrastructure from the cloud. However, achieving the bents of cloud infrastructure by transferring infrastructure control to a third party needn't necessarily result in a net loss of security may also Benet from scale economies. In particular, cloud providers security measures with up-front costs that would be able in local hosting environments, amortizing these costs over myriad machines or tenants.

IV. Conclusion

One of the best technological advancements to come to eLearning is cloud-computing, which will significantly streamline the learning process and infrastructure, making it easier on students, teachers, and administrators as they strive towards academic excellence.

Cloud-computer essentially allows all members of an academic community to store information in a central cloud location; basically, they don't keep separate files on their hard drives. Instead, the online program keeps them on a server to which all members have access. Users upload versions of the files to the shared server, and that server keeps the information secured based on what specific sharing settings each person uses. Because the information is located on a central computer, it actually creates many more opportunities for others to use that information.

Acknowledgement

For the acknowledgement use the unnumbered section layout.

References

[1] Budnikas, G., Misevi˘ cien˙ e, R. (2010). Use of internet-based facilities in innovative IT course. E-Education: Science, Study and Busines.In: Proceedings of 3rd International Conference on Advanced Learning Tech - nologies and Applications, 138–139.

[2] Cloud computing in education (2010). UNESCO Institute for Information Technologies in Education. http://www.microsoft.com/education/solutions/cloudcomputing.aspx.

[3] Keir, J. (2011). Investigation into Google Apps and Microsoft's Live@Edu. http://ist.uwaterloo.ca/projects/GoogleAppsMicrosoftLive/charter.html.

[4] Karazinas, E., Kardzys, E., Matulaitis, E. (2010). KTU EMTC Moodle paslaugos. E-Education: Science, Study and Busines.In: Proceedings of 3rd International Conference on Advanced Learning Technologies and Applications, 107–110 (in Lithuanian).

[5] Sourya, B. (2011). How Can Cloud Computing Help In Education? http://www.cloudtweaks.com/2011/02/how-can-cloud-computing-help-in-education/#.

[6] Jones, Ch., Sclater, N. (2010). Learning in an age of digital networks. International Preservation News, 55, 6–10. http://oro.open.ac.uk/24116/2/learning_in_an_age.pdf.

An Approach to Define Interfaces for Mobile Telemetry

Ivaylo Atanasov[1], Ventsislav Trifonov[1], Evelina Pencheva[1]

***Abstract* – The mobile telemetry has a lot of applications ranging from green technologies through telemedicine to security. The paper studies the generic functions of mobile telemetry to define an abstraction and to design Application Programming Interfaces (APIs). The mobile telemetry APIs for acquisition of measurements data are implementation protocol independent. The paper proposes Web Services access to measurement data which allows 3rd party applications to fetch data in secured and managed manner. Web Services access is also proposed for mobile telemetry system entities that provide data in the repository.**

***Keywords* – Application Programming Interfaces, Web Services, Database abstraction layer.**

I. INTRODUCTION

Mobile telemetry uses mobile data for remote measurement and reporting. The range of applications is from green technologies and security applications to telemedicine. The requirements to mobile telemetry protocols are currently subject of intensive research studies. Different aspects of telemetry functions and their implementations in mobile agents are discussed in [1], [2] and [3]. The main issues in the design of a mobile telemetry protocol concern the reliable data transfer and data security. Despite of the differences in application domains, any mobile telemetry protocol requires generic communication functions. We put the accent on the comparison between the generic functions of existing solutions. Townsend, Abawajy and Kim [4] present an approach to monitoring moving objects by utilizing mobile technologies and Short Message Service (SMS) to transmit data from a patient to the doctor. The SMS is pervasive messaging technology but it does not provide delivery guarantees and standardized delivery receipts. That is why the authors extend SMS with a message validity period, failover options and redundancies. However the proposed solution is not suitable for real-time applications because of unpredictable delays introduced by the SMS Centre. Cibuk and Balik [5] demonstrate a model-based solution approach for mobile telemetry. Their model is based on the communication medium of the Internet and transport layer in TCP/IP model. TCP is not suitable for real-time applications because it provides a reliable message transfer at the cost of retransmissions which implicitly increase the transport delay. Further, as application data may be sensitive, it needs to be transferred in a secure way, but the proposed protocol does not include any security services. In this paper, we suggest an approach to design Application Programming Interfaces (APIs) for mobile telemetry. The APIs define an abstraction of mobile telemetry functions and are independent of the implementation protocol. The API may be used in different mobile telemetry application domain.

Another aspect of mobile telemetry is the external access to measurements data. Traditionally, all database vendors provide their own interface tailored to their products which leaves it to the application developer to implement the code for the database interfaces. Information abstraction layers reduce the amount of work by providing a consistent API to the developer and hide the database specifics behind this interface as much as possible. An effective information encapsulation layer provides benefits such as coupling between the object schema and data schema, an ability to evolve either one, provisioning of a common place to implement data-oriented business rules, and increasing application performance [6]. There are a number of approaches that industries have employed when they want to expose their databases as Web Services. Abstraction layers with different interfaces in numerous programming languages such as Drupal 7 Database API, Java API for XML Web Services, SOAP with Attachments API for Java, Java DB, Java Data Objects, and NewsKnowledge Web Service API provide a standard, vendor-agnostic abstraction layer for accessing database servers. In parallel, intensive research is conducted on applying advanced technologies for the information abstraction layer. Jayasinghe [7] suggests an approach for exposing a database as a Web Service using Axis2. Yang, Zhang, and Zhao [8] analyze database connection mechanism of a WEB application system based on a three-tier architecture, and the corresponding relationship between main steps and auto-generated code used to achieve database access in the Dreamweaver development environment. Qu, Feng and Sun [9] analyze the difficulty of the united access of the distributed and heterogeneous biological information database and present an approach based on web service and multi-agent, which adds intelligence to the united access. Sellis, Skoutas and Staikos [10] describe efforts towards database integration and interoperability, based on Web Services and ontologies. In this paper, we define Web Services interfaces that provide an abstraction of the access to database storing mobile telemetry measurements.

The paper is structured as follows. Section II describes generic architecture for mobile telemetry and identifies common functions. Section III presents an approach to the definition of APIs for mobile telemetry. Section IV describes an approach to design Web Services interfaces for open but secured access to measurements database. Section V presents an example of Web Services usage. The conclusion summarizes the contributions.

[1]The authors are with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mails: iia@tu-sofia.bg; vgt@tu-sofia.bg; enp@tu-sofia.bg

II. GENERIC ARCHITECTURE FOR MOBILE TELEMETRY

The mobile telemetry system is consisted of a central control unit that handles information gathering from multiple mobile sources and a set of mobile agents capable of domain data monitoring and measurement reporting.

The Control unit (CU) plays a central role in registration of mobile agents. It is responsible for management of the measuring and reporting. The CU has to store data for mobile agents and operation-related data of the telemetry system. The CU needs to provide interfaces to a measurement data repository. The mobile agent (MA) belongs to the class of embedded devices, equipped with sensor(s), positioning module, data transmission module and power supply module. The main requirement to the MA is to operate using as low energy as possible.

The access to measurements database may be provided by Web Services. Web Services is pervasive technology that enables building of network of services. The technology allows flexible implementation of more sophisticated applications by combining the existing Web Services with the Parlay X ones.

If Web Services are used for open access to measurements data, the owner of the measurements repository has to deploy a gateway that exposes APIs toward 3rd party applications and Database Management System (DBMS) operations toward the measurements repository as shown in Fig.1. Internal Web Services API may be used by the CU to supply data to the gateway.

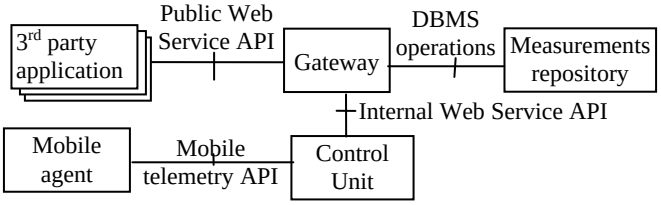


Fig.2 Functional entities involved in open access to mobile telemetry data

Common functions for mobile telemetry applications include the following: MA registration, communication security, operational mode management, CU notifications to MA, measurements reports, and error handling. The MA may be configured to operate in different operation modes. Periodic monitoring and reporting requires the MA to make measurement and send reports periodically. Triggered reporting operation mode requires monitoring for certain criteria and submitting reports on their occurrence. The third mode of operation is reports upon request. The CU induced reporting requires MA to perform measurements and report on demand.

III. MOBILE TELEMETRY API

The proposed approach to design APIs for mobile telemetry defines two interface packages: one supported by the mobile agent and another one supported by the central unit. The generic interface **MobileAgent** is inherited by interfaces which provide functions for the MA client (**MaClient**) and interfaces which provide functions for the MA server (**MaServer**). The interfaces of the MA client have prefix 'MaC', while the interfaces of the MA server have prefix 'MaS'. The generic interface CentralUnit is inherited by interfaces which provide functions for the CU client (**CuClient**) and interfaces which provide functions for the CU server (**CuServer**). The interfaces of the CU client are named by prefix 'CuC', while the interfaces of the CU server are named by prefix 'CuS'. Fig.2 shows the structure of APIs for mobile telemetry.

The MA supports the following interfaces. The **MaSPriodicReport** interface is used by the CU to configure the MA operational mode for period measurements reporting. The interface provides methods for initiation, modification and termination of periodic reporting. The **MaSTriggered-Report** interface is used to configure the MA operational mode for triggered measurements reporting. The interface provides methods for initiation, modification and termination of reporting in case of occurrence of specific events. The **MaSOnDemandReport** interface is used by the CU to request a measurements report ad hoc. The **MaCMea-surementsReport** callback interface is used to receive results of measurements reporting. The **MaCRegistration** is callback interface used to receive results of the registration request. The **MaSAdministration** interface is used for administrative purposes. It provides methods which allow CU to request authentication, to send new temporary identities and to send notifications to the MA.

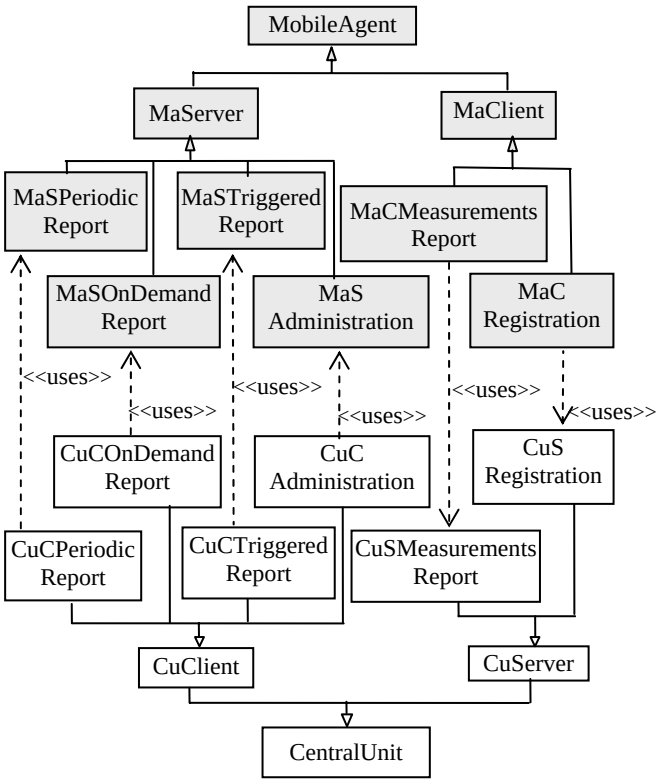


Fig.2 Mobile telemetry interface structure

The CU supports the following interfaces. The **CuCPeriodicReport, CuCTriggeredReport** and **CuCOnDemanReport** are callback interfaces. They provide methods which may be used to receive results of operation mode configuration requests. The **CuSMeasurementsReport** interface is used by the MA to report measurements periodically, triggered and on demand. The **CuSRegistration** interface is used by the MA to maintain its registration. The **CuCAdministration** is callback interface which is used to receive the results of the authentication requests, of new temporary identities allocation requests and of notification requests.

IV. Web Service Access to Measurements

Two Web Services may be defined to maintain consistent information flows between the CU and the gateway and between the gateway and 3rd party applications.

The **Monitoring Management** Web Service allows the CU to supply data to the gateway. It is internal for the mobile telemetry system and is not for public usage. The **Monitoring Management** interfaces are divided into two categories: implemented by the gateway and implemented by the CU. The interfaces implemented by the gateway are **MeasurementsManagement** interface, **Authentication** interface, and **CUNotificationManager** interface. The CU implements the **CUNotification** interface.

The **MeasurementsManagement** interface supports the following operations:

- **submitMeasurements()** operation is used by CU to submit measurements to the database;
- **modifyMeasurements()** operation is used to update previously submitted measurements data;
- **deleteMeasurements()** operation is used to delete previously stored measurements data;
- **queryMeasurements()** operation is used to query for measurements data. The query may include expressions that define the selection conditions.

To make the mobile telemetry system scalable the security related parameters may be stored at AAA server (not shown in Fig.1). The **Authentication** interface supports the following operation:

- **getAuthenticationData()** operation is used to request the authentication parameters related to given MA and stored at the AAA server. The parameters are required by CU to allow building the requests for authentication to the MAs.
- **updateAuthenticationData()** operation is used to store the new temporary identities of the CU and the MA in case of successful authentication.

The **CUNotificationManager** interface supports the following operations:

- **startCUNotification()** operation sets up the notifications in order to inform an authorized user through the administrative interfaces of the CU that measurements which met the defined criteria occurred. When such measurements are submitted, the user is notified about corresponding criteria.
- **endCUNotification()** operation terminates the previously installed triggers for notifications based on given set of criteria.

The **CUNotification** interface is implemented by the CU and supports the following operation:

- **notifyAlertingMeasurements()** operation allows to receive notifications when submitted measurements reach previously defined thresholds.

The **Measurements Access** Web Service is part of the public Web Service which allows 3rd party applications to access measurements data. The Web Service also provides subscription/notification interfaces which allow a 3rd party application to get a notification when particular thresholds in measurements data are reached.

The **DataAccess** interface is implemented by the gateway and it supports the following operation:

- **readMeasurements()** operation allows an authorized 3rd party application to read stored measurements data. The operation specifies the selection conditions.

The **AppNotificationManager** interface is implemented by the gateway and it supports the following operations:

- **startAppNotification()** operation is used by a 3rd party application to register its interest in receiving notifications about specific events in the measurements dataset.
- **endAppNotification()** operation is used by the 3rd party application to indicate that it is no longer interested in receiving notifications.

The **AppNotification** interface has to be implemented by the 3rd party application and supports the following operation:

- **notifyAppMeasurements()** operation allows to receive a notification after specific events defined by **startAppNotification()** operation occurred in the dataset.

The Web Services technology does not provide standardized means for authentication and authorization. Framework interfaces are defined as a part of **Measurements Access** Web Service. These interfaces provide authentication functionality for external access to measurements database. The 3rd party application authentication and authorization parameters may be stored at the AAA server.

The **GWFramework** interface is implemented by the gateway and it supports the following operation:

- **gwAuthentication()** operation is used by a 3rd party application to authenticate the gateway.

The **AppFramework** interface is implemented by the 3rd party application and it supports the following operation:

- **appAuthentication()** operation is used by the gateway to authenticate the 3rd party application.

The **ServiceAgreementManagement** interface is implemented by the 3rd party application and it supports the following operations:

- **signServiceAgreement()** operation is used by the gateway to request the 3rd party application to sign an agreement on the service;
- **terminateServiceAgreement()** operation is used by the gateway to terminate an agreement for the service.

Fig.3 shows the message sequence when a 3rd party application authenticates with the gateway.

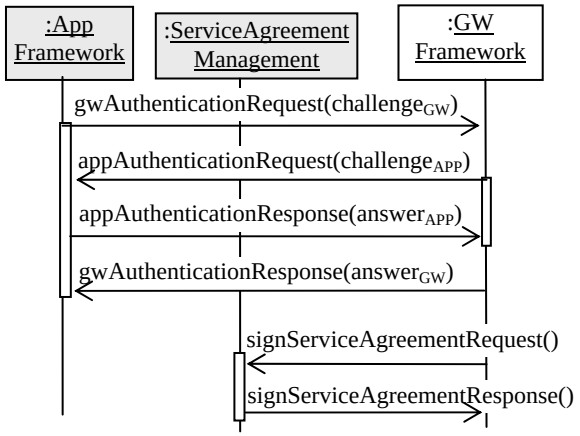


Fig.3 Third party application authentication and service agreement

V. USE CASE

An example application that provides ubiquitous access to local CO_2 emissions levels may use the service brokering function provided by the gateway and the Parlay X **Terminal Location** Web Service defined in 3GPP TS 29.199-9.

Parlay X Web Services are intended to provide simple interfaces for open access to functions in the public communication network. The Parlay X **Terminal Location** Web Service provides access to the positioning information through request or notification of a change in the location of a terminal or notification of terminal location on a periodic basis. The terminal location is expressed through a latitude, longitude, altitude and accuracy.

Fig.4 shows the sequence diagram for a hypothetical application "Can I breathe here?"

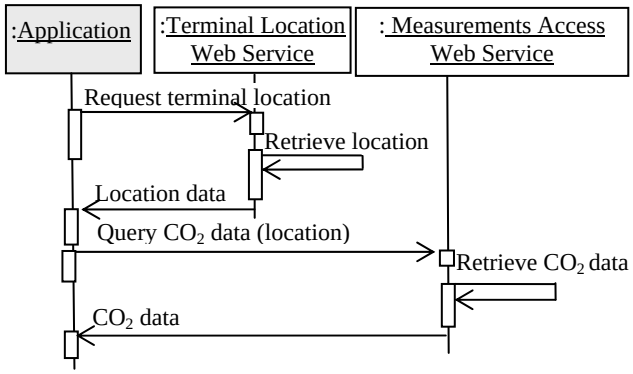


Fig.4 Hypothetical application "Can I breathe here?"

The "Can I breathe here?" application displays at the mobile subscriber terminal a map with subscriber's position and local CO_2 emission levels. It uses the Parlay X **Terminal Location** interfaces to query about terminal location and **Measurements Access** Web Service interfaces to access the CO_2 measurements data. For the application to determine the location of a mobile device, it provides a mobile subscriber's terminal address and desired accuracy, and then it receives the location for the terminal requested. When querying the CO_2 measurements database, the application provides the desired location of the measurements and receives information about current CO_2 levels at its location.

VI. CONCLUSION

The paper provides a new structural approach to definition of Application Programming Interfaces for mobile telemetry. Derived from generic functions the APIs provide an abstraction of mobile telemetry that is independent of the protocols used for implementation. The open access to measurements data allows 3rd party applications to access data in a secured manner. The proposed Web Services for open access to measurements data accelerate the development of new attractive applications and shorten the time to market. The suggested solution for internal access to measurements repository though Web Services interfaces provides an abstraction of DBMS operations which is an additional level of flexibility i.e. the changes of the storage technology will not impact on the overall function of the system. The future work will be focused on API design, implementation and test.

ACKNOWLEDGEMENT

The research is in the frame of Project DDBY02/13/2010 funded by NSF, Ministry of Education Youth Science, BG.

REFERENCES

[1] M. Alenazi, S. Gogi, D .Zang, E. Cetinkaya, J. Rohrer and J.Sterbenz. "ANTP Protocol Suite Software Implementation Architecture in Python", International Telemetry Conference ITC'2011, Conference Proceedings, pp.1-10, 2011.

[2] D. Forthoffer, M. Throop and B. Tilman. "Mobile Telemetry System", Patent Application Publication, Pub. No.US 2011/-166741 A1, 2011.

[3] B. Chi, J. Yao, S. Han, X. Xie, G. Li, and Z. Wang. "Low-Power Transceiver Analog Front-End Circuits for Bidirectional High Data Rate Wireless Telemetry in Medical Endoscopy Applications", IEEE Transactions on Biomedical Engineering, vol. 54, issue 7, pp. 1291–1299, 2007.

[4] B. Townsend, J. Abawajy, and T Kim. "SMS-Based Medical Diagnostic Telemetry Data Transmission Protocol for Medical Sensors", Journal on Sensors, 2011.

[5] M. Cibuk, and H. Balik. "A novel solution approach and protocol design for bio-telemetry applications", Advances in Engineering Software, vol. 42, issue 7, pp. 513-528, 2011.

[6] S. Ambler. "Encapsulating Database Access: An Agile "Best" Practice", http://www.agiledata.org/essays/implementation Strategies.html.

[7] D. Jayasinghe. "Exposing a Database as a Web Service", http://www.developer.com/db/article.php/3735771/ Exposing-a-Database-as-a-Web-Service.htm, 2001.

[8] J. Yang, Z. Zhang, and Y. Zhao. "Analysis on Database Connection Mechanism of WEB Application System in Dreamweaver", ICITA'2010 Wuhan, China, 2010, Conference Proceedings pp. 1-4, 2010.

[9] X. Qu, J. Feng, and W. Sun. "United access of distributed biological information database based on web service and multi-agent", Control and Decision Conference, Yantai, Shandong, China, Conference Proceedings, 2008.

[10] T. Sellis, D. Skoutas, and K. Staikos. "Database interoperability through Web Services and ontologies", 8th IEEE International Conference on BioInformatics and BioEngineering, 2008, Conference Proceedings pp.1-5.

Architecture of Automated System Software for Testing Petrol Engines

Georgi Krastev[1]

Abstract – **The paper presents an approach to developing systems for automation of laboratory investigations, already implemented at the University of Ruse. The automated microcomputer system is for testing of internal combustion engines by means of an eddy current dynamometer generates various family curves of engines under test.**

Keywords – **Laboratory experiments; Scientific experiments; Automated systems; Software.**

I. INTRODUCTION

When solving a number of highly complicated technical problems in the field of internal combustion engine testing, the immediate participation of a man in the performance of some experiments is not possible. In these cases the real experiments can be replaced by laboratory ones on the basis of Automated Systems for Scientific Investigations (ASSI).

This paper focuses on Automated Systems for Scientific Investigations software that has been developed in University of Ruse for Testing of petrol Engines. The system was developed noting the advantages and disadvantages of similar systems [1-5].

II. FUNCTIONS AND OPERATION MODELS OF ASSI

ASSI for testing of petrol engines operates in the following modes:

- Adjustment: - Technological process and control system configuration. The configured data base that includes some elements of the relation data bases such and data searching is updated, if necessary. With the help of the data base, the parameters of the technological processes simulations, the tasks, the translations, the control system adjustments, etc. are configured;
- Measurement: - Registration of the main parameters of the tested object. The operator has the possibility to examine the most important information concerning the process conditions, in particular - inlet/outlet points whose values exceed the pre-set-limits;
- Control: - The system's automatic control is implemented on the basis of the information coming from the object and the tasks set by the operator. The face panels of industrial regulators and the programmable controllers which have been developed, simulate the actual units. They enable the operator to observe and control all tasks and system adjustments in real time mode. The digital inputs and outputs as well as states of the alarms are displayed as Boolean indicators;
- Simulation of the running processes in real time mode: - It provides a possibility for investigating the influence of different signals (i e., change of colours text messages, audible alarms, etc.) over the operator on emergency;
- Testing: - This mode serves for performing different verifications of the correct functioning of any unit and the system as a whole on adjusting, for controlling in the process of preventive maintenance, for searching the reasons that caused error conditions during control program running, and also for adjusting the parameters of the hardware components.

ASSI implements the following functions:

- Acquisition and initial processing of analogue and digital data of the engine's main parameters;
- Automatic stabilization of the control contours of the supply, the advance angle of ignition and loading;
- Carrying out of any control algorithm given as a program;
- Automatic registration of object's states;
- Signaling and registration of extreme and emergency states;
- Start/stop control of the tested object;
- Interaction allowing the intervention in the control and the display of operative information (task, current regulated quantity, extreme values and adjustment parameters and time registration);
- Documentary records of the technological process conditions and updating the date base of the object under test;
- Effective self - diagnostics of the system including current control of the hardware/software state and displaying of error messages.

III. ASSI GENERAL VIEW

The general view of the program system is shown on fig. 1 in a hierarchical structure on two levels: operator's station and controlling microcontroller.

3.1. Software of the Operator's Station

It helps the automation of the preparation procedure of the tests and processing of the results. The software is developed and set for operation in WINDOWS environment.

[1]Georgi Krastev is with the Faculty of Electrical and Electronic Engineering and Automation at University of Ruse, 8 Studentska str., Ruse 7017, Bulgaria, E-mail: gkraste@ecs.uni-ruse.bg.

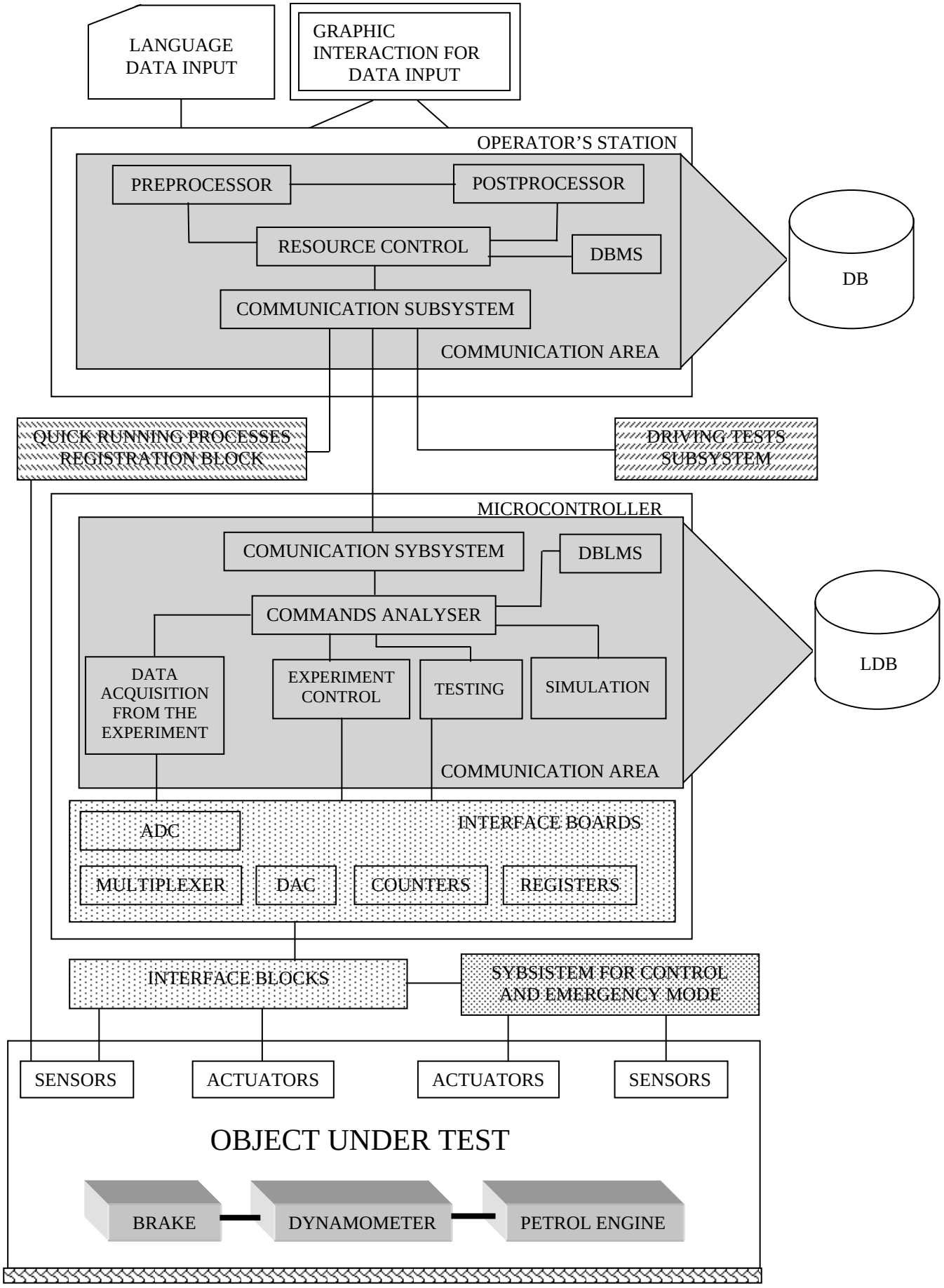


Fig. 1 Architecture of Automated System Software

By means of the pre-processor are given the sequences of operations for carrying out of the testing, for determining the facilities and resources leading to effective achievement of tests' goal. The information referring to the aim, test tasks and the additional data for planning are entered by means of an input language for operation in a scenario mode or interaction mode.

The post-processor implements the visualization both of the input data and the results obtained. It formats all the necessary protocols of the testing, its printing and plotting of the graphic information.

The communication area includes the entire information of the system: - input and invariable data; information as a consequence of pre- and post- processor operation; results acquisition. The information in this area is accessible for all modules using the Data Base Management System (DBMS).

The system applied in this research is dBASE for WINDOWS. The operator's station is in connection with the other units of the system via the following interface buses: specialized parallel bus; interface bus for successive data transmission through line RS 232C; general purpose interface bus (GPIB) - IEE 488.

The connection of the operator's station with the subordinate microcomputer can be made using one of the three above stated interface buses.

3.2. Software of the Subordinate Microcontroller

The software of the subordinate microcontroller is composed of the following basic modules: module for data acquisition form the experiment; experiment control module; testing and simulation modules. All modules communicate via communication area. The invariable information is organized in a Local data base (LDB).

The communication area of the microcomputer includes the entire local information of the system: the invariable data, internal logical keys (flags), internal arithmetic variables, values of timers, etc., as well as the results obtained by the time of their transfer to the operator's station. The information of this area is accessible for all modules using the Data Base Local management System (DBLMS).

The simulation subsystem is designed for of different disturbances, control actions, for obtaining and maintenance of the desired scheme of imitation that determines the technology of testing. The subsystem for monitoring and emergency operation is independent and uses the information from the common sensors and its own (availability of flowing water through the brake, outlet eater temperature after the brake, availability of cooling liquid in the engine, luboil pressure, extreme temperatures, etc.) and actuators for initiations of different interlockings of the engine on emergency.

For driving test subsystem the car is equipped with a special microcomputer (PAD). Prior to the tests it is necessary to enter the results should be transferred to the post-processor of the processing station.

IV. Conclusion

The comparison of the developed ASSI with similar systems is not a simple and easy solving task, because the comparison criteria are not uniform, they concern different aspects, starting with the modes of operation and duration of the testing procedure. It is difficult to generalize and present all comparison criteria by a single criterion of a concrete mathematical expression.

The system can be used both in the practice and research area. It can be also used for investigation of other types of engines, e.g. diesel engines. Then it is necessary to provide additional equipment, invl. suitable sensors and actuators. No need to make any changes in the software package.

References

[1] Formalized description of technological processes - basis if the program modules package for developing ASTPC (Automated System for Technological Processes Control) for testing of engines. Enginebuilding, 2010. N3, p. 33-34.
[2] Prospectus papers of AVL Company.
[3] Prospectus papers of Motor test Company
[4] Prospectus papers of SHENK Company.
[5] Ralf Erik Plewe. Die Funktionalitat von Steuergeraten fur elektronisch geregelte Verteilereinspritzpumpen Systeme. MTZ motortechnische zeitschrift, 2003, N12, p. 628-632.

Mazes – Classification, Algorithms for Finding an Exit

M. Todorova[1], N. Nikolov [2]

Abstract - **The paper presents a classification of various types of mazes according to three basic criteria. The report analyzes, explores, and researches an algorithm for finding an exit by a line-following robot in a maze. The paper outlines the basic algorithms for controlling an object and suggests possible modifications for achieving better results.**

Keywords - **maze, robot, reflective sensor, exit finding algorithm, dimension classes, tessellation classes, routing classes**

I. INTRODUCTION

An algorithm for finding a route in a maze is a method aiming the execution of predefined tasks for passing from an initial state through several consecutive intermediate states to a final state. The transition between the separate states from the start to the finish does not have to be predetermined. There are existing probability algorithms allowing for a random choice [1].

Maze navigation is one of the main tasks in robotics. Finding the exit of the maze and reaching it as quickly as possible is the main goal. The information provided by the external sensors is a prerequisite for accomplishing the task successfully.

The paper introduces some algorithms for steering a robot moving in a maze.

II. MAZE CLASSIFICATION

Mazes can be organized along three classification criteria:

- Dimension;
- Tessellation;
- Routing;

The dimension classes are shown in Fig. 1.

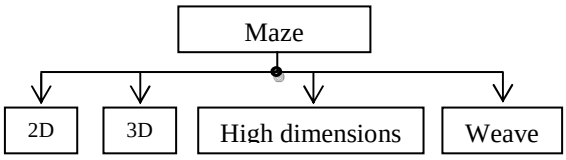


Fig. 1. Dimension Classes

- 2D – most mazes belong to that class. They do not have any overlapping passages;
- 3D – mazes with multiple levels. Passages can go in six directions – left, right, up, down, forward and backward.
- Higher dimensions - 4D and higher dimension mazes;

The tessellation classes according to the geometry of the individual cells that compose the maze are nine. The classes are shown in Fig. 2.

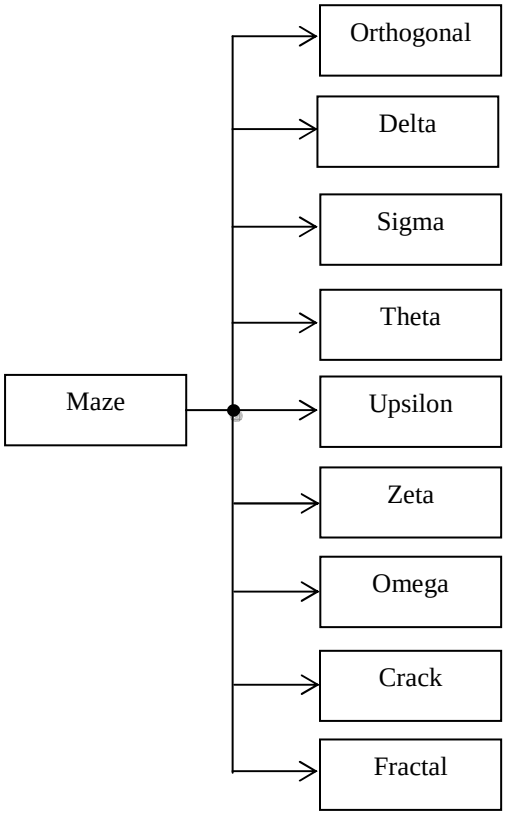


Fig. 2. Tessellation Classes

- Orthogonal – a standard rectangular grid;
- Delta – the cells are triangular. Each cell can have three neighbouring cells;
- Sigma – each cell is a hexagon. Each cell can have up to six neighbouring cells;
- Theta –The start (exit) of the maze is in the centre and the finish (entry) is on the outermost concentric circle. Each cell can have four or more neighbouring cells;

[1]Maya P. Todorova, Technical University – Varna, Computer Sciences and Engineering Department, 9010 Varna, Bulgaria, E-mail: mayasvilen@abv.bg

[2]Nedyalko N. Nikolov, Technical University – Varna, Computer Sciences and Engineering Department, 9010 Varna, Bulgaria, E-mail: ned.nikolov@tu-varna.bg.

- Upsilon – the cells are octagons or squares. Each cell has either eight or four neighbours;
- Zeta – a rectangular grid with diagonal interlocks (passages);
- Omega – all mazes with a consistent non-orthogonal tessellation. Mazes with randomly shaped and sized cells;
- Crack – mazes without any consistent tessellation, but rather have walls or passages at random;
- Fractal – a maze composed of several smaller mazes.

The routing classification refers to the types of passages between separate elements within the defined geometry. The routing classes are shown in Fig. 3.

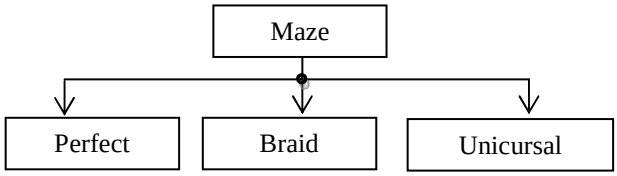


Fig. 3. Routing Classes

- Perfect – mazes without loops or closed circuits. The maze has one solution;
- Braid – mazes without dead ends. The maze consists of passages coiling around and running back into each other;
- Unicursal – mazes without any junctions. The maze consists of a single passage spanning from the start of the maze to the finish. When a U-turn is made the passage leads back to the beginning [3].

Almost all mazes can be described as a directed graph with a finite number of states and finite number of possible continuations of each state.

III. ALGORITHM FOR FINDING AN EXIT IN A MAZE

A.„Black Line Follower"

The algorithm involves determining the exact orientation of the robot towards the outlines of the maze. This is achieved with the use of reflective sensors placed 5 -6 mm above the surface. The robot has five reflective sensors. The sensor reads a value of 1 when on a light surface and 0 when on a dark surface. The algorithm is applicable for mazes without closed circuits. There are eight possible situations that have to be taken into account and worked out – left turn, right turn, left or right turn, to the left or forward, to the right or forward, the four compass directions, dead end, end of the maze. The main execution stages are the following:

- following the line until a junction is reached;
- establishing the type of the junction;
- deciding a follow-up action depending on the type of the junction.

In order to achieve faster exit from the maze, the process of movement has to be split into two stages. The first stage is a "route reconnaissance" passing through the maze. During that stage the shortest route to the exit is established and stored into memory. The second stage involves following that route and reaching the exit in the shortest time possible.

The turns made during the initial passing are stored. Non–optimal turns have to be corrected. When a part of the maze requires a left or right turn the robot makes a turn without having to make a decision – there is no alternative way. These turns are not stored. The U-turn is also without an alternative but has to be stored because a dead end is reached. The shortest route does not include dead ends. That's why the last turn has to be corrected in order to be avoided during the second passing. The movement of a robot in a maze is traced experimentally so that the shortest route is found. The passing is carried out following the "left hand" rule. The maze with its entry point and exit indicated is shown in Fig. 4.

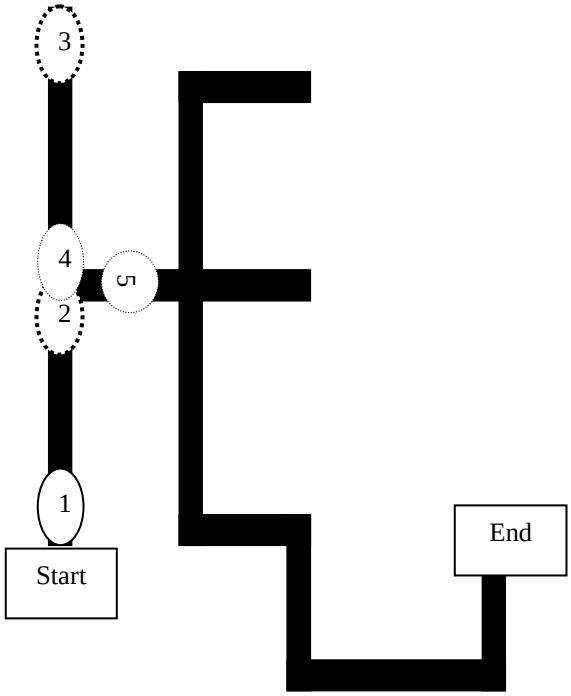


Fig. 4. Maze without Overlapping Passages

The robot is at the beginning of the maze (1 in Fig. 4) and starts its forward movement. The "decisions" memory is empty. The first junction is reached (2 in Fig. 4). The next direction is forward, the "left hand" rule is applied. The choice is stored into the memory and its record reads S. Moving forward, the sensors data change from 11011 to 11111. Therefore, a dead end is reached (3 in Fig.4) and U-turn has to be made. It is written into the memory and the robot goes on with its forward motion until the next junction (4 in Fig.4). Upon reaching the junction, a decision is made for a left turn and L is written into the memory. The memory record is SUL.

Reaching a dead end means that the previous turn was wrong. A U-turn record in the memory means that corrections have to be made for avoiding reproduction of the wrong turn. An analysis of the direction choice just before and after the U-turn has to be made. In the case above these are the directions Forward/Straight (S) and Left (L). Therefore, in order the

dead end at the first junction to be avoided, a right turn has to be made. The incorrect decision (SUL) is erased from the memory and changed into R. The memory record is R. After the corrections the robot continues moving through the maze. Determining the shortest route involves substituting ineffective decisions using the Substitution Table I.

TABLE I
INEFFECTIVE DECISIONS SUBSTITUTIONS

LUL	S
LUS	R
SUL	R

Following these basic rules as shown in Table 1, the robot will succeed in finding the shortest route in any maze without overlapping passages.

B. „Wall Follower"

The "Wall Follower" algorithm is described in Fig. 5.

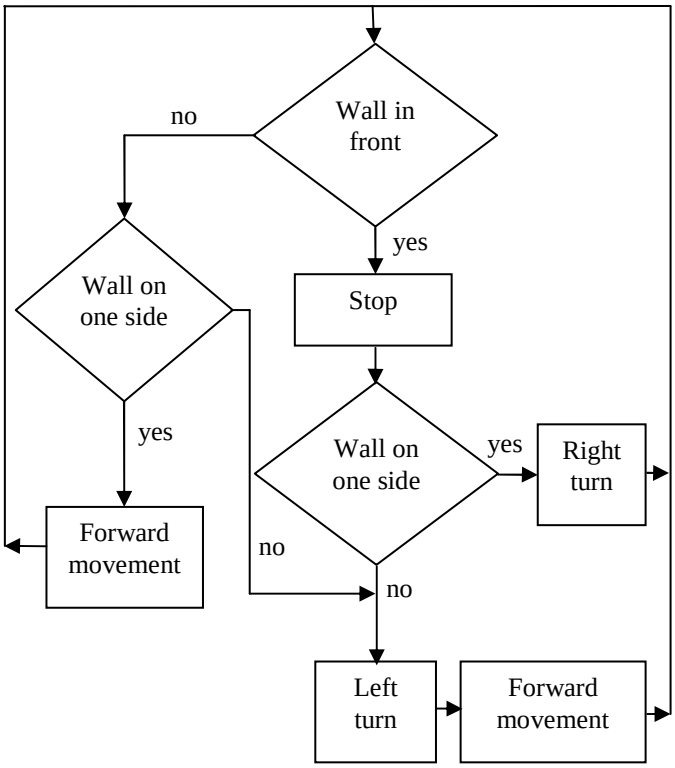


Fig. 5. "Wall Follower" Algorithm

Using that algorithm, the robot always finds the exit of the maze [5].

The algorithms can be realised with MPLab.

IV. CONCLUSION

Good knowledge of the various types of mazes helps for the correct choice of maze solving algorithm. The successful application of the algorithms for robot's movement in a maze depends on:

- the type of the maze;
- the choice of the algorithm for passing through;
- using improvements in the maze movement algorithm;
- the information provided by the external sensors.

The main goal of the presented algorithm is reaching the exit of the maze in the shortest time possible.

REFERENCES

[1].http://en.wikipedia.org/wiki/Algorithm?&oldid=295366239

[2]. R. Dofr, R. Bishop, Modern Control System, Peason Education, 2008.

[3]. http://www.astrolog.org/labyrnth/algrithm.htm

[4]. Aleksandar, P. S. *„Проблемы Гильберта"*. Moskva : s.n., 1969.

[5]. Дипломна работа, дипломант И. Иванов, научерн ръководител доц. д-р инж. Недялко Николов

[6]. www.pololu.com

Integration of Biometrics to the E-Health

Milena Stefanova[1] and Oleg Asenov[2]

Abstract – **This report presents an approach for application of finger vein biometric identity management in the healthcare sector. It is emphasized that the applying of finger vein recognition and identity management is the cost effective way that ensures an incessant and integrated protection of users in e-health services. The basic concept and examples for integration of vein identity management solution into typical healthcare IT-services are presented in the paper.**

Keywords – **Biometrics, Finger Vein Identification, e-Health, e-Government.**

I. Introduction

Information and Communication Technology (ICT) applications for health are referred to as "e-Health". These applications include means and services generally aiming at facilitating and improving healthcare. E-Health supports services and activities facilitating the supply of health-related goods and services. E-Health also includes the provision of health-related knowledge infrastructure and/or collaboration platforms and it depends on the existence of the necessary IT infrastructure [1].

Present-day e-Health solutions should provide for:

- Flexible and scalable web-based solutions;
- Tools to help security and privacy management;
- Integration to current clinical workflow.

By application of finger vein biometric in the health care, the approach to enable secure access is offered and in this way to improve privacy and the security management.

Predicting the development of health insurance is that people will take a more active role in managing and controlling their health record information. According to one of all the future health care network predictions, users will access their own e-health records over the Internet with secure logins and encrypted transmissions. These features readily exist at present. The possibility of their practical application is that people can consider their own medical records as important as electronic banking records.

In the paper, the place of biometric identity of e-government services access is examined. The important advantages of the finger vein over all other technologies are offered. The application model and requirements to IT-resources are described. The possible decision with biometric identification in healthcare sector is presented.

[1]Milena Stefanova works at the Faculty of Mathematics and Informatics, St. Cyril and St. Methodius University of V. Tarnovo, 3 G.Kozarev str., 5000 Bulgaria, E-mail: m_stefanova@abv.bg.

[2]Oleg Asenov works at the Faculty of Mathematics and Informatics, St. Cyril and St. Methodius University of V. Tarnovo, 3 G.Kozarev str., 5000 Bulgaria, E-mail: olegasenov@abv.bg.

II. E-Government And Biometric Identity

Each successful project of e-government is associated with the opportunity every citizen to be identified in order to get access to their own personal pages quickly, easily and reliably. Depending on the type of information and services offered in e-government projects usually there are two levels of identification defined:

− Identification without the need of verification of presence;

− Identification with necessity of presence confirmation.

The solutions for biometric identification are usually applied to the second scenario, they are also applicable to the first one if there is a requirement of a high level of protection. Identification of the individual where the presence is not required can be designed by a unique numeric identifier on a portable electronic carrier which is linked to the data on individuals in a secure database.

Identification, which requires obligatory presence, can be made by presenting the same unique identifier and verification of holder by an official (this brings a subjective element) or by the application of technology that requires the presence of the individual [4]. Recognition for such purposes must comply with several criteria:

− Adequate protection against violent crime: theft, forgery or counterfeiting;

− Uniqueness and consistency in time (the uniqueness is similar to those in DNA (Deoxyribonucleic acid) recognition);

− Contact-less method (convenience and hygiene reasons);

− Minimal digital size of used template;

− Ensuring the presence in the identification;

− Minimum time for identification (speed).

Biometric authentication by scanning the individual geometrical position of the veins in finger or palm of the hand is a method that meets all the **necessary criteria**:

− *Protection* – veins are hidden in the human tissue and their removal / tampering is virtually impossible, even through complex surgery, in addition, the method itself requires the presence of live (with circulation) tissue for identification, thus preventing the possibility of criminal abuse.

− *Uniqueness* – this method is second (after DNA) in terms of uniqueness in identification – It is unique even in Twins and remains constant in adults in the course of time.

− *Contactless* – the identification is done by non-contact scan image by thermal camera.

− *The smallest size* – digital imprint is stored in only 400 B.

− *Requirement of presence*. Biometric identification through this method guarantees presence. This would allow the real treatments and procedures to be controlled as well as a high level of protection to be secured in cases of identification of presence.

− *High speed of identification* – the identification time is less than 1 second for up to 20 000 users.

III. THE TECHNOLOGY

Identification functions on the principle of finger vein structure and each finger of each person is unique [10]. Based on this discovery made in 2006, a vein-code technology for reading and recognizing people has been developed and this technology has **major advantages** in comparison with other technologies:

− Reading the vein-code is *contactless* – fingers and sensors are not in contact;

− Practical *zero-error in detection* and recognition. It does not depend on the quality of the finger skin, the dirtiness of the skin or the presence of surface injuries;

− *The time* for reading and identification (in the system with comparison by 1:1 method) *is less than one second*;

− Recognition takes place only when blood flows through the veins and this type of identification *can not be falsified* (medically proven that it is not possible to imitate the same vein-code through plastic surgery). Vein-code is recognized as a method for rapid identification of individuals by FBI in December 2009.

The presented device has been developed by Japanese-American consortium Hitachi-M2SYS. The software based on this method of biometric identification is compatible with different types of biometric devices [6] and it allows:

− Avoiding duplication of medical records and eliminating language barriers and poor literacy;

− The process of identification of the patient is optimized, and thus the work efficiency of employees in health care is improved;

− Identification of the patient and also verification of their health status;

− The use of the patient biometric identification during doctor's rounds is quicker than usual and the medics have no doubt that the prescribed treatment is for the exact patient;

− Because of the uniqueness of the individual biometric templates, the software prevents duplication of patient medical data; it alerts the medical employee that the data of the patient have already been introduced in the system;

− It integrates easily with the existing hospital-patient software; the system can work up to 24 hours without the need of new program code or integrating data base.

The technology has been developing and finger-vein reading sensor for mobile devices [8] and for tablets [7] has already been designed.

IV. IT-RESOURCES AND MODEL APPLICATION

The digital template, used for biometric identification, is in size not larger than 400 bytes, which provides easy transportation including low-speed communication.

To access the system, clients should have only a web browser and biometric reader, shown on Fig.1 [5], where identification is required to verify of presence. Table I presents the fixed specifications of the biometric reader [5].

For the purposes of application the method of biometric vein-code identification in protection against illegal malicious user access with replacement of identity, when accessing personal e-health records, an architectural model for the implementation of user applications is considered [2,3]. That model [2,3] for the performance of the protection through biometric identification with vein-code for e-health applications is used.

In the presented model, there is a virtual Active X Bridge built and it runs as an executable code on the servicing Web-Server for access from the user workstation to the Host Application. In the model, Client Manager is launched as part of a web-service, supporting Active X gateway. Before the Client Manager permits or denies right of entry of the user, the e-health records are not accessed.

Fig. 1. The Hitachi USB Finger Vein Biometric Authentication Unit

The users must identify themselves on the client computer with biometric reader, connected by USB 2.0 port. Through ActiveX Bridge the registered biometric template is transmitted to the Bio-Plug-in Server to compare with the already stored patterns in the Finger Vein Database. Once permissions are matched and configured– the e-health record is accessible and it is interpreted by the client's browser. In case of mismatch– the application is rejected and medical data are not accessed.

TABLE I
FINGER VEIN READER SPECIFICATIONS

Item	Specifications
Capture System	Infrared LED + Camera
Interface	USB 2.0
Dimensions	59 (W) x 82 (D) x 74 (H) mm
Weight	96g
Power	DC5V +/-5% <500mA (Power from USB bus)
USB Cable	1.8 m
Certifications	FCC, CE

The system allows working with virtually unlimited number of people. In the presented biometric identification, the vein-code cannot be forged or faked, which means that the person is the one who actually has the right to work remotely with the information or input data. The used vein-code ensures that the user who applies for access to the resources of e-healthcare has been certified in person. No person can pass their personal vein-code to a third party. Practically it is impossible to catch the use of the template because every time a scanning over communication interface template is sent, it is encrypted with a different key.

V. Decision on Health Care

The suggested solution is due to a detailed survey and analytical research of the active regulation algorithm of processing of health clinical paths in the healthcare sector in Bulgaria. The analysis is carried out from the point of view of tracing "the way" of the patient from the GP-doctor to the specialist. An optimized model of algorithm for health clinical path information processing, which provides considerably greater extent of authenticity of input data, has been offered. At each phase of the process the presence of the patience is certified by an identification of unique vein-code of one or more fingers of the hand. The registering of the vein-code gives an opportunity to monitor the time and place of patient movement in the process as well as their real "physical" participation in this process.

Thus the application that has been worked out makes the process of information processing of health clinical paths more objective by applying the technology of biometric identification; it leads to decreasing the amount of "paper" documents for processing and giving accounts of this process. The application contains three fundamental items for managing the main activities in the healthcare by biometric identification with the help of the vein-code:

A. Registration of patients in the uniform database (Fig. 2):

− The necessary data in predefined fields are completed;
− The picture is added – from a file or from connected to the place of work camera;
− The input data are recorded;
− For better security, two fingers are registered;
− The input information is recorded once again.

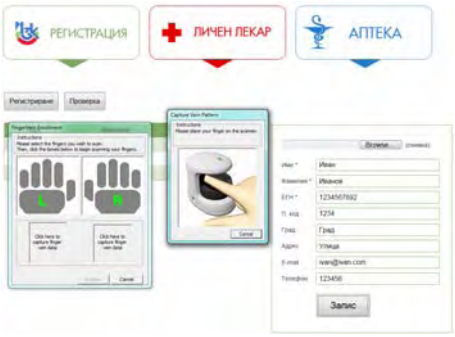


Fig. 2. Patient's registration

B. Registered patient visits their GP-doctor:

− After the identification and examination treatment and a recipe are prescribed;
− In addition to records of recipes details of conditions, drug intolerance, treatments performed, etc can be kept;
− It allows verification of the recipe situation for the registered patient: which have been fulfilled and which have not.

C. Registered patient visits Pharmacy (Fig. 3):

Fig. 3. Patient visits Pharmacy

− After identification of the patient, fulfillment of recipe is registered;
− Similarly, records of any treatment can be implemented, monitored and kept.

VI. Conclusion

Using the highly secure authentication principle of finger vein verification, corporations and healthcare institutions can ensure user identity and data security and minimize data security breaches and cyber crimes taking place at enterprise worldwide [9]. Biometrics can play an important role in authentication applications, since they are strongly linked to the holder, and difficult to forget, lose or give away. It is important that the biometric systems are designed in the way, which can resist the attacks, working in security-critical applications, especially in unattended remote applications such as e-health sector.

References

[1] Progress Consulting S.r.l. and Living Prospects Ltd., *Dynamic health systems and new technologies: e-Health solutions at local and regional levels*, Comittee of the Regions of the European Union, 2011.
[2] M. Stefanova, "Model for user's identity protection by access to web-applications," V International scientific conference "Innovations in technology and education", Kemerovo / Belovo, May 2012.
[3] M. Stefanova, T. Stefanov, and O. Asenov, "Access to e-government' services through vein-code biometric identification," CISTI'2012 (7th Iberian Conference on Information Systems and Technologies), Madrid, June, 2012.
[4] W. Chuck, *Vein pattern recognition: a privacy-enhancing biometric*, CRC Press Taylor & Francis Group, 2010.
[5] http://m2sys.com/finger-vein-reader.htm
[6] http://m2sys.com/rightpatient-biometric-patient-identification.htm
[7] http://www.biometrictribe.com/?p=471
[8] http://www.fujitsu.com/global/news/pr/archives/month/2011/20110419-01.html
[9] http://www.hitachi.com/rd/yrl/conf/2011/ijcb/index.html
[10] http://www.hitachi.eu/veinid/documents/veinidwhitepaper.pdf

Psychology of the Perpetrators of Computer Criminal Acts and Review of Legal and Economic Consequences for the Community

Zaklina Spalevic, Jelena Matijasevic, Dejan Rancic

Abstract - **There are different categories of perpetrators of computer criminal acts. In this paper, we have devoted attention to motivational analysis of perpetrators of computer crimes, and to analysis of the psychological profile of hackers, given that a motive is significant circumstantial fact. We have highlighted the features of hacker culture and hacker ethics, with review of legal and economic consequences of committing these offenses. Because, it is clear that society may adequately oppose to a certain phenomenon, only if consider all of its properties and if come in all aspects of its specificity.**

Keywords - **Computer crime, The motives of performing criminal acts, Hackers, Hacker culture, The Hacker ethic**

I INTRODUCTION

There is no single sphere of life, from production, trade and service provision to the national defense and security in the widest sense in which computer does not have practical application. Nowadays we are all aware of the enormous significance of computer use in contemporary societies and of the fact that there is not a single area of human activity in which computers are not being used. However, the conclusion that there has not been a single technical and technological accomplishment that has not been misused in various ways is pretty devastating. Phases of development in which the invention was susceptible to misuse, groupings of persons who committed such actions and different intents of misuse represent specific characteristics [1].

At the beginning of implementation of computer technology, computers were not be eligible for great abuse, because their application were not be mass, so it was focus of interest of a very small number of users - IT professionals. What opened the door to expanding opportunities to misuse of computer technology in various applications, is its rapid development, simplifying its use and availability to wide range of users [2, p. 852].

Computer technology has very rapidly development. Also, such level of development has education and training those people who would abuse it. Lately, in the media, very often appears some information that an individual (or group) illegally accessed to a computer system, and came up with some data, or created the ability to threaten or triggers such systems like nuclear potentials with great powers. This phenomenon is not characteristic only for developed Western world. It is becoming characteristic of our - Balkan region. In this paper, we have devoted attention to motivational analysis of perpetrators of computer criminal acts, and the analysis of the psychological profile of hackers.

[1]Zaklina Spalevic is with the Faculty of Law for Economy end Justice at University Business Academy, Geri Karolja 1, Novi Sad 21000, Serbia, E-mail: zaklinaspalevic@ymail.com.

[2]Jelena Matijasevic is with the Faculty of Law for Economy end Justice at University Business Academy, Geri Karolja 1, Novi Sad 21000, Serbia.

[3]Dejan Rancic is with the Faculty of Electronics at University of Nis, Aleksandra Medvedeva 14, Nis 18000, Serbia.

II THE MOTIVES OF PERPETRATORS OF COMPUTER CRIMINAL ACTS IN THE FIELD OF CYBERCRIME

There are different categories of perpetrators of computer criminal acts, with respect to a variety of criminal acts that they commit and considering motives which impel them to engage in such activities. Individuals who illegally access to another computer or network without further criminal motives are significantly different from an employed in a financial institution or bank, who with abuse of information technologies accessed to other computers and make substantial damage.

The emotional mood, needs, interests and attitudes are an integral part of the mental processes that we call motives. Motivated behavior involves all aware of the elements that occur under the influence of needs, goals, aspirations and interests of people in general. The motive is the current psychological phenomenon. The motive is the incentive for crime. It means incentive as internal cause, and also reason as external category, which affect the character, instincts, feelings, perceptions of people. Because of the treatment an appropriate incentives and cause to the personality of the offender, he sets to himself a goal he wants to accomplish making a criminal act. The needs, interests, customs, beliefs, traditions, instincts, passions, desires and feelings lead to motive for the crimes [3,4].

Motive is an important indicator of many classic forms of criminality, and even computer criminal acts. Motive as a clue becomes prominent in setting up versions of suspected persons, regardless of whether it is a case of a single perpetrator or a case of a group of perpetrators, where the method of elimination is used so as to remove suspicion from innocent persons [5, p. 323].

Obtaining illegal financial gain by committing computer criminal acts is one of the most common motives found in perpetrators of these criminal acts. However, this motivation can be induced by various wishes of the perpetrator, such as unjustified gain, possibility of repaying a debt, an adequate status in society, satisfying certain personal vices and the like. Revenge, inferiority complex, economic competition, the desire for self-approval and achieving a certain success, as

well as envy, hatred, jealousy, enthusiasm for one's own knowledge and skills and even political motives in some cases can all be possible motives for committing computer criminal acts [6].

In this context we may emphasize the parallels with organized crime. Bearing in mind that "the desire for profit is the main motive of most criminal offenses committed in the field of organized crime", we can notice a certain similarity of some activities in the field of cyber crime and activities in the field of organized crime.

So, when it comes to computer offenses, we note that a heterogeneous array of motives beyond, in scope and variety, most of the motivational systems of other groups of criminal offenses. Motivational factors of computer crime perpetrators that appear most frequently as dominant [6,7,8] are: intellectual challenge, curiosity and adventurousness, fun, a sense of omnipotence, the need to triumph, dizziness own knowledge and skills, compensation and personal sense of social inferiority, elitist, revenge, internal pressure (hacking) rules, prestige (reputation). The *offender sharing computer crimes based on the structure* motivational.

III TYPES OF THE COMPUTER CRIMES OFFENDERS BASED ON THE MOTIVATIONAL STRUCTURE

There is a general division of perpetrators of such acts into malicious ones, who commit crime so as to obtain financial gain or just cause damage, and into perpetrators who are not motivated neither by obtaining gain, nor by causing damaging consequences, but simply find pleasure in unauthorized penetration into a well-secured information system.

Malicious perpetrators of computer crimes are mostly motivated by greed. Data from practice indicate a definite set of characteristics that form their criminal profile: about 80% of them are first-time offenders, 70% of them have been working for more than five years for the company which is the damaged party; they belong to the age group below 30; they are mostly male, highly intelligent; they generally have several years of business experience and are considered as conscientious workers that don't cause any problems while fulfilling their work tasks; their degree of technical competence surpasses technical qualifications required for their work position; the perpetrators do not consider themselves thieves or criminals in general, but just borrowers [2, p. 388].

Computer criminal acts motivated by greed are very common in banking, financial corporations and insurance companies. Statistical data on the perpetrators of computer crime in the area of banking indicates the most common occupations of the perpetrators: 25% are persons who have special authorization and responsibilities for IT systems; 18% are computer programmers; 18% are employees who have access to the terminals; 16% are cashiers, 11% of them are operators – IT Specialist, and 12% are persons outside the affected corporation, including the service users [2, p. 388].

The second group of perpetrators of computer criminal acts find deep pleasure in the very act of breaking into multiple security IT systems. The higher the security of the system is, the higher is the challenge to engage in such activities Here we are dealing with so-called hackers. Regarding professional affilation, they are usually computer programmers, operators or highly qualified informaticists, and sometimes they are just people with computers as hobby.

Given the fact that the second group of perpetrators of computer criminal activity raises a lot of attention, causes much contraversy and mixed reactions and that even the computer networks of governments of modern countries were targets of these perpetrators, we will further examine the hacker profile in the following text.

IV PSYCHOLOGICAL PROFILE OF HACKERS

Although there are a variety of prejudices against hackers, it is clear that all hackers share the following features (based on different analyses of this specific group of perpetrators of computer crimes): a high IQ, consuming curiosity and the ease of intellectual abstraction. They have an increased ability to absorb knowledge and they pay attention to a variety of details which are irrelevant to the "ordinary people". Hackers are not interested in just one area; on the contrary: they tend to be involved in any subject that stimulates intellectual effort. On the other hand, hackers are afraid of control and do not want to deal with anything binding or authoritative. Similarly, they have no ability of emotional identification with other people, according to many authors. They often tend to be arrogant and impatient with people or things they believe are wasting their time.

Still, there is one thing some of them are exceptionally good that – social engineering. Social engineering denotes the ability of disclosing confidential information by manipulating people, but not breaking into a computer. This method is based on the assumption that man is the weakest link in the chain of security. It is most often used by telephone or the Internet and it makes people reveal their confidential information (such as passwords used to access accounts and credit card numbers) or do illegal things.

Hackers believe that many of their illegal acts are justified and ethically correct. The psychologist Lawrence Kohlberg has developed a three-level theory to explain moral development in normal people. The first level deals with avoiding punishment and receiving rewards, the second level comprises social rules and the third one includes moral principles. Each of these level contains two phases. Computer criminals have only evolved through the lowest three phases of the Kohlberg model: two phases of the first level and the first phase of the second level [7].

Sarah Gordon, an expert on the psychology of computer criminals, conducted a survey of the Dark Avengers, notorious creator of the virus. He consistently refused to acknowledge responsibility for his creations and, like traditional computer criminals, blamed the victims. Dark Avenger said that human stupidity, but not a computer, spreading the virus. Trying to justify the creation of destructive viruses, he said that most of the computers that were attacked did not contain any important information, and therefore his malicious program has not made any real damage. At the hearing, said that he hates when people have more computers than him, especially

when those resources are not used to what he considered constructive [9].

Hackers have their own specific culture in which they are recognizable. Although they do not need to have a prejudices, it is clear that every culture, including the hacker culture, has its own distinctive characteristics in terms of key characteristics, mode of communication, relations among members, behavior, habits, etc. The hacker culture began to develop in early 60s of last century. After the 1969th was merged with the technological culture which included the founders of the Internet. Over time, it assimilated all cultures associated with technology and computers, and by the 1990s hacker culture is almost equal to what is called "Open Source Movement." The central pillars of the hacker culture in which it develops are: Internet, World Wide Web, the GNU project, Linux operating system and all the hacker creations. From the 1990s until today, the hacker culture gets some recognizable symbols: Tux, the Linux penguin, the BSD Daemon, Perl Camel, and the hacker emblem.

Hackers have also developed a specific way of communication, which is another important characteristic of them. Due to the fact that they are much more successful in written communication than in face-to-face, interpersonal communication, they have adopted „leet speak“. Leet speak is an encrypted form of writing in which letters are represented by numbers, symbols and other signs that resemble the letters. The basic function of this form of communication is to exclude „outsiders“ from the communication, i.e. to make a clear difference between the language of this group of people and the language of the the majority. Leet is not to be confused with the so-called AOL language found on the Internet. The primary function of AOL language is to shorten written forms of some words, while the purpose of the leet speak is to make traditional language incomprehensible to people who do not belong to this group.

V ABOUT HACKER ETHICS

There is no definitive and generally accepted definition of the hacker ethics. In a way, every person has their reasons and justifications for the things they are doing. In the same way, hacker ethics does not exist in the form of written, official document anywhere, although several authors have presented its entries.

According to Jargon File, hacker ethics is: a.) the belief that the dissemination of information is a powerful, positive characteristic and that it is the ethical duty of hackers to share their knowledge by creating free programs and enabling access to information and computer sources whenever it is possible; b.) the belief that breaking into a computer system for fun and research is ethically correct, as long as the hacker commits no theft, vandalism or reveals confidential information [10].

Both beliefs are widespread among hackers, and most of the hacker population, among other things, engaged in writing free software. Also, many hackers breaking into computer systems for fun, and many even going so far that have unauthorized access to system and send by e-mails free advice and guidance on how to improve the irregularities and imperfections to system-operators.

With the development of technology over time, the approach to determining the hacker ethics has changed. Two following approaches particulary stand out: The Original Hacker Ethics and the Hacker Ethic of 90s Hackers.

Steven Levy, the representative of The Original Hacker Ethics singled out six key principles of the hacker ethics in his 1984 book "*Hackers: Heroes of the Computer Revolution"*. Those principles are: access to computers and anything which might teach us something about the way the world works – needs to be unlimited and total; all information should be free (public); mistrust toward authority – promotion of decentralization; hackers should be judged by their hacking, and not by false criteria such as degree, age, race, sex or position in the society; computers are used to create art and beauty; computers can change life for the better. All of these principles suggest that hackers obligation is to remove the border, decentralize power, to evaluate people based on their capabilities and to improve lives through computers.

On the other hand, The Hacker Ethic of 90s Hackers is essentially contradictory to the Original Hacker Ethics, because it advocates the opinion that the activity of hackers should be safe, that it should not damage anything, that it should not threaten anyone either physically, or mentally or emotionally, and that it above all should be fun for most people who practice it.

All previously stated principles of hacker ethics suggest certain duties, type of conduct, refraining, attitudes and needs. The extent to which the ethics is accepted and in what way it is interpreted was depicted in the classification of hackers on White, Black and Gray Hackers, which is based on adherence to and compliance with the principles of the hacker ethics.

VI CONCLUSION

In this paper we tried to present the most important segments related to the personality and actions of the perpetrators of computer crimes. Explaining the motivational structure and psychological aspects of personality, as well as aspects of the basic division with an essential overview of the principles of the hacker ethic and specificity of the hacker culture, we are closer to their most important characteristics. Successful steps in elimination the negative effects of certain phenomenon include not only an understanding of the phenomenon but also understanding the offenders of the crime and other illegal behavior in the computer crime field. Relevant fact is that this area includes both, the legal and economic standpoint. It is necessary to establish an adequate legal framework related to the abuse of computer technology, considering that the financial effects of these abuses can significantly damage the quality of economic operation of a state. On the other side, besides the economic effects on the economy in many countries, also is very important moral aspect of the use of computer technology. Certain principles in this area must be respected for the effective use of ICT sector. An important guarantee for quality legal aspect are adopted legislative provisions, which in a very efficient way representing the state's attitude towards the communication

culture with computer technology, and also include efforts to protect economy in countries from variety abuses in this area. Also it includes punishment of such behavior with adequate sanctions.

It should be noted that in addition to criminal acts which are directed against the security of computer technology and information system elements, there are a number of traditional criminal offenses which, with the help of using computers and computer components, offenders made faster, easier, it is more difficult to enter the track of them, and the consequences are far serious [1].

It is perfectly clear that the society can adequately confront a certain negative phenomenon only if all of its characteristics and specificities are recognized. Given the fact that the means of the misuse of computer technology are becoming increasingly advanced and more complicated to detect, and that it is very difficult to be step ahead of these criminal activities, it is necessary to keep raising public awareness about this phenomenon and to constantly work on finding the most adequate solution to various criminal activities in this field.

REFERENCES:

[1] J. Matijasevic, Z. Spalevic, "Specific characteristics of computer criminal offenses with regard to the law regulations", ICEST 2010, Conference Proceedings, Vol. 2., PO VII.1, pp.643-646, ISBN: 978-9989-786-58-7, Ohrid, Macedonia, 2010.

[2] Z. Aleksic, M. Skulic, *Criminalistics*, Faculty of Law, University in Belgrade and Public Enterprise „Official Gazette", Belgrade, 2010.

[3] D. Modly, N. Korajlic, *Crime glossary*, Center for Culture and Education, Tesanj, 2002.

[4] S. Petrovic, *Computer Crime*, Ministry of the Interior of the Republic of Serbia, Belgrade, 2002.

[5] B. Banovic, *Providing evidence in the criminal process of economic crimes*, Police Academy , Belgrade, 2009.

[6] *The socio-psychological profile of the perpetrator of a computer crime*, Faculty of Informatics and Computing, p. 5, http://www.dir.singidunum.ac.rs/ (2012, March 05).

[7] Main problems related to the Cybercrime, 10th United Nations Congress on the Prevention of Crime and the Treatment of Offenders,http://www.justinfo.net/UPLOAD/docs/argentina.htm (2011, November 29).

[8] The World of the Hackers; http://www.svethakera.com (2011, November 10).

[9] J. Matijasević, S. Ignjatijevic, "Cybercrime in legal theory, the concept, characteristics, consequenses", INFOTEH®-JAHORINA 2010, Conference Proceedings, Vol. 9, Ref. E-VI-8, p. 852-856, ISBN-99938-624-2-8, East Sarajevo, Serbian Republic, Bosnia and Hertzegovina, 2010.

[10] J. Matijasevic, M. Petkovic, "Criminal offenses against the security of computer data - the analysis of Criminal Law provisions and the importance in the context of combating cyber crime", International Scientific and Professional Conferences „Criminological and forensic research" 2011, Conference Proceedings, pp: 598-609,Vol. 4, No 1, ISBN 978-99955-691-1-2, Banja Luka, Serbian Republic Bosnia and Hertzegovina, 2011.

P2P WIRELESS NETWORK BASED ON OPEN SOURCE LINUX ROUTERS

Hristofor Ivanov[1] Miroslav Galabov[2]

Abstract – **In this paper we present our work towards deploying a community wireless network with ad hoc communication and routing between its elements. We describe our network model and implementation of wireless routers, while motivating decisions and pointing out open issues. The main advantage of our approach is the low deployment cost and inherent flexibility in terms of adapting the network configuration with little or no human intervention, which in turn can be exploited to support the dynamic addition, removal and mobility of network elements.**

Keywords – **Wireless Network, Open Source Firmware, Linux Routers**

I. INTRODUCTION

Algorithm design and evaluation in the field of wireless networks is performed using network simulators, such as ns-2 and NCTUns [1] [2], in order to systematically investigate system behavior under different assumptions, operating conditions and environmental settings. But it is also important to deploy and experiment with real-world networks. One key reason is that mathematical tools, even when used in conjunction with elaborate failure models, have limitations and cannot capture the full behavior of physical systems, such as the transmission anomalies in an inhabited area or the actual performance of commercial hardware. Real implementation and testing is thus needed to validate theoretically studied systems. Another, perhaps more important, motivation is that a testbed can actually be *used* to run not only test programs but also *real applications*. It is through such application-driven usage that unexpected system behavior is often discovered or new ideas emerge in terms of system and application functionality.

The deployment of infrastructure-based wireless networks has been straightforward and cost-efficient since wireless access points based on the 802.11a/b/g/n standards have entered mass production. However, this is not the case for ad hoc wireless networks given that wireless routers with ad hoc capabilities are hard to find in the market and are also quite costly. Another problem is that most such platforms are proprietary and closed so that is impossible to change internal settings let alone reprogram the network elements, for example to install a new routing or power management algorithm. This has led research groups to the development of wireless routers and networks based on personal computers and laptops [3]. While this achieves the desired flexibility in terms of software development and testing, it restricts the scope of deployment inside a single building or within an area of few neighboring buildings. Hence a PC-based network is not suited for urban environments where an outdoor, rooftop installation is usually needed to achieve good connectivity. In turn this poses additional requirements such as size constraints, weather protection, power supply and consumption and heat management. Last but not least, using PCs merely for the purpose of implementing wireless routers is too costly and would bring any large-scale deployment effort to a standstill. In this paper we present our work towards developing a community wireless network which employs low-cost, off-the-shelf 802.11b/g/n wireless routers running a custom Linux distribution. The network operates in ad hoc mode, thereby resulting in great flexibility and reduced administration. In the next sections, we discuss our motivation, give an overview of key hardware and software features of the router platform, present the current configuration and functionality, and discuss our experiences so far.

II. MOTIVATION AND REQUIREMENTS

In terms of functionality provided to the end user, we wish to deploy a wireless network of sufficient scale that will be used to run applications in a real-world environment outside of the laboratory. In addition, we wish to augment popular outdoor areas of the city with wireless connectivity to be exploited by mobile computers such as laptops and handheld devices. Besides supporting the typical suite of Internet applications such as e-mail, telnet, ftp and web browsing, we are interested in exploring peer-to-peer, groupware and ubiquitous computing applications. Notably, we do not want to limit participation only to students and faculty, and are strongly interested in attracting other communities like schools, local authorities, or even businesses, perhaps at a later deployment stage. On another dimension, we aim to create a testbed for implementing and evaluating algorithms and software, in the area of networking and distributed systems. Furthermore, besides using the network as a “dump” data carrier, our desire is to be able program the network elements in order to control their operation in a flexible way. We also wish to be able to install application-specific components, possibly on the fly, making the network itself an “active” part of the middleware or application's architecture. We believe that the dual nature of a “living laboratory” approach where a testbed is also used by students in their everyday lives will inspire them to become more actively involved in this area of technology. From a more practical but nevertheless crucial perspective, we aim for a simple, open and autonomous participation model that will encourage our students, but also users from other communities, to adopt the system. We are looking for an approach that is easy to implement, setup, manage and extend,

with as little human intervention as possible; maintenance is neither something we like to nor can afford to spend many resources on. At the same time, one has to strike a balance between performance and flexibility, while keeping the cost of network deployment low. The latter is particularly important because we want to adopt a community-driven approach where each participant covers (at least in part) the costs of installation. We feel that this is necessary to guarantee survivability without relying on a constant inflow of funds, which is hard to achieve in practice, especially in a non-profit academic environment.

III. Network Model

Our approach is illustrated in Figure 1, showing an indicative configuration that comprises several stationary and mobile network elements with typical deployment options to support fixed terminal devices. It is motivated and described in more detail below. Due to the fact that a large part of the city is densely built with many tall buildings standing next to each other, network elements (labeled as ad hoc routers in Figure 1) must come primarily in the form of stationary devices installed on roofs and balconies in order to achieve better connectivity. Their primary role is to provide an IP-plug, which can be used to connect one or more local client devices to the rest of the wireless network. At the same time, they may serve as hot-spots providing IP connectivity to mobile devices in range. We also wish for mobile devices themselves to serve as active network elements.

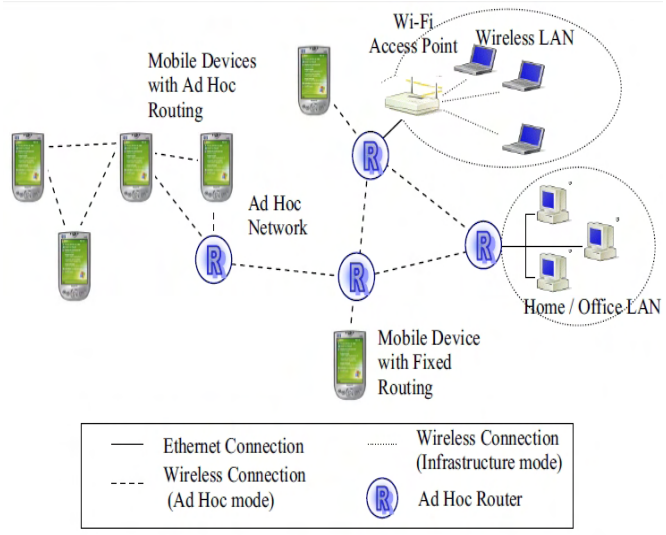


Figure 1. Network Architecture Model

This option can be used to eliminate the need for installing a stationary network element at home, and so that users may share part of the network's load even when on the move, if so desired, in accordance to a true community spirit. It should also be noted that mobile network elements can be particularly convenient for conducting on-site experiments of the type "what will happen if we place a new network element at location X?" for the purpose of teaching, demonstration, testing and configuration planning. A pure ad hoc and peer-to-peer networking approach was chosen to support this functionality. This is because we wish to be able to add new network elements without having to reconfigure the ones that have already been deployed. Even though such changes can be performed using a combination of remote configuration tools and scripting, they still require human intervention and assume that all network elements are up and running during the update process. Automatic adaptation when network elements are removed or do not respond is even more important. Failures are likely to occur occasionally, be it due to software glitches, hardware problems, power outages, or people resetting the equipment installed in their homes by accident. Moreover, in the case of mobile network elements topology changes are the rule rather than the exception thereby making manual reconfiguration practically impossible. In terms of end-user devices (terminals), we support two options: fixed and mobile terminals. In the first case, a device such as a personal computer or IP-enabled appliance connects to the ad hoc wireless network via a local area network. Several terminals can be connected to the network in this fashion, via the same network element. In the second case, a mobile device such as a laptop or handheld computer connects to the network via an ad hoc wireless connection to any nearby network element of the system. Mobile terminals may optionally assume the role of network elements with routing functionality, thereby enhancing robustness and increasing the bandwidth of the system. The distinction between fixed and mobile terminals is at the level of the IP protocol layer software and transparent for the applications residing on terminals.

IV. Implementation of the Ad Hoc Wireless Router

The choice of the wireless technology to be used was "naturally" trivial, with 802.11b/g/n being the only practical option given its wide industrial adoption, high-speed potential, and support for both infrastructure and ad hoc operation modes. But the quest for a suitable router platform proved somewhat more adventurous. As already mentioned, we desire a platform that can be programmed, ideally from scratch, in order to have as much development flexibility as possible. We also do most of our software development work in the Linux system environment, the open-source mentality and strong community support being the main reasons for this preference. Another restriction was that the device should fit in a reasonably sized weather-proof container without running into problems due to low external temperatures or –far more likely– overheating during summer. With PC platforms being out of the question due to their high price, big size and indoors-only deployment scope, our initial approach was to search for an embedded system board with interfaces that can be used to add Wi-Fi cards. A similar approach is followed in other projects, such as the WAND project [7]. We found that the WRAP Wireless Router Application Platform, designed by Pascal Dornier (PC Engines GmbH), met our requirements. Consequently we built a prototype wireless router using a WRAP board and two Prism 2.5-based Mini PCI wireless interface cards, running embedded Linux (Figure 2a).

Figure 2a. WRAP Linux router

Figure 2b. TP-Link TL-WR1043ND *Linux router*

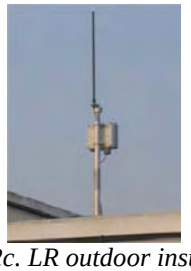


Figure 2c. LR outdoor installation

Unfortunately, although we were satisfied with its performance, the overall cost of the package turned up to be more than what a typical student could afford. Changing our strategy, we turned our focus on the readily available wireless products in the market and searched for platforms that would meet our requirements. We decided for the TP-Link TL-WR941ND and TL-WR1043ND broadband wireless routers (Figure 2b), which come with a Linux-based firmware and source code published under the General Public License. The TL-WR941ND device (v 2.0) features 4MB Flash memory, 32MB RAM, Atheros AR9132@400MHz, 10/100/1000 Ethernet controllers and an TP-Link TL-SG1008 8 port 10/100/1000 switch. The switched ports are separated into two different Virtual LANs (vlans), one for the LAN interface of the router comprising of 4 switched ports, and another intended for external Internet connectivity (WAN port).

The enhanced TL-WR1043ND version 1.8 features 8MB of Flash memory, 32MB RAM, Atheros AR9132@400MHz. Both devices have on-board interfaces for connecting to indoor or outdoor antennas. Experimentation in the lab with TL-WR941ND and OpenWRT proved that this particular hardware/software combination (henceforth referred to as the LR platform) met our requirements, and was therefore chosen as a basis for developing our own ad hoc wireless router and community network. Routing of IP packets in the ad hoc network is dynamic, as a function of topology changes. We decided to use the Optimized Link Source Routing (OLSR) protocol, due to its inherent flexibility, scalability potential and reduced administration overhead characteristics [4][5]. OLSR is a proactive, table-driven routing protocol based on the issue of Multipoint Relays (MPRs). It is considered to be well suited for large and dense networks with low mobility rates [6]. We tested a number of different solutions to choose a stable and highly configurable implementation.

The internal router configuration is as follows. The LAN interface, which is attached to the 5-port switch, is identified as eth0. The switch separates the ports into two Virtual LANs, one comprising of 4 ports (LAN segment), and another assigned to the WAN port. The LAN segment is addressed as vlan0 while the WAN port is identified as vlan1. Finally, the Wireless interface is identified as eth1. In the default configuration, eth1 and vlan0 are bridged together. We decided to separate these interfaces because we assume that devices connected to the router's LAN are terminals that do not (have any reason to) support ad hoc routing. A diagram that illustrates the router setup approach is shown in Figure 3.

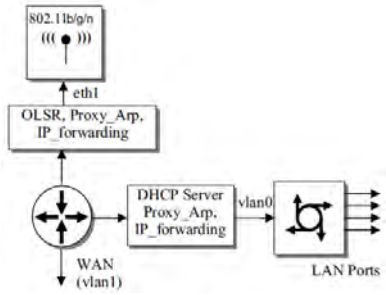


Figure 3. The internal router configuration

V. Address Management

An important requirement of our network is that routers should be able to join, quit and reappear, perhaps at another location, at arbitrary times, without the individual network elements having to be explicitly reconfigured by an administrator. At the level of IP, this means that address assignment and routing should be performed dynamically and in a location-independent manner.

On switched local area networks, it is possible to automatically assign IP addresses to a priori unknown hosts via the Dynamic Host Configuration Protocol (DHCP) [8]. This requires that the DHCP server is reachable in one hop from the requesting host, which is definitively not the case in the envisioned network. One solution is to use DHCP relay [9] to forward requests to a central server. However, this would make our system fragile as it introduces multiple single points of failure, namely the server itself plus every ad hoc router that is used to create the network path between the host and the server. Decentralized IP address assignment algorithms have been proposed to address this problem in a distributed fashion, including [10] [11] [12] [13], yet no implementation seems to be readily available.

Since we wanted to start the deployment we decided to adopt a static IP addressing plan, as a temporary and easy to implement solution. A registry of IP addresses and subnet addresses was created for the ad hoc network, and each router is statically assigned an IP address which uniquely identifies it on the network. Furthermore, each router is given a subnet network address for 32 IP addresses to be used in the context of its local LAN. The first address is assigned to the router's LAN interface, which leaves room for 29 IP-enabled devices that can be connected to the LAN interface of each router. Along this scheme, each router features a DHCP server that

manages a pool of 29 addresses and provides automatic IP address assignment for the LAN interface, where we also activated Proxy ARP and IP Forwarding using the proc file system. Hence devices that connect to a router through the LAN interface are automatically configured. In turn, OLSR running on the WLAN (vlan0) interface is configured to advertise the subnet corresponding to the LAN segment to the rest of the system.

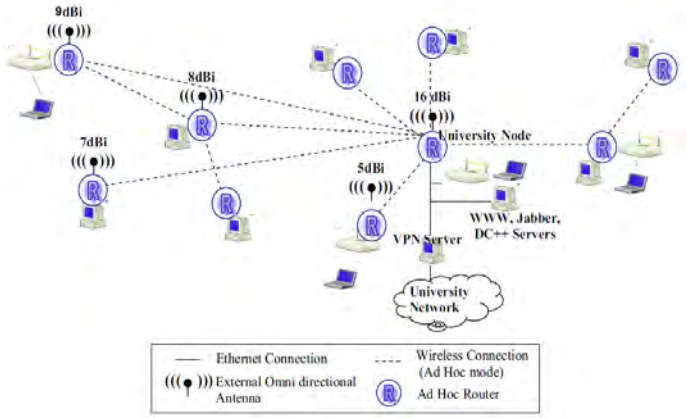


Figure 4. The Ad Hoc wireless network testbed covering.

VI. Conclusion

The network model is kept open so that many different communities are free to join, contribute to and exploit it. However this does not imply that all participants should be allowed to access the same set of services. Public Internet access, via the University's infrastructure, is a typical example. Of course, nor should any participant be able to retrieve account names and passwords transmitted over the ad hoc network. Security and access control is implemented using a Virtual Private Networking (VPN) approach. VPN technologies provide an efficient solution with central control of user access mechanisms. Our implementation employs the Point-to-Point Tunneling Protocol (PPTP) [14] with MPPE128 encryption [15], which is supported by almost every operating system including those running on most handheld devices. The VPN server is installed on a dual-CPU (2400MHz) personal computer with 2048MBs of RAM. Linux was the operating system of choice and PoPToP was used as the PPTP server [16]. Measurements performed on the local network showed a packet encryption/decryption rate of 950Kbytes/sec, which we consider sufficient for the purpose of initial deployment. The VPN server is equipped with two Ethernet interfaces. The first interface connects to the University LAN through which access to the Department's servers and public Internet is provided. The second interface connects to the LAN interface (vlan0) of an ad hoc router that connects the VPN server to the community wireless network. The network segment between this particular router and the VPN server was assigned an individual network address and the OLSR daemon on the router was configured to advertise the corresponding network address. Using the IP Masquerade feature provided by the Linux kernel (IP Tables) we configured the VPN server to perform Network Address Translation (NAT) between the computers attached to the VPN and the University network. NAT was necessary since the IP address pool of the University cannot be used for the purposes of the ad hoc network for reasons of security and scalability. With this setup, students and faculty may connect to and access the University network from any terminal connected to any ad hoc router. We have successfully tested Linux, Windows XP/ Vista and Windows 7 PPTP implementations over the wireless network.

References

[1] Lee Breslau et al., “Advances in Network Simulation”. IEEE Computer, 33 (5), May 2000, pp. 59-67.
[2] S.Y. Wang, C.L. Chou, C.H. Huang, C.C. Hwang, Z.M. Yang, C.C. Chiou, and C.C. Lin, "The Design and Implementation of the NCTUns 1.0 Network Simulator", Computer Networks, Vol. 42, Issue 2, June 2003, pp. 175-197.
[3] Daniel Aguayo, John Bicket, Sanjit Biswas, Glenn Judd, Robert Morris, “Link-level Measurements from an 802.11b Mesh Network”, SIGCOMM 2004, Aug 2004
[4] T. Clausen et al., “Optimized Link State Routing Protocol”, IEEE INMIC Pakistan, 2001.
[5] T. Clausen, P. Jacquet, “Optimized Link State Routing Protocol (OLSR)”, Request for Comments 3626 (Experimental), Network Working Group, Internet Engineering Task Force (IETF).
[6] P. Trakadas et al., “Efficient Routing in PAN and Sensor Networks”, ACM Mobile Computing and Communications Review (MC2R), Vol. 8, Num.1, January 2004, pp 6-9.
[7] Stefan Weber, Vinny Cahill, Siobhan Clarke and Mads Haahr. Wireless Ad Hoc Network for Dublin: A Large-Scale Ad Hoc Network Test-Bed . E RCIM News, 2003, vol. 54.
[8] R. Droms, “Dynamic Host Configuration Protocol”, Request for Comments 2131 (Standards Track), Network Working Group, Internet Engineering Task Force (IETF).
[9] Matthew J. Miller, William D. List, Nitin H. Vaidya, “A Hybrid Network Implementation to Extend Infrastructure
Reach”, Technical Report, January 2003.
[10] MansoorMihsin and Ravi Prakash, “IP Address Assignment in a Mobile Ad Hoc Network”, MILCOM 2002.
[11] Nitin H. Vaidya, “Weak Duplicate Address Detection in Mobile Ad Hoc networks”, Proceedings of 3rd International Symposium on Mobile Ad Hoc Networking & Computing, 2002.
[12] S. Nesargi and R. Prakash, “MANETconf: Configuration of hosts in a mobile ad hoc network”, INFOCOM 2002.
[13] C.E. Perkins, E.M. Royer, and S.R. Das, “IP address autoconfiguration for ad hoc networks”, Internet Draft, IETF
Working Group MANET, (Work in Progress), July 2000.
[14] K. Jamzeh et al., “Point-to-Point Tunneling Protocol (PPTP)”, Request For Comments 2637 (Informational), Network Working Group, Internet Engineering Task Force (IETF).
[15] G. Pall, G. Zorn, “Microsoft Point-to-Point Encryption (MPPE) Protocol”, Request For Comments 3078 (Informational), Network Working Group, Internet Engineering Task Force (IETF).
[16] PoPToP Project, URL: http://www.poptop.org/

[1] Hristofor Ivanov, St. Cyril and St. Methodius University of Veliko Turnovo, Veliko Turnovo 5000, Bulgaria, E-mail:globul@mail.bg
[2] Miroslav Galabov, St. Cyril and St. Methodius University of Veliko Turnovo, Veliko Turnovo 5000, Bulgaria, E-mail:lexcom@abv.bg

An Approach to Optimization of the Links' Load in the MPLS Domain

Veneta Aleksieva[1]

Abstract –**The problem with the optimization of MPLS (Multi Protocol Label Switching) network is discussed in terms of finding such label switching paths (LSP) for all flows in the network, which would provide minimal load on the links in it, as collaboration of the protocol of the third layer (IS-IS or OSPF) and MPLS is reported. The proposed analytical model is reduced to optimization problem which is solved using linear optimization. Researches have been made for this algorithm and the results are presented.**

Keywords – **MPLS, optimization problem, load minimizing on the links.**

I. INTRODUCTION

Nowadays, network services are proposed to customers in an environment of strong competition, so in general it is not possible and justified fully or multiple booking even at the critical points because it would significantly increase the cost of network services. Therefore, an important activity related with the availability of the network services is the restoration of services in their planned quality after accidents and failures. This rises the necessity of using network protocols, allowing application of quick and automated solutions for network recovery after failures, as the most popular solution nowadays for the NGN networks is the MPLS technology [2],[8].

This IP/MPLS network offers universal transport media, providing the need of reliability and QoS, which is achieved by the two methods – recovery and protection. The recovery takes less network resources than the defense and it is used more often by ISPs. But which link will be alternative in case of primary link' fails is a subjective decision, not always optimal and in some cases after the switching of traffic on backup connection, it appears that it has not sufficient capacity to absorb all the traffic and some packets are dropped and not delivered. Therefore solutions, in which the alternative links must have sufficient capacity to be possible to minimize packet's loss, are being searched. However, these solutions must take into account the willingness of ISPs to minimize the cost of money and the amount of occupied resources.

II. RELATED WORKS

In recent years, several studies focus on the recovery path in the network because users need paths with low latency, high throughput and a small number of lost packets [1],[3],[4],[5],[9],[10],[11],[12]. Many critical applications require greater flexibility than the one provided by the current Internet routing. Due to these facts, some authors [5],[7],[11] use configuration files of the network devices to investigate and remove anomalies in the traffic and the reasons, which are rising the recovery of the path. Other authors [3], [12] propose solutions to overcome this problem by predicting the behavior of network traffic. In some cases, however, route recalculation can take minutes to switch traffic on it. During this period, loss of many packets is possible. In addition, periodic failures can cause repeated rerouting, causing routing instability. To overcome this problem in previous years, researches of the ISP were focused on routing algorithms with multiple paths routing [9],[10]. Studies show that it is possible to be provided a set of alternative paths in scalable network as with protocols from third layer of the OSI model, so as with MPLS. In most cases, the searched solutions are targeted at already built and operating networks, but a little focus on the first, off-line phase of pre-defined LSPs in the network.

Several authors propose different analytical models for the evaluation of quality indicators for MPLS network [13], the viability of the MPLS network [6], evaluation of service quality in the MPLS network (in terms of average delay and number of lost packets for different classes of services)[14],[15], structural synthesis of the MPLS network with constraint (quality of service to achieve a minimum price)[16]. Typical for these studies is that they examine the behavior of MPLS in an ideal environment in which accidents do not occur and does not require switching a packet of back-up links. The authors report latency factors in MPLS switches and distribute flows simultaneously on two factors - the average latency and the number of packet loss.

One factor not considered in the above solutions is the chosen by the administrator of the network policy for the classification of traffic. Another factor not considered in these decisions is the retain of optimal spare capacity for each port, which if necessary, by using this link as an alternative would absorb both peak traffic on this line and the temporarily transferred from other link traffic on it. The capabilities of MPLS port should include adequate reserve capacity above the average peak load in order to take on short peaks in traffic. However, if the spare capacity is too large, the higher expenses are not justified. Therefore, the revision of the average peak load and comparing it with the capacity of the port can provide opportunities to reduce its size. Different organizations follow different principles, but as a general rule of practice is considered the average peak load to be 60-70% for MPLS port (or up to 30-40% if the port is part of the backup configuration)[17]. These percentages are obtained based only on empirical experience.

[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: VAleksieva@tu-varna.bg

Routers in the MPLS network receive packets not only with MPLS labels. They exchange information for the building of correct layer 3 routing tables. Meanwhile, with signaling protocols, they exchange information for MPLS labels with which to manage the routing of LSP. Part of the IP packets can be routed without labeling based on the routing decisions of layer 3.

The most frequently used signaling protocol is RSVP-TE , in which's configuration are set as the primary and back-up links. In determining the LSP is included the priority of the flow, as it gives the new LSP the right to replace the existing one if the new flow has higher priority than the old. This allows high-priority LSP to be set optimally, whether there are existing reservations on the links on this LSP, but only if those reservations have lower priority. When needed a rerouted LSP, LSPs with higher priority are more likely to find an alternate route than those with lower priority. But these packages in lower priority flows may be lost due to insufficient capacity of the connections.

As an indicator of the network congestion is used the maximum load of the link. When the static routing is used and the traffic is linearly increasing, the minimization of this parameter will provide the maximum linear growth of the traffic before its need of rerouting.

III. Definition of the Problem

The purpose of this study is finding the optimal primary and backup LSPs in the MPLS domain to achieve minimal loaded in links and minimal loss of traffic at a failure.

Let G = (N, E) is a finite undirected graph, representing the MPLS network, where N is the number of nodes (routers), E is the number of edges(links). On each of the edge (i,j)$\in$ E is assigned a weight c_{ij}, representing link capacity in Mbps. Each node can be both source and destination of traffic. Source-node s(f) and the receiving node t(f) may not have direct contact with each other, but for the purposes of the model is necessary to exist at least one route between them. Let F is the set of source-receiver pairs and d^f_{st} is the maximal requested by the administrator traffic that passes by the pair $f_{st} \in$ F. Apparently

$$d^f_{st} \geq 0 \quad (1)$$

For d^f_{st} also can be set the number u_{st}, which limits the bandwidth of the edge:

$$0 \leq d^f_{st} \leq u_{st}. \quad (2)$$

Let X(m,k) is the routing matrix whose elements are rational numbers in the interval [0,1]. They represent part of the flow of f_{st} , which is routed by the protocol of the third layer (IS-IS or OSPF) on the edge (i, j). The rows of the matrix correspond to the edges and the columns of source-receiver pairs. The sources generate a flow which flows in the direction of edges, possibly through the intermediate points and gets to the receivers. Quantitative measure of the total flow that will run on the network is determined by these requirements.

The load $H^{(k)}_{\Sigma}$ on edge (i,j) is presented as a total sum of the loads $h^{(k)}_{ij}$ of all the flows in this edge, where k is priority of flow:

$$H^{(k)}_{\Sigma} = \sum_{j=1}^{n}\sum_{i=1}^{n} h^{(k)}_{ij} \quad (3)$$

For the purposes of the study are examined the following variables of the model:

u_{max}: maximum load on all arcs

is^f: the part of the requested traffic d^f, routed from IS-IS/OPF, [bit/s]

w^f_{ij}: the part of the requested traffic d^f, transmitted by LSP, and routed from MPLS

$\overline{T}_{cp}$: average time of delay in MPLS network flow with priority k as at the links, and in the routers

LP_k – probability of packet loss with priority k, which is calculated as:

$$LP_k = 1 - \prod_{s,t \in E} (1 - P^{(k)}_{los(s,t)}) \quad (4)$$

Where $P^{(k)}_{lose(s,t)}$ - the probability of all virtual channels, on which is configured to transmit the flow with priority k, are filled with a capacity of flows with higher priority or equal to this (i.e. from 0 to k-1)

The problem of finding optimal LSPs in MPLS network, minimizing the load on the links between nodes is formulated as follows: it is necessary to find such a distribution of flow $f_{ij}^{\ q}$ thus minimizing the maximum flow μ_{ij} based on flows and on priorities , collaborate working with the third layer protocol (IS-IS or OSPF) and MPLS, while load edges is minimized. This optimal distribution of flows is expected to increase the spare capacity of links.

Solving such a problem in structural synthesis of the network is available in [16], but their solution is only to newly-developed network. As a final result is the lowering of its price. Solving this problem in terms of planning, but non-constructed network is expected to achieve the same result. Unlike the quoted decisions, this proposal can be applied to a functioning network with real traffic. When a failure giving rise to recovery at peak times of load occurs, the increasing of spare capacity on links (in particular those used as backup) will reduce congestion and minimize the packet loss of the lower priority traffic.

Then the problem of finding optimal LSPs in the MPLS network, minimizing the load on the links between nodes can be expressed analytically as following:

Target function: $min\ u_{max}$ (5)

Restrictive conditions:

$$\sum_{f \in F} is^f x^f_{ij} + \sum_{f \in F} w^f_{ij} \leq u_{\max c_{ij}}, \ (i,j) \epsilon E \quad (6)$$

$$\sum_{(i,j)\in A} w^f_{ji} - \sum_{(i,j)\in A} w^f_{ij} = \begin{cases} -d^f + is^f, if\ i = s(f) \\ d^f - is^f, if\ i = t(f) \\ 0, in\ the\ other\ cases \end{cases}, \ i \epsilon N, f \epsilon F \quad (7)$$

$$LP_k \leq LP_{k,fixed} \quad (8)$$

$$T^{\kappa}_{cp} \leq T^{\kappa}_{fixed} \quad (9)$$

$$w^f_{ij} \geq 0 \ , (i,j)\ c\ A,\ f \epsilon F \quad (10)$$

$d^f \geq is^f \geq 0$, $f \epsilon F$ (11)

Objective function reflects the maximum value of the load that has to be minimized. When solving the model for each f is appearing an optimal set of paths in G from source node s(f) to the receiving node t(f), i.e. optimal number of LSPs.

IV. Solving the Optimization Problem

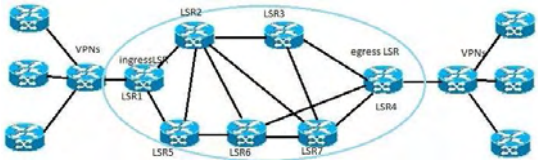


Fig. 1. Experimental MPLS Network

Restrictions on the amount of data in the Solver of Excel (up to 200 cells), exceed the maximum amount of currently existing networks (largest MPLS backbone network is realized in Canadian Telecom - 100 LSR, for comparison, in Bulgaria consists of 50 LSR). It is presented a suggested solution based on the experimental network, which is presented on the fig. 1.

Cost of all links is 1000Mbps and purpose is to reach to less than 60% of each link capacity (600Mbps). Starting matrix is presented on Fig. 2. On Fig. 3. and Fig.4 are presented target cell, constraints, cells which be changed, options. On Fig. 5. is presented final matrix with optimal decision. On Fig. 6. and Fig. 7. are presented answer report.

	LSP1-2	LSR1-3-2	LSP1-3	LSP1-2-3	LSP2-3	LSP2-1-3	LSP2-1	LSP2-3-1	LSP3-1	LSP3-2-1	LSP3-2	LSP3-1-2	sumHij
lsr1-lsr2	200	300	0	0	0	0	0	0	0	0	0	0	650
lsr1-lsr3	0	0	150	350	0	0	0	0	0	0	0	0	600
lsr2-lsr3	0	0	0	0	350	150	0	0	0	0	0	0	950
lsr2-lsr1	0	0	0	0	0	0	250	250	0	0	0	0	600
lsr3-lsr1	0	0	0	0	0	0	0	0	300	200	0	0	650
lsr3-lsr2	0	0	0	0	0	0	0	0	0	0	400	100	900
													4350

Fig. 2. Part of Starting Matrix

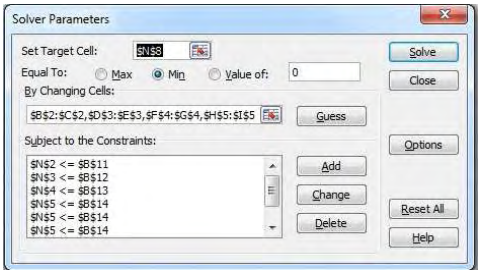


Fig. 3. Solver Parameters

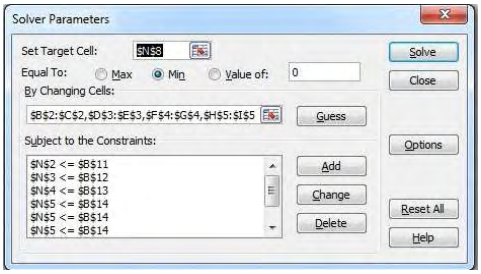


Fig. 4. Solver Options

	LSP1-2	LSR1-3-2	LSP1-3	LSP1-2-3	LSP2-3	LSP2-1-3	LSP2-1	LSP2-3-1	LSP3-1	LSP3-2-1	LSP3-2	LSP3-1-2	sumHij
lsr1-lsr2	500	0	0	0	0	0	0	0	0	0	0	0	500
lsr1-lsr3	0	0	500	0	0	0	0	0	0	0	0	0	500
lsr2-lsr3	0	0	0	0	500	0	0	0	0	0	0	0	500
lsr2-lsr1	0	0	0	0	0	0	500	0	0	0	0	0	500
lsr3-lsr1	0	0	0	0	0	0	0	0	500	0	0	0	500
lsr3-lsr2	0	0	0	0	0	0	0	0	0	0	500	0	500
													3000

Fig. 5. Part of final matrix

Microsoft Excel 12.0 Answer Report
Worksheet: [Book2]Sheet2
Report Created: 3/13/2012 6:54:08 AM

Target Cell (Value Of)

Cell	Name	Original Value	Final Value
$N$8	sumHij	4350	3000

Adjustable Cells

Cell	Name	Original Value	Final Value
$B$2	lsr1-lsr2 LSP1-2	200	500
$C$2	lsr1-lsr2 LSR1-3-2	300	0
$D$3	lsr1-lsr3 LSP1-3	150	500
$E$3	lsr1-lsr3 LSP1-2-3	350	0
$F$4	lsr2-lsr3 LSP2-3	350	500
$G$4	lsr2-lsr3 LSP2-1-3	150	0
$H$5	lsr2-lsr1 LSP2-1	250	500
$I$5	lsr2-lsr1 LSP2-3-1	250	0
$J$6	lsr3-lsr1 LSP3-1	300	500
$K$6	lsr3-lsr1 LSP3-2-1	200	0
$L$7	lsr3-lsr2 LSP3-2	400	500
$M$7	lsr3-lsr2 LSP3-1-2	100	0

Fig. 6. Optimal Decision-Report

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$N$2	lsr1-lsr2 sumHij	600	$N$2<=$B$11	Binding	0
$N$3	lsr1-lsr3 sumHij	600	$N$3<=$B$12	Binding	0
$N$4	lsr2-lsr3 sumHij	600	$N$4<=$B$13	Binding	0
$N$5	lsr2-lsr1 sumHij	600	$N$5<=$B$14	Binding	0
$N$5	lsr2-lsr1 sumHij	600	$N$5<=$B$14	Binding	0
$N$5	lsr2-lsr1 sumHij	600	$N$5<=$B$14	Binding	0
$N$6	lsr3-lsr1 sumHij	600	$N$6<=$B$15	Binding	0
$N$7	lsr3-lsr2 sumHij	600	$N$7<=$B$16	Binding	0

Fig. 7. Optimal Decision-Report

V. Experimental Results

For the purpose of the experiment each node has incoming traffic as shown on the unit LSR1 and it is entered by 3 types of traffic with different priority (0,1,2) and for these three types of traffic output node can be any node, as shown node LSR4. For the experiment will be considered only incoming flows LSR1, reaching and output of the network LSR4. These three types of traffic pass through the same channels (LSR1-LSR2-LSR3-LSR4). By increasing the intensity of traffic with priority 0 remains less spare capacity for other types of traffic on this road, which at one point leads to the arrest of the rest packets in queues of the nodes in order to leave capacity for the high priority traffic.

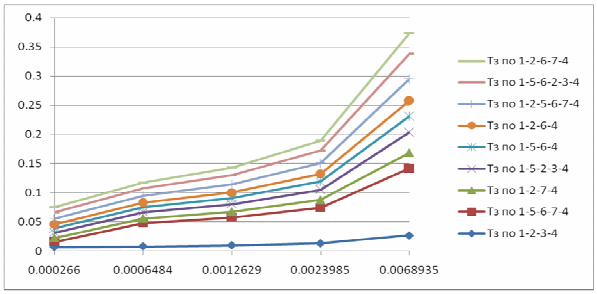


Fig. 8. Average Time to Pass a Package from Traffic with Priority 2 before the Optimization of the Load Experimental MPLS Network

During unloading the path, however (eg routing of lower priority traffic without overlapping arcs on the path LSR1-LSR5-LSR6-LSR4), a balanced load of links in the network is given, which leads to optimal loading of links in the network and shorter retention times in it.

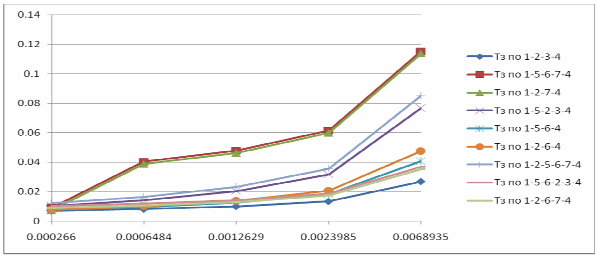


Fig. 9. Average Time to Pass a Package from Traffic with Priority 2 after the Optimization of the Load Experimental MPLS Network

Thus, Fig. 8. shows the average time for passage of packets of the traffic in the experimental lower priority MPLS network when it is not achieved optimization of load links, and Fig. 9. - after achieving the optimization. It is clear that when achieving a balanced load, average retention time of the longest roads is decreased by almost 3 times when the network is intensive load and is almost the same with the minimal selected network's load and at the shortest routes.

VI. Conclusion

The problem of optimizing the MPLS network is discussed in terms of finding such LSPs for all flows in the network to provide the minimum load on any links in it and is proposed an analytical model for solving this task. Account is taken of the collaborate work of the the third layer protocol (IS-IS or OSPF) and MPLS, so that the load on the links between the nodes in the network would be minimized. A solution is presented with Solver at Excel. There have been experimental studies of that algorithm and the results are presented.

References

[1] Anand A., C. Muthukrishnan, A.Akella, R.Ramjee, Redundancy in Network Traffic: Findings and Implications, Proceeding of SIGMETRICS, ACM NY,USA, 2009, pp.37-48

[2] L. Andersson, R. Asati, Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field, February 2009, http://tools.ietf.org/html/rfc5462

[3] Cunha I.,R.Teixeira, D. Veitch, C.Diot, Predicting and Tracking Internet Path Changes, Proceeding of SIGCOMM 2011, Computer Comunication review, vol.41, Number 4, october 2011, pp.122-133

[4] Gupta A., A. Akella, S. Seshan, S. Shenker, J,Wang, Understanding and Exploiting Network Traffic Redundancy, Technical Report 1592, UW-Madison, April, 2007

[5] Mahimkar A., H. Song, Z.Ge, A. Shaikh, J.Wang,J.Yates,Y.Zhang, J.Emmons, Detecting the preformance impact of upgrades in large operational networks, Proceeding of ACM SIGCOMM 2010, pp303-314,NY,USA,2010

[6] Mossavari M., Y.Zaychenko, The analysis and optimization of survivability of MPLS Networks, International Journal "Information Theories&Applications",vol.15,2008

[7] Phillipa G., J. Navendu, N. Nachiappan, Understanding Network Failures in Data Centers: Measurement, Analysis, and Implications, Proceeding of SIGCOMM 2011, Computer Comunication review, vol.41, Number 4, october 2011, pp.350-361

[8] Rosen, E., A. Viswanathan, R. Callon. Multiprotocol Label Switching Architecture. RFC 3031, IETF, January, 2001, http://www.ietf.org/rfc/rfc3031.txt

[9] Sahar A. El Shazely, Omayma A. M. Abdel Mohsen and Khaled Shehatta, Enhancing MPLS Network Fault Recovery Using P-Cycle with QoS Protection, IEEE CCC Code: 1-4244-1430-X/07/$25.00 ©2007 IEEE

[10] M. Tacca, Kai Wu, Andrea Fumagalli, Jean-Philippe Vasseur, Local Detection and Recovery from Multi-Failure Patterns in MPLS-TE Networks, IEEE ICC 2006 proceedings, vol.9, 11-15.06.2006, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4024203

[11] Turner D., K.Levchenko, A.Snoeren, S. Savage. California fault lines:understanding the causes and impact of network failures. Proceeding of ACM SIGCOMM 2010,pp 315-326, NY,USA,2010

[12] Zohar E., I. Cidon, O. Mokryn, The Power of Prediction: Cloud Bandwidth and Cost Reduction, Proceeding of SIGCOMM 2011, Computer Comunication review, vol.41, Number 4, october 2011, pp.86-97

[13] Зайченко, Е.Ю., „Комплекс моделей и алгоритмов оптимизации характеристик сетей с технологией MPLS", Системні дослідження та інформаційні технології, 2007,№4,с.58-71

[14] Зайченко Ю.П., Е.Зайченко, Инструментальный комплекс алгоритмов и программ оптимального проектирования сетей с технологией MPLS, ITHEA, с.94-100, 2009

[15] Зайченко, Е.Ю., Зайченко Ю.П., Лавринчука А.Н., Инструментальный комплекс алгоритмов и программ оптимального проектирования сетей с технологией MPLS, 2010, http://it-visnyk.kpi.ua/?page_id=1266&lang=ru

[16] Зайченко, Е.Ю., Зайченко Ю.П., Структурный синтез сетей с технологией MPLS, International Book Series "Information Science and Computing", ITHEA,2010,с.83-89

[17] http://www.infoweek.bg/tab_it/it.php?show_category=2&show_sub_category=2&id=666

Performance Study of Virtualization Platforms for Virtual Networking Laboratory

Hristo Valchanov[1]

***Abstract* – Virtualization is a modern software technology that quickly spreads in various areas of the IT sector. However, its use in higher education is still insufficient widespread. Virtualization platforms provide flexible and efficient utilization of existing infrastructure. They offer the capability to integrate advanced topics into courses in a way that gives students control so that they can perform hands-on activities that would be infeasible on physical computers. Due to, the teaching will become more adequate to the rapidly changing world of IT industry. This article presents a comparative analysis of performance of virtual platforms for building a virtual networking laboratory.**

Keywords – **Virtualization, Virtual Platforms, Virtual Laboratory, Higher Education.**

I. INTRODUCTION

Virtualization is a modern software technology that quickly finds spreads in various areas of the IT sector. However, its use in higher education is still insufficient widespread. The traditional method of teaching requires students to attend the laboratory in individual laboratory rooms. University laboratories typically involve a fixed number of computers and network equipment. Each computer has installed a separate operating system and software according to the material. The existence of different courses in different semester, requires the installation and maintenance of various software packages. The software packages have specific requirements to the capabilities of the hardware, which limits their use only in certain computer labs. However, these computers and software require specific engineering support, and need to invest in new hardware.

Using virtualization platforms (VPs) can resolve a number of existing problems. This can be realized in several ways. Firstly, these platforms provide cost-efficient environments for training and research in the form of virtual laboratories. Second, the use and sharing of hardware with a different purpose can reduce the need for investment in new equipment. At the same time VPs will reduce the cost of equipment maintenance in terms of reinstalling operating systems and software packages.

Virtualization allows providing a single computer to every student. The implementation of remote access to virtual infrastructure will reflect in the quality of teaching, as students will be able to access virtual laboratories at any time. This will compensate for the insufficient number of computer labs and workstations.

[1]Hristo Valchanov is with the Department of Computer Science and Engineering at Technical University of Varna, 1 Studentska Str., Varna 9000, Bulgaria, E-mail: hristo@tu-varna.bg.

The paper presents a comparative performance analysis of virtualization platforms, suitable to build virtual network university laboratories.

II. VIRTUALIZATION TECHNOLOGIES

The virtualization is defined as an „abstracting a computer's physical resources into virtual ones with the help of specialized software" [1].

The virtualization platform virtualizes or transforms the hardware resources of very popular x86-based computers, including CPU, RAM, hard disks and network controllers. It creates a fully functional virtual machine (VM) that can run its own operating system and applications just like a real computer [2]. Each virtual machine contains a complete system, eliminating potential conflicts. Virtualization works by inserting a thin layer of software directly on computer hardware or the host operating system. This software is a virtual machine - monitor or hypervisor which allocates hardware resources dynamically and transparently. Multiple operating systems can run simultaneously on one physical computer and share hardware resources with each other. By encapsulating the entire machine, including CPU, memory, operating system and network devices, virtual machine is fully compatible with the all standard x86 operating systems, applications and drivers. We can run multiple operating systems and applications at the same time on one computer, and each user can have access to the resources he needs and when he needs.

A. Virtualization Environments

There are two main types of virtualization: hosted virtualization and bare-metal environments (Fig.1).

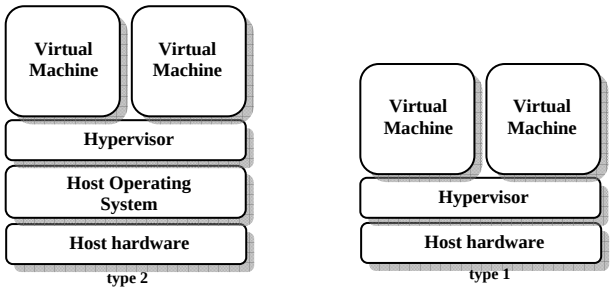


Fig.1 Virtualization environments

In hosted environment the hypervisors (type 2) are software applications running within a conventional operating system environment of the host. The hypervisor controls the resources that are allocated from the operating system on the lower level. This type of hypervisors are generally used in systems where there is a need for different input/output devices that

can be maintained by the host operating system and client systems with low efficiency. Examples of such hypervisors are: Parallels Workstation, Microsoft Virtual Server, VMware Server and VMware Workstation.

In bare-metal environments the hypervisors (type 1) are software systems that run directly on the host hardware. They use a hardware control for monitoring the guest operating system. These types of hypervisors are the preferred approach to virtualization because they are running directly on hardware, thus achieving higher efficiency and performance. Examples of such hypervisors are: Citrix XenServer, VMware ESX and Microsoft Hyper-V.

B. Types of virtualization

Depending on the image of the software, virtualization software can be divided into two categories: server virtualization and desktop virtualization.

The server virtualization (Fig.2a) allows consolidation of multiple servers on a single high-performance server machine. Thus can reduce the number of physical servers and hence the cost of maintaining the equipment and the power. The hypervisor isolates the individual VMs, thus protecting against improper interference and changing configurations, processes and other resources. Examples of server virtualization are: VMware ESXi, Microsoft Virtual Server and Xen Server Enterprise.

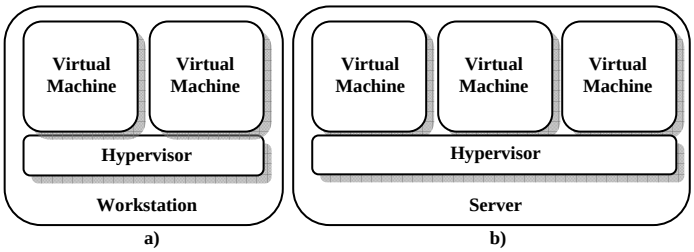


Fig.2 Desktop (a) and server (b) virtualization

Desktop virtualization (Fig.2b) allows virtualization on a desktop OS. One or more VMs can run on desktop machine. The virtual machine accesses resources via hypervisor. Desktop virtualization allows use of existing computers without the need to buy more powerful and expensive servers. Examples of such type of platforms are: Microsoft Virtual PC, Oracle Virtual Box, VMWare Workstation, Xen Server (Free Ed).

III. REQUIREMENTS FOR VIRTUAL LABORATORY

A virtual laboratory is a tool to which students have a remote access via the Internet and they use to conduct specific laboratory tasks [3, 4]. The choice of virtualization platform is determined by the nature of training. In Table 1 are shown the training courses related to network technologies teaching in Department of Computer Science at Technical University of Varna.

The table shows the need of different OS and different software packages. The choice of virtualization platform must consider the following factors:

- The performance of the equipment;
- The cost of the virtualization platform;
- The maintenance of devices and software.

TABLE I
NETWORKING COURSES

Course name	Operating system
Computer networks	Linux
Network infrastructures	Windows Server 2008
Network administration	Linux
Distributed programming	Linux
System Administration	Windows Server 2008

A. The performance factor

The power of the used computer equipment is essential for the effectiveness of training. Modern software packages bring ever greater requirements on computing resources. Unfortunately, the subsidies for university laboratories are insufficient to purchase the necessary equipment. From this perspective, the best solution is to use existing desktop PC computers in laboratories instead of buying powerful servers. The creating of virtual network infrastructure requires efficient management of network resources with access to devices on a computer.

B. The virtualization environment cost

Using a cost-effective platform for teaching and learning process is a direct consequence of the problem with the universities subsidizing. The market offers a wide variety of virtualization platforms with different status of use.

A much better solution is offered by the VMWare is an ESXi server with numerous features. Unfortunately its price model is not acceptable for the realization of our intentions. Several tools are provided with cost from $5000 to $12,000. Some certain features are shareware. Despite the rich features of this platform, the cost limits its use for our purpose.

The limited resources of desktop computers (in comparison with those of the server machine) are a prerequisite for choosing a platform for desktop virtualization. The VMWare offers VMWare Workstation [5] with enough features, but its use is shareware. This will need reinstallation after the trial period, which is not desirable to break the learning process.

Offered by Microsoft Virtual PC [6] is freeware, but requires installation under Windows, with is very big limitation.

Another representative of desktop virtualization environments is Oracle VM Virtual Box [7]. It is also freeware and can be installed under different OS, supporting multiple guest OS.

A possible solution is the use of Xen platform [8]. Apart from being distributed under the GPL, an important advantage is that it allows starting and managing virtual machines from two popular types of virtualization: paravirtualization and full virtualization. Xen also supports multiple guest OS.

C. Maintenance

Maintaining a virtual laboratory infrastructure is vital to the effectiveness of the learning process. There are different courses in different semesters and they require different

software packages. However, students work with administrative privileges, thus often leads to problems and crashes of the OS as a result of deleting configuration files or applications. From this perspective, the recovery system must be fast. These requirements are covered by the technique of snapshoot. With this technique it is possible to save a snapshot of the guest OS with fast recovery if necessary.

D. Comparison

Table 2 shows the base features of the presented platforms for virtualization. The presented platforms have similar features, so crucial for the selection test will be performance tests.

TABLE II
VIRTUALIZATION PLATFORMS FEATURES

Feature	VMWare Workstation	Virtual PC	Oracle VirtualBox	Xen
Product use	shareware	free	free	free
Host OS	Windows, Linux	Windows based only	Windows, Linux	Bare-metal
Snapshot	Yes	no	yes	yes
USB 2.0	Yes	yes	yes	yes
Virtual net	Yes	limited	yes	yes
Seamless	Yes	yes	yes	yes

On the basis of the features presented most closest to the criteria for virtual network laboratory platforms are VirtualBox and Xen. These two platforms were selected for testing experiments on real desktops computers in a real laboratory.

IV. EXPERIMENTAL STUDY AND RESULTS

The analysis of the performance of virtual machines is based on the benchmark tests performed respectively on Linux, and Windows Server 2008 in a computer laboratory. Performance tests were made on both platforms, respectively the Oracle VirtualBox and the Xen hypervisor. Software used for testing is: Unix Bench for Linux virtual machines and Performance Test 7 Passmark for Windows virtual machines. Tests were performed in parallel on the computer configuration model HP Desktop 500B with parameters: CPU Intel Core Duo E5800 3,2GHz, 2G DDR3 RAM, G1 Express chipset, Intel GMA graphics.

A. Performance test for Linux VM

The purpose of the benchmark UnixBench is to provide a basic indicator of the performance of Unix-based systems [9]. The set of tests were used to test various aspects of system performance. Their results are compared with values of the BASELINE System, i.e. baseline assessments which are used to calculate the index. Then these indices are combined to generate an overall index of the system. The UnixBench consists of several individual tests that are aimed at different aspects of performance. In Table III are shown the results for VirtualBox, and in Table IV - for Xen hypervisor.

TABLE III
UNIXBENCH FOR VIRTUALBOX

System Benchmarks Index Values	Baseline	Result	Index
Dhrystone 2 using register variables	116700	12871170	1102.9
Double-Precision Whetstone	55	2351.4	427.5
Execl Throughput	43	490	114
File Copy 1024 bufsize 2000 maxblocks	3960	37510.5	94.7
File Copy 256 bufsize 500 max blocks	1655	9552.4	57.7
File Copy 4096 bufsize 8000 max blocks	5800	140643.4	242.5
Pipe Throughput	12440	40775.2	32.8
Pipe-based Context Switching	4000	7452.1	18.6
Process Creation	126	1177.4	93.4
System Call Overhead	15000	986973.3	658
System Benchmarks Index Score		137.8	

TABLE IV
UNIXBENCH FOR XEN

System Benchmarks Index Values	Baseline	Result	Index
Dhrystone 2 using register variables	116700	26366570	2259.3
Double-Precision Whetstone	55	3294.7	599
Execl Throughput	43	1328.6	309
File Copy 1024 bufsize 2000 maxblocks	3960	304156.8	768.1
File Copy 256 bufsize 500 maxblocks	1655	82718.2	499.8
File Copy 4096 bufsize 8000 maxblocks	5800	678869.9	1170.5
Pipe Throughput	12440	483750.5	388.9
Pipe-based Context Switching	4000	61750.9	154.4
Process Creation	126	2799.9	222.2
System Call Overhead	15000	502553	335
System Benchmarks Index Score		516	

E. Performance test for Windows Server 2008 on VirtualBox

The determining of the performance of a virtual machine with Windows Server 2008 is implemented by PassMark PerformanceTest 7.0 [10]. Different tests are done: performance of the CPU (Fig.3), a test of memory read / write (Fig.4) and input-output operations of the virtual HDD (Fig.5).

Fig.3 CPU performance test on VirtualBox

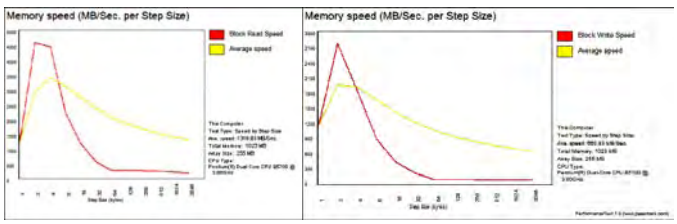


Fig.4 Memory read/write test on VirtualBox

The test includes integer math operations, floating point operations, multimedia instruction, compression, encryption,

sorting strings, SSE, 3DNow instructions. Based on the results of each test an overall assessment of the CPU is formed.

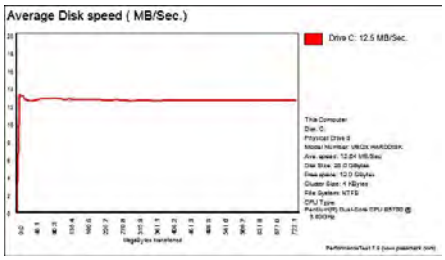


Fig.5 Virtual HDD test on VirtualBox

E. Performance test for Windows Server 2008 on Xen

Similar test are made on Xen hypervisor platform (Fig.6-8).

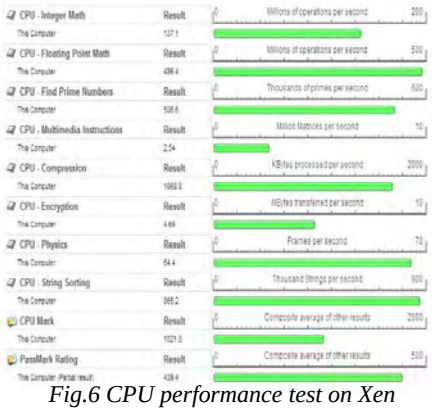


Fig.6 CPU performance test on Xen

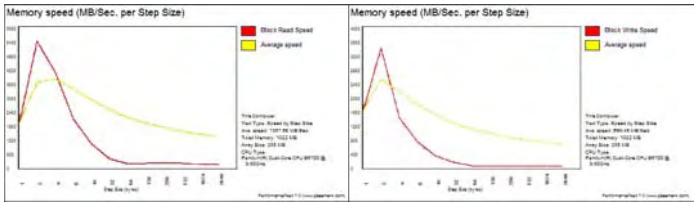


Fig.7 Memory read/write test on Xen

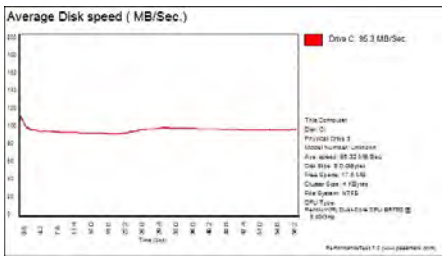


Fig.8 Virtual HDD test on Xen

F. Summary results

In Tables V and VI are shown the summarized results from the tests. The curves have similar shapes indicating that both platforms have similar mechanisms of resource management.

The analysis of the performance of virtual machines on both platforms showed that Xen platform has better indicators than VirtualBox. On this basis it can be concluded that virtualization platform Xen hypervisor is a more suitable choice for the building of virtual networking laboratory on the existing infrastructure in University.

TABLE V
SUMMARY RESULTS FOR LINUX VM

System Benchmarks Index Values	Xen	Virtual Box
Dhrystone 2 using register variables	2259.3	1102.9
Double-Precision Whetstone	599	427.5
Execl Throughput	309	114
File Copy 1024 bufsize 2000 maxblocks	768.1	94.7
File Copy 256 bufsize 500 maxblocks	499.8	57.7
File Copy 4096 bufsize 8000 maxblocks	1170.5	242.5
Pipe Throughput	388.9	32.8
Pipe-based Context Switching	154.4	18.6
Process Creation	222.2	93.4
System Call Overhead	335	658
System Benchmarks Index Score	**516**	**137.8**

TABLE VI
SUMMARY RESULTS FOR WINDOWS SERVER 2008 VM

Tests	Xen	Virtual Box
CPU	1021.8	984.6
Read	1367.68 MB/s	1318.83 MB/s
Write	699.45 MB/s	660.83 MB/s
HDD transfer	95.32 MB/s	12.54 MB/s

V. CONCLUSION

This paper provides an introduction to virtualization technologies and discusses the use of such platforms to build a virtual networking laboratory. Using desktop virtualization enables efficient use of available computers in the labs, thus reducing the cost of infrastructure but saving the performance. The experiments show that the choice of Xen hypervisor satisfies the requirements for virtual infrastructure. Goal of future work is actually building a virtual network laboratory and examine the implementation of virtualization to improve the quality of education at the university. The work will improve the university teaching methodology, will bring new learning techniques and will enrich the experience of both students and lecturers.

REFERENCES

[1] Drews, J. E. “Going Virtual,” Network Computing, Vol. 17, No. 9, p. ES 5, 2006.
[2] Kroeker K. The Evolution of Virtualization. Commun. of the ACM, vol.52, No3, pp.18-20, 2009.
[3] Lunsdorf D. Virtualization Technologies in Information Systems Education. Journal of Information Systems Education, vol.20, pp.339-348, 2009.
[4] Fecilak P., Jacab F. Solutions for virtual laboratory. In Proc of ICNS 2011, pp.314-319, 2011.
[5] VMWare Workstation. http://www.vmware.com/products/workstation/.
[6] Microsoft Virtual PC. http://www.microsoft.com/windows/virtual-pc/.
[7] Oracle VirtualBox. http://www.virtualbox.org/.
[8] Xen hypervisor. http://xen.org/products/xenhyp.html.
[9] UnixBench. http://code.google.com/p/byte-unixbench/.
[10] Passmark test. http://www.passmark.com/products/pt.htm.

Integration of Video Conference into eLearning Platform Based on Moodle for the Vocational School

Student authors: Ilche Acevski[1], Valentina Acevska[2],
Mentor: Linda Fahlberg-Stojanovska

***Abstract* – In vocational education students should acquire skills required in the labor market. In Macedonia the practical work takes place mostly in the schools, and very little in the companies. On the other hand video conferencing can provide distance learning in real-time discussions with educators and experts in the particular field. This paper will cover two parts. The first part dealt with the limitations and potential of the technology used for video conferencing, what server is required, what software is needed and so on. And the other part will involve researching whether students and teachers are ready to use video conferencing.**

***Keywords* – eLearning, server, technology, video conferencing.**

I. INTRODUCTION

How to enrich an eLearning environment such as Moodle with the possibility of video conferencing, faster creation, sharing and monitoring of effective educational experiences for students and teachers. In recent years a very popular question is how students from vocational schools will acquire the necessary skills? What if students do not have many opportunities to perform some of the practical work in real companies. Does video conferencing can provide distance learning in real-time discussions with educators and experts in the particular field, and thus students to acquire communication and IT skills, but also specific skills needed for the labor market.

The basic idea is derived under strange circumstances. If you install Adobe package and then run Microsoft Word, you can see an icon for starting a video conference or on-line meeting through Adobe Connect. What is Adobe Connect? Adobe Connect is the solution for web conferencing, web meetings, learning and can provide extremely rich interactions. Through this system you can participate in online web conferencing in real time, to use resources such as audio (microphone), video (webcam), presentations (in PDF or PPT, which are converted to Flash), and to share them on your computer screen or whiteboard. Even when you cannot attend a meeting or you want to get back later to check something, Adobe Connect allows you to see recording of what happened, like videos on YouTube [1]. But this is a closed source software, and should be paid.

Is it possible to set up something similar based on open source and integrate it into Moodle?

It is possible if you have a server under your control, and if you could include support for multiple simultaneous users. Installation, configuration and setup is not so easy, but with more engagement and energy it can be done.

Such a solution, perhaps more powerful than Adobe Connect is OpenMeetings. Code and manuals are hosted on Google Project Hosting. OpenMeetings same as Adobe Connect, has the ability to set up audio and video conferences, to convert presentations into flash format, to record meetings, to share documents on a whiteboard, to share screen etc. But the most important thing is that runs across all platforms. OpenMeetings is a free browser based software that allows you to set up instantly a conference in the Web. You can use your microphone or webcam, share documents on a white board, share your screen or record meetings. It is available as hosted service or you download and install a package on your server with no limitations in usage or users [4].

In [2] is described spread of video conferencing as a tool for connecting students in an efficient and economical way to "real world" outside the classroom walls. Additions of the book offers a wealth of resources and checklists for making the experience of video conferencing successful.

“Historically, the cost of maintaining dedicated telephone lines for ISDN videoconferencing connections has been a major obstacle to adoption in schools. Line fees can range into the hundreds of dollars per month for a resource that may be used only once or twice a month. Now that IP connectivity is more readily available and codecs are offering higher rates of compression (thus more available bandwidth), more content providers are offering IP-delivered programs. As a result, more schools can consider equipment purchases. Getting started with video conferencing may require a school to invest approximately three to five thousand dollars.” [3].

What is the main purpose of this paper? This paper will cover two parts. The first part dealt with the limitations and potential of the technology used for video conferencing, what server is required, what software is needed and so on. And the other part will involve researching whether students and teachers are ready to use video conferencing.

Student autors:
[1]Ilche Acevski is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola,Macedonia, E-mail: iacevski74@gmail.com.
[2]Valentina Acevska is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola, Macedonia, E-mail: valle.mk@gmail.com.
Mentor:
Prof. Linda Fahlberg-Stojanovska is with the Faculty of technical sciences, Ivo Lola Ribar bb, 7000 Bitola, Macedonia, E-mail: lfahlberg@gmail.com.

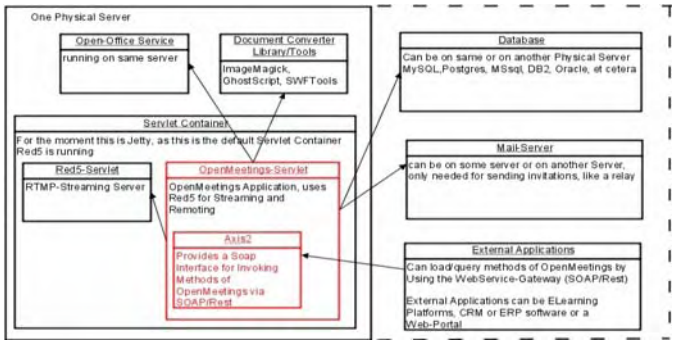


Fig. 1. Technologies used to integrate OpenMeetings

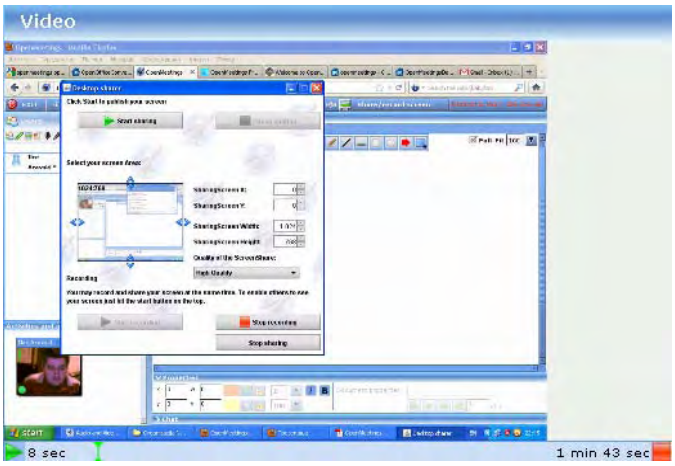


Fig. 2. Showing previously recorded video conferencing room

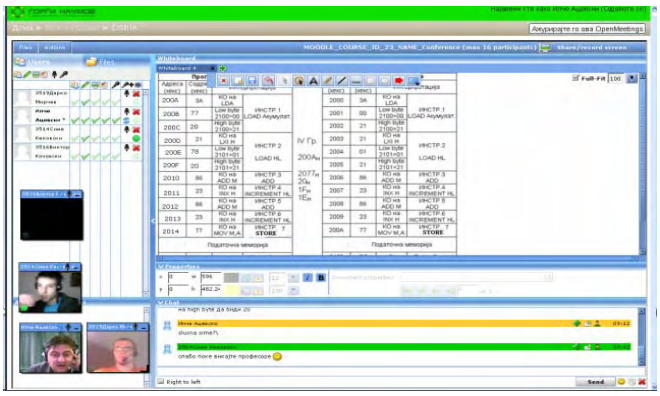


Fig. 3. Importing documents into whiteboard in Conference room

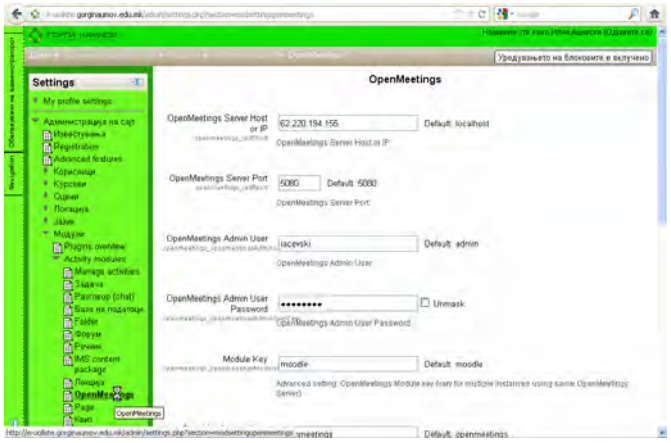


Fig. 4. Configuration of the OpenMeetings plugin

II. METHODOLOGY OF RESEARCH

A. Technologies

Technologies used to integrate OpenMeetings is shown at Fig. 1. Openmeetings uses Red5 server as media streaming server. By default OpenMeetings uses the integrated Apache Derby database. For production environment you should consider using MySQL, Postgres or for example IBM DB2 or IBM DB2. Let see some other open source applications needed for OpenMeetings to work with all features: FFMpeg codec and SoX (Sound eXchange) to record, convert and stream audio and video, shown at Fig. 2, ImageMagick for image uploading and importing to whiteboard, GhostScript and SWFTools for importing of PDFs into whiteboard, OpenOfficeConverter for importing of .doc, .docx, .ppt, .pptx and all Office Documents into whiteboard, Fig. 3. [5].

B. RED5 vs FMS

OpenMeetings is suitable for e-learning platform, as well as commercial Adobe Connect, but the license for Adobe FMS-Flash Media Streaming server is very expensive. Open source Red5 server written in Java is used by OpenMeetings instead of Adobe's Flash Media Server. FMS features can be written only by Adobe FMS team. Red5 features can be adjustable.

FMS uses Real Time Media Flow Protocol -RTMFP and Red5 work with Real Time Messaging Protocol-RTMP, RTMP over HTTP Tunneling-RTMPT, RTMP Secured-RTMPS and RTMP Encrypted-RTMPE.

FMS uses FLV and H.264 for Streaming Audio/Video, and Red5 work with FLV, F4V, MP3, MP4, AAC and M4A.

FMS uses FLV, MP3, AAC and Speex for Live Stream Publishing, and Red5 work with Sorenson, VP6, h.264, Nelly Moser, MP3, Speex, AAC and NSV.

AMF (Action Message Format) Objects are used by RED5 for communication with the flash clients. There are 2 techniques for communication: Flash-Remoting and RTMP. In Flash Remoting technique, this objects are sent over HTTP (new connection for every request cause latency), and in RTMP technique AMF Objects are sent over TCP (one connection for the whole session cause unnecessary open connection). Red5 uses port 1935 for RTMP, 8088 for RTMPT and 5080 for HTTP [6].

From the above mentioned can be concluded that Red5 can be a good substitute for the expensive FMS.

C. Integration into Moodle

One of the main reasons for which Openmeetings are used in this paper is the easy integration into eLearning platform based on Moodle. Once the OpenMeetings is installed on your server can be easily integrated into Moodle. You can get the latest Version of the Plugin for Moodle from [5], and unzip it to Moodle's mod directory. After that, you will find the Plugin's configuration that is shown at Fig. 4.

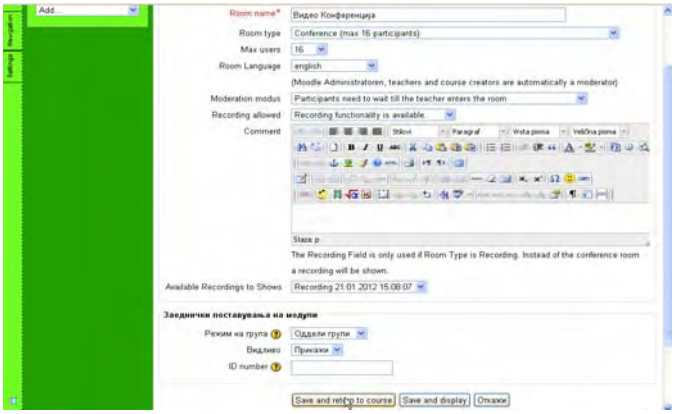


Fig. 5. Creation of OpenMeetings activity

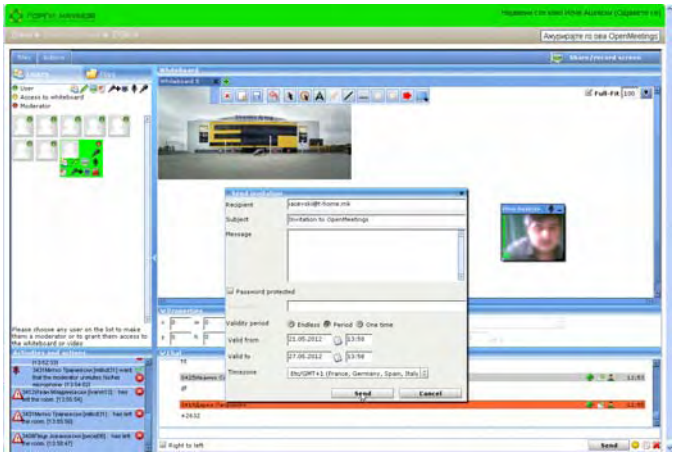


Fig. 6. Student privileges and guest invitation in Audience room

After configuring, OpenMeetings appears as an activity in all courses in Moodle platform, and every teacher can create this activity for his students. With the integration in Moodle all members of the courses are also members of OpenMeetings. During creation of OpenMeetings activity, Fig. 5, teachers are moderators by default. Every student can request some privileges from the teacher, who can assign specific privileges to the student, for each individual user and for each possibility separately, that is shown at Fig. 6. *Sent invitation* from the *Actions* menu is used to send an invitation via email to an outside guest, as an expert in the field to whom students can ask questions. Namely in the video conferencing rooms access have only teachers and students who attend the course. With this invitation, an outside guest will receive an email with link which will lead to the conference, and with a period when the meeting will be scheduled, Fig. 6.

D. Hardware and Operating System Requirements

Openmeetings has been tested twice for this paper.

The first time it has been installed on a computer with Windows XP operating system, AMD Processor 3000 + 1,81 GHz and 1GB of RAM, to feel the software and see if it can be used for nonprofit purposes and restricted conferencing.

The second time, Openmeetings has been installed on a much better Linux-Mint server machine with Intel Core2Duo Processor and 2GB RAM and tested with Moodle.

E. Who will use video conferencing?

All that is mentioned before will be without context, if there are no real use of it. For this purpose a research was done through surveys of teachers and students. These surveys were designed to show whether students and teachers are ready to use video conferencing. Both surveys were in native language, for the students (http://tinyurl.com/753rh8p) and for the teachers (http://tinyurl.com/7durqyk), and was created by Form in Google Docs.

From total of 71 teacher, the survey for the teachers was answered by 50 of them. There was 17 questions in survey, but 3 of them were particularly important for this research. Other questions were for educational quality, and maybe for another researches. This questions were: “8. Would you like to trie, some extra or additional classes to hold through video conference if it is provided by the school?”, “10. Would you like to attend a short training for using ICT, …, using video conferencing, …?”, “17. Would you held a video conference with a group of students and expert from the appropriate field in which the students could ask questions?”

The survey for the students was answered by 213 from total of 792 students (8-10 from each class). There was 17 questions (similar with that of a teachers survey, but prepared for the students), and 3 of them are relevant for this research: “7. Would you like to try some extra or additional classes to hold through video conference if it is provided by the school?”, “8. Do you think that the use of ICT in ways that are mentioned in the above questions would raise your skills and achievements as a student?”, “17. Would you participated in a video conference with an expert from your vocation, if the teacher provide it, to be able to ask him questions?“.

III. Results of Research

The first time on a computer with Windows XP, For several participants at a conference this hardware meets the requirements that refers to the processing and the time delay. But the translation of documents, can kill the process. The machine very often can spontaneously reboot in this case and shut down the meeting.

The second time, on a much better Linux-Mint server machine, using Openmeetings is at first tried when all users are in the same LAN (using OpenMeetings in the classroom), and time delays are awful - up to half minute with 3-4 or more participants in *Conference* (max 16 participants), shown at Fig. 3 and *Aaudience* (max 32 participants) room, Fig. 6. It makes the application unusable in some cases. In Webinar (max 150 participants) room, Fig. 7 this delays are much smaller with more participants, about few seconds.

When the users are at home, all with their own Internet bandwidth, the results are much better and mostly depend on OpenMeetings server. In our case test was done with 7 participants in Conference room, and more than 10 participant in Audience and Webinar room. Results were pretty much better than that from the School LAN, with about 1-2 second delays for audio/video and screen sharing.

Let’s see the results from teachers survey. Original results (in Macedonian) can be seen at http://tinyurl.com/73o6rur.

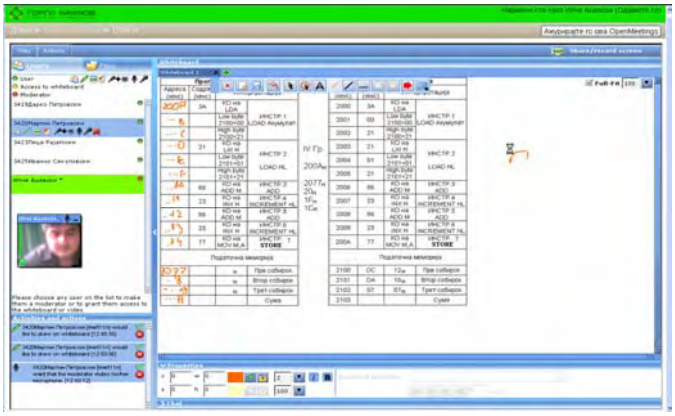


Fig. 7. Webinar (max 150 participants) room

Results shows that about 82% of surveyed teachers would like to held a video conference with expert from the appropriate field. Results are shown on Fig. 8, for every important question from the teachers survey.

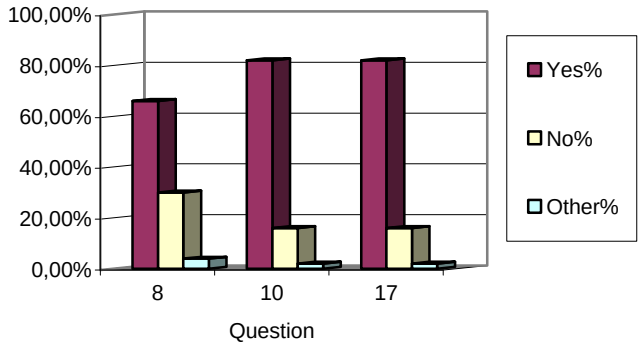


Fig. 8. Results from the teachers survey

Results shows also, that about 76.06% of surveyed students wished to participate in a video conference. This is shown on Fig. 9. Original results from the students can be seen at http://tinyurl.com/7ldnavj.

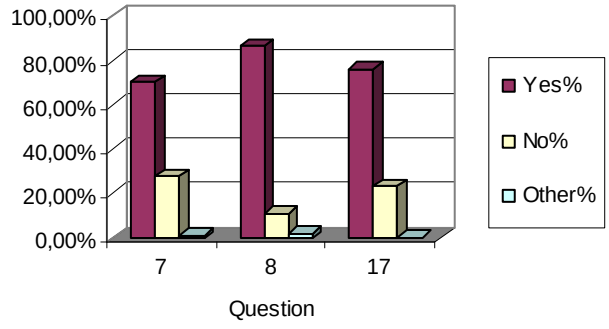


Fig. 9. Results from the students survey

IV. Discussion

From the experience with Windows XP can be conclude that: setup and installation is fairly simple; during documents conversion setup problems arise, and the hardware does not meet requirements.

Of course that under Linux there are quite a lot of tricks when installing, but the following conclusions can be obtained: setup and installation are not so simple; the hardware is good for document converting and for application to work; by adding on more than 10 users, each with its own camera, problems can occur depending of the hardware at the client side; real problems can occur with the flow and there is need of QoS. Hardware and bandwidth can be limiting factors. The router on which the test is done has no QOS. Another note for the end: It is better to use headphones when using OpenMeetings. If speakers are used it may cause audio feedback with a delay, which further reduces quality.

Also, from this research can be concluded that students and teachers are ready to use video conferencing. For this purpose, tutorials were prepared in written and video format.

At http://gorginaumov.edu.mk/?page_id=1703 can be seen video tutorials. Also, tutorials in written format can be seen at http://www.issuu.com/ilceacevski/docs/video_konferencija.

V. Conclusion

During Integration of Video Conference into eLearning Platform, a lot of different techniques were used. OpenMeetings is a powerful tool written in Java, and uses open source applications. With Openmeetings you can use your microphone or webcam, share documents on a multi-whiteboard, share your screen, record meetings, send invitation via email to an outside guest, as an expert in the field that are studied in a course etc. OpenMeetings uses Red5 server as a good substitute for the expensive FMS. There is a need of powerful server machine, wide bandwidth and QoS. Once the OpenMeetings is installed on your server can be easily integrated into Moodle.

At the other hand, through teacher's and student's surveys can be concluded that students and teachers are ready and wish to use video conferencing. But, even though that the tutorials are prepared in written and video format, there is need of quite a job for fairly implementing of this project, and must be other activities which directly or indirectly encourage teachers to change the way they work.

Acknowledgement

Our thanks to many students, teachers and principal at SOTU Gorgi Naumov - Bitola, who helped with this research.

References

[1] Adobe, Using Adobe Acrobat Connect Pro 7, Adobe Systems Incorporated, San Jose, California, USA, 2008.
[2] Cole, C., K. Ray, and J. Zanetis. Videoconferencing for K12 classrooms: A program development guide. ISTE, 2004.
[3] Merrick, S., Videoconferencing K-12: The state of the art. Innovate 2, 2005. [Accessed 26 Fevruary 2012] Available: http://www.innovateonline.info/index.php?view=article&id=24.
[4] OpenMeetings 1.6: Tutorial for (new) Users, January 2010
[5] Openmeetings website, [Accessed 26 Fevruary 2012] Available : http://code.google.com/p/openmeetings.
[6] Red5 website, [Accessed 26 Fevruary 2012] Available: http://code.google.com/p/red5.

System for Multi-variant Multi-parametric WEB-based Test Control

Vladimir Karailiev[1], Raicho Ilarionov[2] and Hristo Karailiev[3]

Abstract – **The necessity of development of a system for WEB-based test control has been examined in the present report. A classification of the multi-variant and multi-parameter models in the developed system has been made. A simple structure and a user interface of the system have been suggested. The realized in the WEB-based system models has got multimedia character and expands the possibilities and the application areas of the test control.**

Keywords – **Test control, Teaching courses, Assessment model, Examination model, Test control technology**

INTRODUCTION – NEEDS OF USING WEB-BASED TEST CONTROL SYSTEM

➢ **Drawbacks of the traditional forms of learning and test control [4]**

- Lack of motivation of the learners and the tutors;
- The traditional form of education requires using buildings, which are not free of charge (heating, lighting, repairs, maintenance, etc.);
- The learning process is carried out considering the possibilities of the medium knowledge learners;
- The terms of the learning are fixed, and it is not possible to decrease them.

➢ **Criteria for the education efficiency**

The efficiency of the education (secondary, college or higher) has been defined by the following three criteria:

- Quality of the learners at the end of the learning cycle. It can be assessed using two main approaches:
  - Making creative characteristics for the practical knowledge and skills and for their theoretical training in the definite science direction. This is made permanently till the end of the study by the lecturers.
  - Making an opinion by the users of the young specialists concerning their quality and eventually recommendations for improving the learning in a distinct direction.
- Duration of the learning. If there are preconditions for individual learning, then a part of the learners will be able to receive the same amount of knowledge and skills in shorter time.
- The spent money for carrying out the whole learning cycle. The less are they the more is the efficiency of the education.

➢ **Necessary conditions for increasing the efficiency of the education [1]**

- Using high information and communication technologies (ICT) as telecommunication technologies, multimedia technologies, Internet technologies, etc.
- Highly qualified, well motivated tutors, who use and apply the new technologies.
- Transforming the consecutive form of secondary education into a parallel-consecutive form. This will allow at completion the secondary education to pass some of the exams from the higher education.

➢ **Reasons, requiring applying the web-based test control**

- There are trends to globalization in the world, information boom in many areas of life. It is necessary the students, specialists and scientists from various geographical sites;
- The students require new learning forms. They want to be more mobile, to complete their education earlier, to be free to distribute and use their time;
- Economical and didactical constraint. The students pay fees which increase continuously and this is a reason for them to work more efficient.

CLASSIFICATION OF THE MODELS IN THE SYSTEM FOR WEB-BASED TEST CONTROL

Depending on where the system for test control is based it can be installed in the following two variants: WEB based and locally based. On the other hand the WEB based system can be installed on a Freeware or Shareware Server. The locally based system is installed on the server of the corresponding department which it serves.

Depending on the ***Number of the variants of the test models*** the system for WEB based test control can be two types [3]:

- One variant test model
- Multi-variant test model

The multi-variant test models in a definite learning course allow making in advance models meeting some criteria. Such criterion for instance is Current or End control, models including questions of higher or lower difficulty level etc. The variety of test models allows the test control system to be more flexible.

Depending on the ***Models of assessment*** the systems for WEB based test control can be the following two types:

➢ **According to the number of the model parameters:**

- One-parameter model of assessment type „0";
- One-parameter model of assessment type „-";
- Multi-parameter model of assessment.

[1]Vladimir Karailiev is with the Faculty of Electrotechnics and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, E-mail: vkarailiev@gmail.com.

[2]Raicho Ilarionov is with the Faculty of Electrotechnics and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, E-mail: ilar@tugab.bg.

[3]Hristo Karailiev is with the Faculty of Electrotechnics and Electronics at Technical University of Gabrovo, 4 H. Dimitar str., Gabrovo 5300, Bulgaria, E-mail: hkarailiev@gmail.com

The ***one-parameter model of assessment type „0"*** may have two sub-models – with one variant of correct answer and with many variants of correct answers, differing with their completeness.

The sub-model with one variant of a correct answer brings points only for one of the variants of the answers. At the sub-model with many variants of the correct answer every of the correct answers brings different number of positive points, and the variants of wrong answers do not bring any points. For the test as a whole only positive or zero sum result is possible.

At the ***one parameter model type „-"*** generally more than one correct and more than one wrong answers are possible. The number of the points for the correct variants of the answers is added to the current result, while the number of the points for the wrong answers is subtracted. A positive, zero or negative total result from the test is possible. A summary for the one-parameter models is given in Table 1.

At the ***multi-parameter assessment models*** with „N" parameters in one test question the following models are possible:

- Model „F" with fixed set of values for each of the parameters;
- Model „V" with variable values for each of the parameters;
- Model „FV" including fixed sets of values and with variable values for the distinct parameters of one test question.

At the multi-parameter assessment models for the each of the parameters it is possible to use only one-parameter model "0" with sub-model „One variant of correct answer" (Table 1).

Then the number of the points, formed for the i test question, including N parameters is expressed by the equation:

$$P_I = \sum_{j=1}^{N} Q_{IJ} \qquad (1)$$

where $j=1$ to N is the number of the questions in one test question.

Q_{IJ} is the number of the points of the i test question for the j parameter.

P_I is the number of the scores of the i test question, considering all N parameters.

$$P_{IJ} = \sum_{I=1}^{M} P_I = \sum_{I=1}^{M} \sum_{j=1}^{N} Q_{IJ} \qquad (2)$$

where $I=1$ to M is the number of the questions in the active test.

P_{IJ} - the number of the points for all M test questions.

- **According to the linearity of the assessment model:**
  - Linear assessment model;
  - Non-linear assessment model.

The assessment model is characterized with a minimum possible mark level

$$C(2)=p.C_{max} \qquad (3)$$

corresponding to the number of points for poor mark, where:

C_{max} – maximum number of points if all the questions have been answered

p – coefficient defining the lowest level of the number of the points, which corresponds to a poor mark.

At the linear assessment model the assessment range is divided into equal intervals for the different marks. At the non-linear model two sub-models are possible – non-linear model with increased resolution at higher marks and non-linear model with increased resolution at lower marks. At the non-linear assessment model the assessment range is divided into non-equal intervals for the different marks, according to the equations in [3].

The possibility to choose various number of parameters and linearity/ non-linearity of the model allow to implement wider area of applications and learning courses, and also simultaneous implementation of a test using many test models for one and the same course. The classification of the models implemented in the suggested test control system is shown in Figure 1.

ARCHITECTURE OF THE SYSTEM FOR WEB-BASED TEST CONTROL

Architectural Model of the Web-based System for Test Control

A simple architectural model of the WEB-based System for Test Control is shown in Fig.2. It is based on the net technology Client/Server [5, 2]. The Web-client represents all the users, who communicate with the WEB server by means of a standard browser using HTTP protocol.

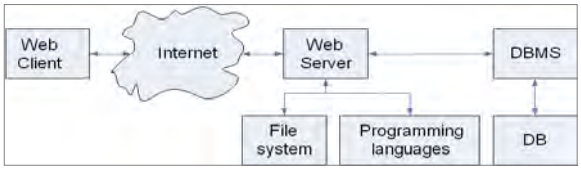


Fig. 2. A Simple Architectural model of Web-based System for Test Control

As a transfer media the global net Internet can be used or a local net Ethernet including some Switches, Hubs, Servers и др. These nets have been built up according to the ISO standard, using 7-level model of OSI.

The basic components of the simple structural model of a Web-based test control system are the following:

- **Web server** – a computer or a set of computers with system and user software connected to the net, which controls the operation of the whole system and the access to the database;
- **Web client** – a computer with respective software, connected to the net and allowing of the user to communicate with the resources of test control system**;**
- **Internet** – the transportation data media of the net using the 7-level standard model of OSI;

TABLE I
ONE-PARAMETER ASSESSMENT MODELS IN THE WEB BASED TEST CONTROL

Type of the answer / Model /Sub-model	True answer Variant scores + 1 to +XX	Wrong answer Variant scores - 1 to –XX
Model „0" Sub-model – One variant of a correct answer	The number of the scores of the answer variant is added to the current result.	The number of the scores of the wrong answer variant is cleared and the current result does not change.
	Only positive, zero or negative total result from the test is possible.	
Model „0" Sub-model – Many variants of a correct answers	The number of the scores of the correct answers is added to the current result.	The number of the scores of the wrong answer variants is cleared and does not added to the current result.
	In case there are either correct or one or more wrong answers, the number of the scores for the current question is cleared. A positive or zero total result from the test is possible.	
Model „-" Many variants of correct and wrong answers are possible	The number of the scores of the correct answer variants is added to the current result.	The number of the scores of the wrong answer variants is subtracted from the current result.
	A positive, zero or negative total result from the test is possible.	

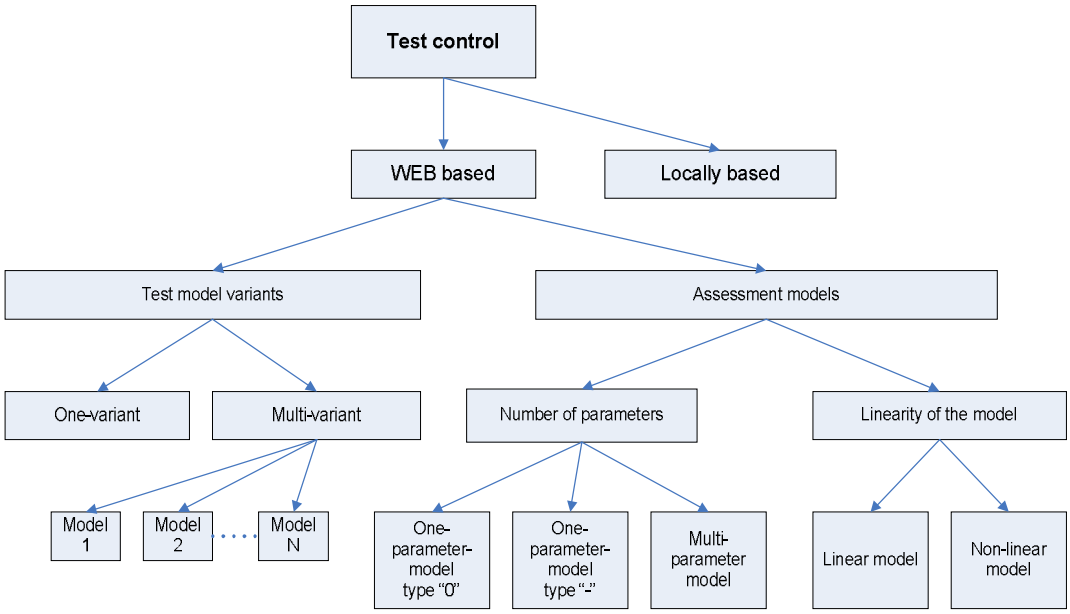


Fig. 1. Classification of the WEB based test control models

➢ **File system** – An hierarchy structure on a memory device, including data catalogue and information itself, used in the system;
➢ **Programming languages** – Web-based programming languages (interpreters and compilers);
➢ **DBMS** (Database Management System) – manages the access levels for the data and implements user inquiries;
➢ **DB** – Database consisting of distinct templates including catalogue data for distribution of the records, lecture courses, tasks, knowledge control tests, files of the tutors and students, schedules for the test process, etc.

The operation principle of the architectural model of the test control system is the following:

The client sends to the server an inquiry via the media, which is processed by the Web server and it is determined whether it is processable by the Web server or not.

The inquiries, which are possible to be processed by the Web server are executed: the required file is taken from the file system; it is processed, and if it is necessary using the corresponding interpreter language and is sent back to the client in HTML format.

Inquiries, which are not possible to be processed by the server, if it is not an inquiry to the DBMS, are taken from the file system and send to the client without change. If the inquiry is for the DBMS and will not be processed by the server, DBMS defines the access privileges of the corresponding client (to which databases and tables the access is possible) and if the client has the access privilege for the corresponding resource, their inquiry is processed and is sent back in a text form answer, if it is necessary.

Basic Functional Features of the Test Control System

➢ **Implementing the knowledge control of the learners by:**

- Developing and inputting tests on the contents of the corresponding courses and implementing the test control.
- Assigning homeworks the students and receiving the results by E-mail, assessing the homework and inputting the mark into the system.

- Conversation with the learner by the means of the Chat to form the mark on a definite learning course.

➢ **Adjusting and development of the necessary models for assessment the results of the learners at implementing the test control.**

- Developing one or multi-parameter assessment models;
- Developing linear or non-linear assessment;

➢ **Creating variants of test models**

- One-variant test model;
- Multi-variant test model;

➢ **Displaying individual data for the learner.**

- Output of an academic reference concerning the current state of the student.
- Output of a creative characteristic concerning the achievements of the student.

➢ **Making group references for the learners.**

- Output of a list of the students, who may receive social scholarship or scholarship for excellent results during the current semester.
- Statistics concerning the results of the students from a course or specialty.
- Statistics concerning the results from all the courses in a concrete specialty.
- Statistics about moving the students from one university to another in the country and from a specialty to another in a university.

USER INTERFACE OF THE SYSTEM FOR WEB BASED TEST CONTROL

Technology for development learning courses and test control

The actions in making tests for a course have been made by the learner, creating the theme from the learning course. The tutor makes the test questions, which have to include all the material in the theme.

A possible technology for development learning courses and test control is shown in Fig.3.

To create **graphical files** – pictures - special software is used (CC1), for instance GIMP, Corel Draw, PhotoShop, etc.

To create an **audio files** special software is used (CC2), for instance Sound Forge, Adobe Audition, Wave Pad, etc.

To create an **audio and video files** special software is used (CC3), for instance Media Studio Pro, Adobe Premiere, Cinelerra, etc.

Y1, Y2, Y3, and Y4 are the corresponding Plugin tools in the toolbar of the HTML editor.

The basic features of the course editor are the following: it is easy to integrate; it has template themes, adjustable HTML output, block elements and attribute options, can be used with plenty of browsers.

File Manager of the system for creating the database for the test control

The questions and answers input are made with the text editor and the file manager of the system. For each answer variant definite number of points is input, depending on whether the corresponding answer variant is complete, correct or not in the chosen assessment model and on the difficulty level.

The tutor file system includes **Directories** for the corresponding courses in which they could manage the files and **Personal site** for announcements and news.

Creating test model, assessment model and time interval for access

The modeling of the test by the lecturer is made during the creation of the test in a table describing from which theme and which level to take a number of questions for a course of lections.

Interface for Implementing the WEB Based Test Control and Verifying the Results

It allows the access to the test for definite students to the corresponding course test for definite time. After completing the test current results and statistics (individual and group) for the students' grades to courses and questions is received.

CONCLUSION

The contributions in the presented work can be summarized as follows:

- A system for multi-variant and multi-parameter WEB based test control, allowing using various assessment models and accumulating statistical information.
- Generating test questions is made using input base of test questions, which allows this to be made in random order for each of the students.
- The number of the answer variants at the various questions may be different and is unlimited. The questions depending on their complexity have weight, and the questions with one and the same weight are assigned to the definite level.
- Availability of one and many variants of models allow more flexibility in using the test control system. Multi-parameter assessment models have been introduced expanding the application areas of the test control system.
- The system for WEB based test control is implemented using Apache Server, the programming language PHP, the database management system MySQL, and can be installed using OS Windows, Linux, etc.

The test control system is verified about its correct functioning for two disciplines for current and final control.

REFERENCES

[1] FLAME – Flexible System for managment of learning and control in MA „G.S.Rakovski", National Seminar, Sofia, May 2005. (in Bulgarian).

[2] J. Meloni, Teach Yourself PHP, MySQL, and Apache in 24 Hours, Published by Sams Publishing, 2003.

[3] H. Karailiev, R. Ilarionov, V. Karailiev. Test Control Module of the WEB-based system SEDL. CompSysTech Conference Proceedings, 2006.

[4] MSCS – Software platform for e-Learning, National Seminar, Sofia, May 2005. (in Bulgarian)

[5] Nabit Akrout, A client server architecture for distance education, Xing Tchnology, USA.

Methods for Assessing Information Sites

Tihomir Stefanov[1]

Abstract –**This paper looks into the methods of website assessment and their effectiveness in the assessment of innovatory online editions. A comparative analysis of the leading Bulgarian informative portals is also made.**

Keywords – **Newspapers, On-line editions, news, news sites, regional media**

I. INTRODUCTION

At present there are numerous methods for the assessment and carrying out of comparative analysis of national, administrative and educational sites. It appears, however, that there is a lack of a unified methodology developed for the assessment of informative news websites, despite the fact that almost every user seeking information online makes use of such resources. This is what has been the motivation for the current research.

The existing methods for evaluation of sites are reviewed in the first part of the report. The second part is dedicated to one of the main carriers of information in the net - on-line news agencies and in the last part of the report there is a comparison of the three of the leading Bulgarian news portals using the pointed methods and some new ones are added.

II. METHODS FOR WEBSITE ASSESSMENT

The most commonly used methods for website assessment can be split into two major subgroups:

- Common;
- Specialised (Private).

The specialised methods are aimed at the assessment of a particular type of websites such as sites on: - Education; - E-trading; - Health sites; - Public administration and E-government. No private method has had its way in the assessment of news sites, operating in connection with the regional print media so far.

The common methods are used to assess random websites. Here are three of the most commonly used methods:

- WEbQEM (Web Quality Evaluation Method) (2003), worked out by a team under the supervision of Luis Olsina from the La Pampa University, Argentina;
- Goor and Murphy (2002);
- Jan Alexander and Marsha Ann Tate (1998).

WEbQEM is based on a full-scale, easily adaptable tree of quality website requirements, which is fully-fledged, has 5 levels of fraction and a total of over 100 elements.

When the **Goor and Murphy** method for website assessment is used, there are fewer and more restricted indicators for objective quality measurement [7]. Goor and Murphy divide the assessment criteria into groups as follows:

Content criteria

The basic forms of control under these criteria are conformity and key word checks – HTML meta tag key words in particular are checked on the basis of their frequency of occurrence within the text throughout the whole site, as well as for correct spelling – its quality is an indicator for the overall website quality.

Technical criteria. *What is rendered an account of is*:

- The time needed for connection to the site and the time required for the data to be received, while the developed Goor and Murphy software calculates the statistically normalised average value of these indicators.
- Link functioning – whether those links are valid or not.
- Analysis of the size of the files comprised in the site.

Organisational criteria – an example of this is the so called internal "diameter" used for interrelated links, internal to the website (“Internal Link Diameter”).

The **Jan Alexander and Marsha Ann Tate** “Web Resource Evaluation Techniques” published in 1998 are based on 5 “traditional assessment criteria” of the various sources of information. According to their originators, like any other source of information, web-resources are subject to assessment on the basis of the above-mentioned five criteria, so long as certain specifications of the information in the web are being indicated [1]. Below are enumerated the criteria and specific questions applied when a website is assessed on the basis of its news content.

A News Web Page is one whose primary purpose is to provide most up-to-date information [2].

AUTHORITY [2]

- Is it clear what company or organization is responsible for the contents of the page?
- Is there a link to a page describing the goals of the company or organization providing the information?
- Is there a way of verifying the legitimacy of the company or organization? That is, is there a phone number or postal address to contact for more information? (Simply an email address is not enough).
- Is there a non-Web equivalent version of the material which would provide a way of verifying its legitimacy?
- If a page contains individual articles, do you know who wrote the articles and his/her qualifications for writing on the topic?

[1]Tihomir Stefanov is with the Faculty of Mathematics and Informatics, St. Cyril and St. Methodius University of Veliko Tarnovo, 3 G.Kozarev str., 5000 Veliko Tarnovo, Bulgaria, E-mail: tiho2000@gmail.com.

- Is there a statement giving the company's name as copyright holder?

ACCURACY [2]

- Are the sources for factual information clearly listed so they can be verified in another source?
- Are there editors and fact checkers monitoring the accuracy of the information being published?
- Is the information free of spelling, grammatical, and typographical errors? These kinds of errors not only indicate a lack of quality control, but can actually produce inaccuracies in information.

OBJECTIVITY [5]

- Is the informational content clearly separated from the advertising and opinion content?
- Are editorials and opinion pieces clearly labeled?

CURRENCY [5]

- Is there a link to an informational page which describes how frequently the material is updated?
- Is there an indication of when the page was last updated?
- Is there a date on the page to indicate when the page was placed on the Web?
- If a newspaper, does it indicate what edition of the paper the page belongs to?
- If a broadcast, does it indicate the date and time the information on the page was originally broadcast?

COVERAGE

- Is there a link to an informational page which describes the coverage of the source?
- If you are evaluating a newspaper page and there is a print equivalent, is there an indication of whether the Web coverage is more or less extensive than the print version?

Additional criteria which enlarge this model and they are tested in practice by the authors, they can be:

AVAILABLE ARCHIVE

- The free archive gives the opportunity of monitoring every site. It is an useful option for the user.

SEARCH OPTIONS

- Facilitates the user in finding information in the site

MOBILE PLATFORM

- The users of mobile phones generate more and more traffic in Internet. The mobile platform is a lighter version of the site. It saves traffic and time for access of the users.

USERS' COMMENTS

- The possibility the user to comment the content of the site or a part of its content gives not only a feedback but it is also a means of communication among its users.

AVAILABLE RSS CHANNELS (Really Simple Syndication)

- By RSS channel the users are informed about new information in the site. It gives an easy connection to the social nets.

The following **practical** steps in the assessment of a particular website [3] can be summed up as follows:

1. *Domain assessment*
- Domain extension (.bg/com/org/net/info/)
- The possession of all extensions and derivatives of the domain
- Popularity of the brand and originality of the name
- Length of domain existence

2. *Popularity within the search engines*
- Hyper links and other sites (Backlinks)
- Index pages (indexed pages)
- Ranking in Google (Google PageRank)

3. *Layout and programming code*
- website design
- Administration/ Content management (the extent of its applicability)
- Programming language (html, flash, php, asp, ajax, java)

4. *Content*
- Volume
- Type (created by users; copyright articles)
- Update
- Staff, needed for its maintenance

5. *Viewing*
- Number of unique users per day (a popular method for the assessment of websites, supported by affiliate networks is the number of unique users per day, multiplied by 10 USD)
- Number of page reloads per user (a main criterion for news sites)
- Cyclic repetitive viewings (monthly, daily, hourly)

6. *Users*
- User profile
- Number of registered users
- Common conditions of use, approved by user during the registration process

7. *Financing*
- Monthly revenue (an average estimate over the past 12 months)
- Type of revenue (coming from advertising, subscriptions, sales or services)
- Number of clients / advertiser
- Monthly expenses

III. On-line News Agencies

In the age of information the most frequently visited sites in the Net are those of the information agencies. The readers find information on the current affairs not only in the country but also worldwide. The strength on the agencies arises from the fact that they not only are being the first to report the news but they also keep up-to-date news coverage. Thus the reader is literally immersed in the news and thanks to the audio-visual materials becomes witness to the current affairs.

In Bulgaria there are a number of websites, but few of them have the capacity to cover and generate news independently. Most websites are far from the concept "informative systems" or the Cohen criteria for defining what we called Informing Science in 1999.

The quality informative websites within the Bulgarian online space are based on news agencies, such as BTA and

BGNES, television companies - bTV and NovaTV or on radio stations, such as Darik and Focus.

Thus, with one and the same technical and human resources news are being covered and information flows generated, yet there are more channels for their transfer, which is effective and at the same time economically expedient.

Fig. 1. Authority – focus-news.net

IV. Comparative Analysis of News Websites

Analysis of the framework and content of websites at http://www.focus-news.net/, http://www.bgnes.com/ and www.dariknews.bg, in accordance with the criteria, outlines in the first part of this paper.

Authority. There should be an organisation, held responsible for the content update. All three organisations have a separate section called "About us" (Fig. 1) at their disposal, where one can get detailed information about the owner of every organization, its contact details and staff members. A significant drawback in bgnes.com is that the name of every author should be placed under each of the publications, whereas the numerous photo materials do have an author included at the bottom.

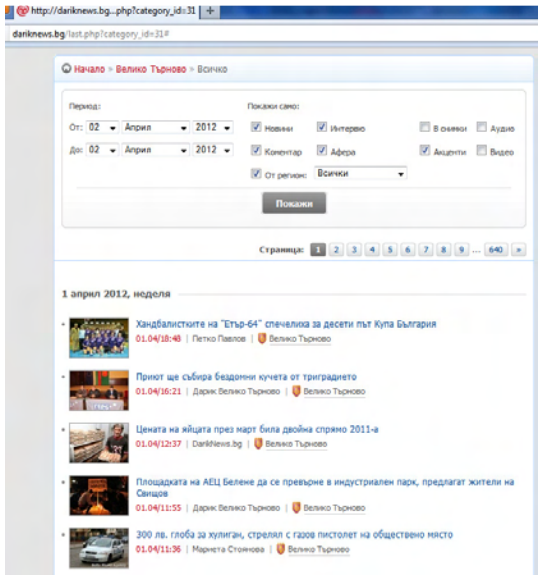


Fig. 2. Search criteria – dariknews.bg

Accuracy. The three websites follow the basic journalist principle of relying on more than one source of information, in all other cases – its full citation should be provided. The texts in dariknews.bg are accompanied by a sound file based on the type of news coverage (a report, an interview, a press conference), while in bgnes.com a video file is being attached. In the work of focus-news.net one can notice a tendency towards covering of current events in progress, by adding or correcting the information in the publications to come.

Objectivity. The information content presented by the agencies can be clearly identified from the advertisements present in their websites. The authors' opinions are always listed in the form of independent comments.

Currency. The three agencies organize their work in such a way as to clearly indicate the date and time of every news publication. This allows the users quickly find the latest update of a particular event.

Fig. 3. Mobile – bgnes.com

Coverage. There is news classification indicated under to their national or regional aspect in all three informative websites. Thus, uses can read the news from a region of their choice. The whole content of the websites is available worldwide.

Viewing. Agencies do not provided public access to their web-counter.

Comparative analysis based on additional criteria:

Available archive. Only dariknews.bg possesses an independent archive, storing old publications, accessible online. The archive is context oriented, directed towards

current news and region being chosen. The archives in bgnes.com and focus-news.net are available against a paid subscription.

Search options. There is no such system within the commonly accessed free section of bgnes.com, and the news search is only possible against a paid subscription. The search options in focus-news.net are not efficient enough and therefore it is not possible to search under specified criteria. The best organized search system is in dariknews.bg (Fig.2), which the searching to be carried out over a specified period of time under various criteria (in news, comments, interviews, regional news, audio and video news only).

Mobile platform. For uses' convenience all three agencies offer the opportunity for access to their content thorough smart phones (Fig. 3). With such access users are automatically redirected to the mobile platform of the respective agency, thus receiving speedy loading and saving of traffic data. Every user has the opportunity to choose the loading of the full (Desktop) version of the website.

Automatic reload. The websites of all three agencies undergo automatically refresh of the information on the user's browser over specified intervals.

Available **RSS** channels (Really Simple Syndication). Only dariknews.bg supports such software mechanism for news exchange.

Users' comments. There exists the opportunity for every news to be discussed and commented by the readers. Only website of dariknews.bg provides this service.

V. CONCLUSION

The news sites are not only one of the main carriers of information on the Internet but, what is more, in the informative century they serve as a channel used by all other means of mass communication (newspapers, radio stations and television).

Such a comparison of news sites has not been offered up to now. A comparative analysis of sites of three news agencies has been done and the existing methods of evaluation have been extended.

The reviewed most common methods of evaluation of sites, as well as the carried out comparison are a base on which the problem to be analyzed and tested in details in future. An automated system for evaluation and analysis of informative sites is being worked out according to the pointed criteria.

REFERENCES

[1] http://lib.nmsu.edu/instruction/evalcrit.html
[2] http://mtateresearch.com/checklist_for_a_news_web_page
[3] http://usm.maine.edu/library/checklist-evaluating-web-resources
[4] http://www.lib.auburn.edu/bi/bookmark.htm
[5] http://www.widener.edu/libraries/wolfgram/evaluate/new.asp
[6] http://www2.vuw.ac.nz/staff/alastair_smith/evaln/evaln.htm
[7] *e-Minicipality,* Sofia, Foundation PIK, 2005, ISBN-954-9456-05-6.

One Approach for Development of Software Modules Adding New Geometric Primitives in 3D Graphics Applications

Emiliyan Petkov[1]

Abstract – **Currently, the main expectations of the abilities of three-dimensional graphics applications increasingly are turning to the use of specific sets of models which perfect the realistic representation of the various objects and processes. There are many developments that provide new or improved models. This study provides an approach to the development of additional software modules that extend the capabilities of three-dimensional graphics applications in the field of modeling and implement new models already in place.**

Keywords – **Computer, Systems, Modeling, 3D, Graphics, Application, Geometric, Primitive.**

I. INTRODUCTION

The creation of photorealistic 3D images requires graphics applications to be developed constantly by adding new and more effective tools [9]. Main part of this process is the means of design that these systems provide. This includes: the coordinate space and viewports, graphics primitives and the models operating behind them, drawing instruments, transactions between graphic objects, modifiers and tools allowing interactive modeling [1].

Efficiency, reliability and functionality of the created or improved models and algorithms can be demonstrated and evaluated in practice if they are implemented in a 3D graphics application (3DGA). This could be done through the development of specialized software modules (SSM) for a specific system [3]. Then an analysis of their work and of the geometric primitives, tools or models that they create, could be performed. The main goal of this research is an approach for development of software modules that add new graphics tools in 3DGA to be offered.

A 3D graphics application can be designed to allow expansion of functional capabilities which further increase its level of usefulness and uniqueness [2, 10, 11]. Such applications have Software Development Kit (SDK) and/or script language. The main part of the toolbox consists of Application Programming Interface (API) which is usually a set of routines written in high level language.

For greater clarity and demonstration of the effectiveness of the basic concept that is given in this approach, a realization of a particular group of models in a specific 3DGA will be considered. The new group of models could be seen in [12, 13]. It represents three quadratic curves: ellipse, parabola and hyperbola, and nine quadratic surfaces: ellipsoid, paraboloid, hyperboloid, double hyperboloid, cone, elliptic cylinder, parabolic cylinder, hyperbolic cylinder and hyperbolic paraboloid. These twelve models of curves and surfaces are represented using NURBS (Non-Uniform Rational B-Spline) functions [13]. Using these new NURBS models of the quadratic curves and surfaces twelve SSM for Autodesk 3ds Max graphics application for design have been developed. They have been approved and received from "*HighEnd3D*" – Internet site for top-of-the-art technologies and 3D applications. They are also available for download as follows: NURBS Conical Arcs 1.4 - http://www.highend3d.com/f/4368.html, NURBS Quadratic Surfaces 1.2 - http://www.highend3d.com/f/4369.html.

[1]Emiliyan Petkov is with the Faculty of Mathematics and Informatics, St. Cyril and St. Methodius University, 2 Theodosii Tarnovski Str., 5000 Veliko Tarnovo, Bulgaria, e-mail: EPetkov@abv.bg.

II. THE SOLUTION

The approach includes four stages. The first one consists a selection of a proper 3DGA and a performing of an analysis of this system in order to determine SDK and geometric models and primitives that it possesses; second stage includes determination of the main features of the new SSM; third one – development of the SSM; and fourth one – comparative analysis. A diagram, showing the main stages in the approach for the development of SSM, is exposed in Figure 1.

A. First stage: Selection and performing of an analysis of a 3DGA which has a toolkit for expansion

Comparative analysis of existing 3DGA and their applications in terms of their main features can be found in [6, 8]. Among them are: 3ds Max, (Autodesk), Maya (Autodesk), Cinema 4D (MAXON), Realsoft 3D (Realsoft Graphics), Blender (Blender Foundation), EIAS (EI Technology Group), form-Z (Autodessys Inc.) , Houdini (Side Effects Software), Softimage | XSI (Avid), SolidThinking (SolidThinking Ltd), TrueSpace (Caligari Corporation).

The first requirement that must be set towards the 3DGA is holding features for expansion. This must be at least one of the three: API, Script language, SDK. The models that are implemented in the 3DGA usually are based on mathematical apparatus. The second requirement that must be respected is this apparatus (or possibility of its creation) to be presented in the 3DGA.

Looking at the presented here example of the twelve new models one could see that the NURBS apparatus stands under.

This means that the representation of NURBS in the API of the chosen 3DGA is the best. For the implementation of the proposed new models of quadratic NURBS curves and surfaces Autodesk 3ds Max graphics system has been selected. The choice has been made taking into account its features. Autodesk 3ds Max is a powerful, comprehensive and well designed 3D graphics application for modeling, animation and rendering [1, 3, 5, 9].

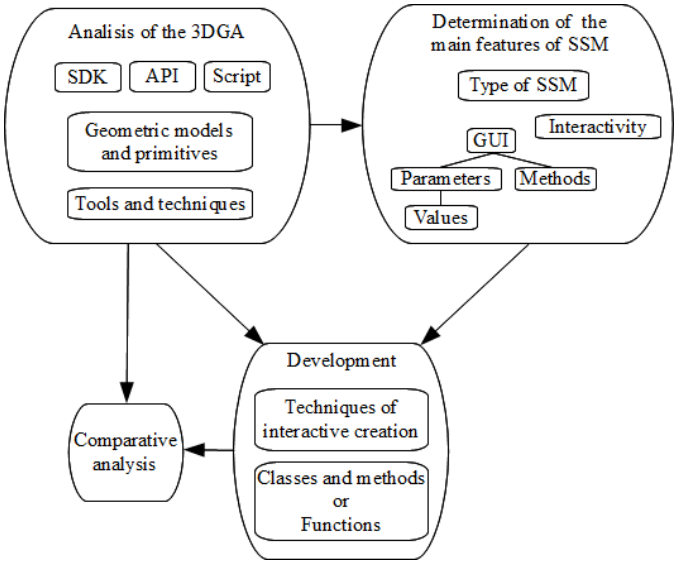


Fig. 1. The main stages in the approach for development of SSM.

An analysis of the types of the geometric (mathematic) models and primitives that are presented in the 3DGA should be performed. Parametric forms and objects should be modeled efficiently as well. There are some types of geometric primitives for design: poly, mesh, patch, subdivided, Spline and NURBS.

For example, all these types of primitives are presented in 3ds Max. Moreover, a fast scan-line algorithm and rendering for high quality production are realized in 3ds Max. It has a full range of effects, several light sources (including photometric lighting), atmospheric effects, calculation of shadows, materials, libraries of textures and creating systems of particles, which provide all the necessary components for creating realistic images and animation.

Autodesk 3ds Max has SDK which is called MaxSDK [2]. MaxSDK is an object-oriented library for creating SSM for the system. MaxSDK provides a fully comprehensive set of classes that developers can combine and expand to create a variety of SSM. In the context of the 3DGA SSM is called plug-in.

Autodesk 3ds Max also features a scripting language called MaxScript. It is specially designed to allow expansion of the system [3]. It may be useful in many ways: automating repetitive tasks, establishing standard procedures, importing and exporting data, creating plug-ins. It may also be responsible for many important features of the program and also be part of the user interface. MaxScript is object-oriented C++-like programming language. MaxScript contains object-oriented API that provides all the functionality and power of scripting applications.

Therefore, 3ds Max has two opportunities for development of SSMs: MaxSDK and MaxScript. Both instruments have the necessary resources for development of the modules according to the requirements that are set out above. If the modules are developed using MaxSDK however, then the SSMs need to be edited and recompiled for each new version of the program in order to run into the system. Furthermore, they should be redistributed and installed by the users. If they are designed using MAXScript for a lower version of 3ds Max (version 4.0 for example), then they will be able to perform correctly and work of each higher version of the application. Based on the fact that versions of the program are released fairly often (every year), MaxScript is more effective way for developing new SSMs.

B. Second stage: Determination of the main features of the new SSM

There are three very important questions that should be taken in mind:

- What tool or primitive does the each SSM represent?
- Which parameters of the model will determine the work of each tool or primitive? This concerns mainly the GUI of each one.
- Will they afford an opportunity for interactive construction in each viewport?

Thus the main features of the new SSMs could be determined.

In the present case NURBS models of quadratic curves and surfaces should be implemented as new graphical modeling primitives in Autodesk 3ds Max. For this purpose, software modules have been developed, which are tools for creating new primitives in the system. The following requirements have been set:

- ✓ They can build curves and surfaces in each of the coordinate plane and orientation depending on the selected viewport.
- ✓ Each of them must represent one drawing primitive (a curve or surface).
- ✓ They must have GUI. This concerns setting the parameters of the curves and the surfaces in their parametric representation of the model.
- ✓ They must provide an opportunity for interactive construction and modeling of curves and surfaces in real time.

Usually, there are different types of modules in 3DGA. Some of them are: *utilities*, *macros*, *plug-ins* and *plugin-scripts*. They can have graphics user interface or not. The main task is the determination of the type of the new SSM. This determines their future structure and behavior.

Some types of SSM are presented in 3ds Max. Concerning the requirement that one SSM should represents one drawing primitive, the type *plugin-script* has been chosen. Plugin-scripts are integrated into the user interface and thus become part of an integrated set of groups of graphical primitives. They are divided into three types: extensions - to expand the existing plug-ins; systems - to create nodes; and

plugin-scripts to create new plug-ins that create new graphic primitives.

Implementation of the proposed models of quadratic NURBS curves and surfaces requires the development of new tools that provide new graphics primitives. Therefore, the third type of plugin-scripts has been chosen, namely plug-ins for creation of new graphics primitives.

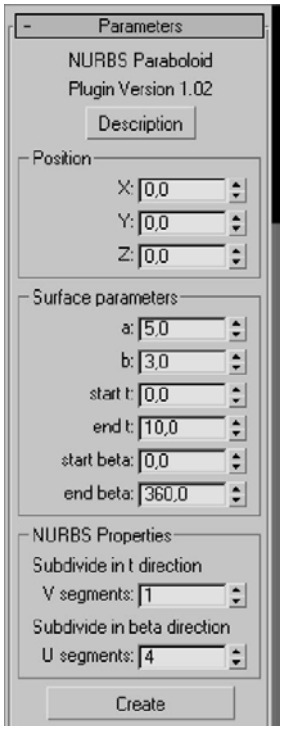


Fig. 2. GUI of a SSM.

Depending on the number and types of parameters, defining the primitive, GUI must be designed. Some 3DGA have separate instruments. In 3ds Max it is called Visual MaxScript. After loading the SSM (automatically or manually) in 3ds Max, two groups of plug-ins NURBS Conical Arcs (containing three primitives) and NURBS Quadratic Surfaces (containing nine primitives) are added to the box “Create”, category “Geometry”. Each SSM has a separated GUI (Figure 2).

The values of the parameters are set in the numeric fields that can receive integer or real numbers. These values should be limited. Usually, there are several events and exceptions that happen during the operation of the SSM. This may be because the user should be aware of something and to correct his/her actions. This is done by message boxes.

C. Third stage: Development of the SSM

Based on the selection of a 3DGA and determination of the main features the development of the new SSM could begin. First, the methods for development of the geometric primitives should be created.

The group NURBS Conical Arcs has a check box in case it is necessary not only NURBS curve to be created, but also a line-shape form. However, the creation of such a form requires more time, so at this point progress control is appeared.

Each plug-in has a button "Create" which the curve or surface is created by. This is an alternative method to the interactive creation in the space. The values of the parameters (length of curve, positions of the focuses, distance between the focuses, eccentricity, etc.), related to specific properties of this object, are displayed in MaxScript Listener window.

Routines that create the new SSMs should be developed. They can be a list of functions added in the main structure of the SSM or a new objected oriented structure that could be used in the plug-in.

A set of routines for the creation of the NURBS models in the new SSM has been developed. They are conditionally divided into two groups. The first group consists the functions that define the main control vertices, the key vectors and all control vertices of the curve or surface. The second one consists the functions which have already the details of the curve or surface and using them create a NURBS object in the scene.

Tools for interactive creation of objects should be developed. Creation of an object could be done by setting values in the GUI but also setting values by mouse movements. This could be done if routines, that are run when events are generated from the mouse, are presented. This depends on the 3DGA. Usually, the interactive creation sets values for a part of the parameters.

The code of the scripts is written by using a command *tool* and two major actions: *mouseClick* and *mouseMove*. These actions determine how the SSM behaves when the user clicks and drags the mouse.

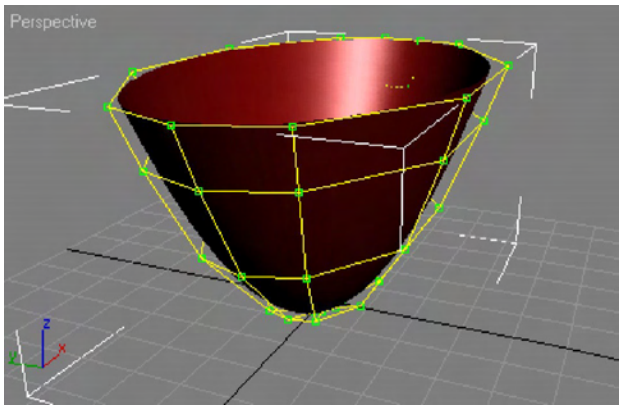


Fig. 3. NURBS Paraboloid created in a selected viewport.

The manipulator of the event *mouseClick* determines what the tool will make in situations when left mouse button is clicked for the first time and when it is released for the last time. The manipulator of the event *mouseMove* determines what the tool will make when the mouse is moved.

In 3DGA the creation of the NURBS objects should be done in all available viewports. Therefore, a routine that concerns the correct creation of the objects, depending on the selected viewport, should be developed (Figure 3).

Often the values of the parameters of the SSM should be localized. This is necessary when the mathematical models require.

The values of the parameters which define the curves and the surfaces are limited by values given by their parametric equations. The permitted values are taken into account in the implementation of the new SSM. These limits can only be subintervals of the intervals given by the quadratic parametric equations of the curves and surfaces.

D. Forth stage: Comparative analysis

Once the SSM have been developed a comparative analysis could be made. The work of the new SSM and the performance of existing tools and techniques in the system can be compared.

For example, a comparative analysis of existing tools and techniques in 3ds Max and the operation of the new SSM have been done. This is reflected both the analysis of their work and the analysis of the results obtained by their work. The main characteristics that are monitored are:

- Number of steps in construction;
- Time of construction;
- Opportunity for local modification;
- Number of control points;
- Level of visualization (for surfaces only).

III. CONCLUSION

In this paper an approach for development of software modules adding new geometric primitives in 3D graphics applications has been presented. This approach includes four stages sown in Figure 1. On the basis of this research and using new NURBS models of quadratic curves and surfaces twelve SSM for Autodesk 3ds Max graphics application for design have been developed. They have been approved and received from "*HighEnd3D*" – Internet site for top-of-the-art technologies and 3D applications.

REFERENCES

[1] 3ds Max 8 Documentation. Autodesk. http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=9861621&linkID=5604642, 2008.

[2] 3ds Max SDK Documentation. Autodesk, http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=7481368. 2008.

[3] Autodesk 3ds Max Detailed Features. Autodesk. http://usa.autodesk.com/adsk/servlet/index?siteID=123112&id=8108755. 2008.

[4] Autodesk Maya API - White Paper. Autodesk Maya Publisher, USA, 2007.

[5] Autodesk Unveils 3D Studio MAX for Windows NT; Technology Innovations Mark Next Generation of Leading Professional 3D Animation Tool. http://findarticles.com/p/articles/mi_m0EIN/is_1995_August_8/ai_17118597. Business Wire, 1995.

[6] Comparison of 3d tools. CG Society of digital artists. http://wiki.cgsociety.org/index.php/Comparison_of_3d_tools. 2007.

[7] ds Max 7 Fundamentals and Beyond Courseware Manual. Elsevier Inc., USA, 2005.

[8] European CG network "TDT 3D". http://www.tdt3d.be/articles_viewer.php?art_id=99. 2008.

[9] Key 3ds Max Features for Design Visualisation Professionals. http://www.autodesk.co.uk/adsk/servlet/item?siteID=452932&id=9898959. 2008.

[10] Maya API White Paper. Alias|Wavefront, Toronto, ON, Canada M5A 1J7, 2007.

[11] Meade, T., S. Arima. Maya 6: The Complete Reference. The McGraw-Hill Companies, USA, 2005.

[12] Petkov Emiliyan. Development and Implementation of NURBS Models of Quadratic Curves and Surfaces, Serdica Journal of Computing, Sofia, 3/2009, pp. 425-448.

[13] Petkov Emiliyan. NURBS Models Development of Quadratic Surfaces for 3D Graphic Systems. Scientific journal "Computer Engineering", Sofia. 2008, pp. 22-31.

Analysis of Platform Dependencies in Software Solution for Auction and Trading in Electric Energy Market

Milos Gajic[1], Marko Djukovic[2], Sasa Devic[3], Branislav Atlagic[4], Zvonko Gorecan[5] and Dragan Tomic[6]

Abstract – **This paper shows possibilities and limitations of mobile technologies (*Android and Windows Mobile*) as well as web solution (*Silverlight*) in the management process of auction trading of power energy in Central - Eastern Europe. Also, there is a brief overview of the process of auction and the information that is of interest to the participants. Moreover, it is shown a short description of these technologies and the architecture of the proposed solutions and their performance.**

Keywords – **Android, Windows Mobile, Silverlight, Auction, Energy Market.**

I. INTRODUCTION

Electric power industry is dealing with problems of generation, transmission, distribution and consumption of electricity. The main aim of its activities is to ensure the required supply of electricity to consumers, the prescribed quality, the necessary levels of safeness, reliability and with the minimum costs. Cessation of large investments in the construction and development of electric power system, it leads to the introduction of deregulation in the vertically-integrated, monopolistic oriented electric power companies and creating the need for an open electricity market in the form of stock exchange. Electricity trade in Europe is done through the auction office established by the association of electric-power companies. The auction office should ensure a smooth exchange of electricity between producers and consumers on a daily, weekly, monthly and yearly basis.

The auction process consists of two steps. The first step is trading of electrical power between producers and consumers, while the second step is validation of transactions taking into account constraints of transmission network. For validation of transactions is developed Power System Analyzer (*PSA*) software support. TSO (*Transmission System Operators*) has enabled communication with PSA through a desktop client called PSAClient. Based on a model of the national network that provides transmission operator, PSA performs the allocation of transmission capacity of interconnecting lines at the Central and Eastern Europe, which includes the following countries: Poland, Czech Republic, Slovakia, Hungary, Austria, Slovenia and a part of Germany.

The auction process is very dynamic and often happens that a certain group of people needs access to current network status, as well as to the results of calculation. Still, due to the nature of their work, they do not have a constant access to the auction process information, which could be a problem. There is a probability that having only a desktop client might be a problem. In today's world of information technology it's almost mandatory to have a browser and mobile access for any network application in order to gain a head start over competition.

In order to overcome these problems it is suggested two mobile solutions based on Windows Mobile (*WM*) and Android platforms and one web solution based on Silverlight technology. Mobile phones allows flexibility, mobility and constant access to auction process, but if user needs detail information it can access them via web application suggested in this paper. All solutions, will be presented here, are designed to work alongside with existing *PSAClient*.

II. ARCHITECTURE OF PSA SOFTWARE SOLUTION

Software applications described in this paper are developed to control auction process for trading of electricity. Each country (the operator) of auction process must have an insight of its electricity network such as information about power, voltage, current, etc. For that purpose is developed a server application called PSA.

PSA performs calculations and analysis of static load flow and tests over random changes in power systems, taking into account needs of liberalized electricity market. In addition to the basic tools to load the input data elements and edit the network model, within PSA application there are tools for automatic testing of random changes, calculations for the network transmission capacity (*NTC*), for the calculations of power transfer distribution factors (*PTDF*), factors for maximum flows, for scaling of production and consumption, as well as to display the results of different calculations.

Each operator sends its transmission network model that consists of:

- all nodes with data of active and reactive power and voltage constraints,
- boundary (X) nodes, fictional nodes that are located in the areas of interconnection lines,
- branches of the transmission network or lines, with parameters and maximum allowable current,
- transformers with a rated primary and secondary voltage.

[1]Milos Gajic is with the TelventDMS D.O.O., Narodnog Fronta 25, 21000 Novi Sad, Serbia, E-mail: milos.gajic@telventdms.com

[2]Marko Djukovic is with the TelventDMS D.O.O., Narodnog Fronta 25, 21000 Novi Sad, Serbia, E-mail: marko.djukovic@telventdms.com

[3]Sasa Devic is with Telvent DMS D.O.O., Narodnog Fronta 25, 21000 Novi Sad, E-mail: sasa.devic@telventdms.com

[4]Branislav Atlagic is with TelventDMS D.O.O., Narodnog Fronta 25, 21000 Novi Sad, E-mail: branislav.atlagic@telventdms.com

[5]Zvonko Gorecan is with TelventDMS D.O.O., Narodnog Fronta 25, 21000 Novi Sad, E-mail: zvonko.gorecan@telventdms.com

[6]Dragan Tomic with TelventDMS D.O.O., Narodnog Fronta 25, 21000 Novi Sad, E-mail: dragan.tomic@telventdms.com

Job of PSA application is to merge individual national network model in a comprehensive joined model, and to calculate the specific technical parameters that are later used as input to the auction software part.

PSA application must be able to perform calculations twenty-four hours a day, seven days a week. It is also developed client oriented desktop application intended for operators of certain countries. *PSA Client* allows a user to send national models, display input data, run specific calculation, get an insight into the results, search by preselected criteria and so on. Architecture of PSA software is presented in the Fig.1.

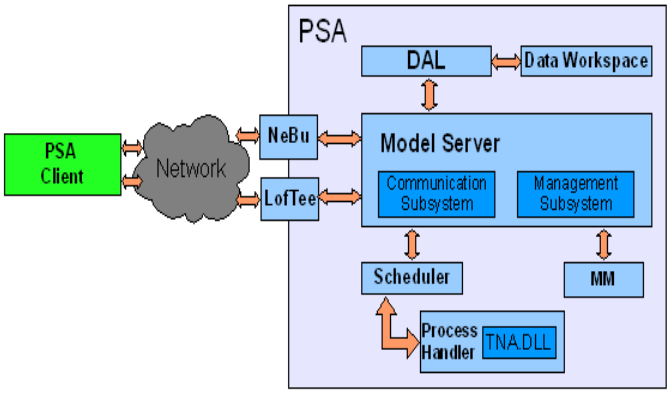


Fig. 1. Architecture of existing PSA solution

III. USED TECHNOLOGIES AND TOOLS

The latest technologies and widely used standards have been used for development of interoperational web services, web and mobile clients which can provide support for the above operations. The following is a brief description of the used technologies and tools.

- SOAP [1] is a protocol intended for exchange of structured data through web services over the network. It is based on XML as the format of messages. SOAP can provide a base layer for web services protocol stack, providing a foundation for building messages based on which web services can continue to build.
- WSDL [1] is language based on XML standard, which provides a model for describing web services. It defines services as collections of endpoints or ports.
- TCP/IP [2] (*Transmission Control Protocol/Internet Protocol*) is the most commonly used protocol for data transmission in computer networks. TCP is responsible for working with data in the transport layer, is suitable for the high-security communication channels of information transmission, but shows poor performance on the communication channels with frequent damage to the data (e.g. wireless).
- HTTP [2] (*Hypertext Transfer Protocol*) protocol is the most commonly used protocol for transmission of information on the web.
- SSL [3] (*Secure Socket Layer*), and it's successor TLS is the standard security technology for establishing a secure connection through unsecured network.
- WCF [4] (*Windows Communication Foundation*), the application interface within .NET environment for the construction of inter-related, service-oriented application. WCF is designed in accordance with the principles of service-oriented architecture (*SOA*), to be able to support the development of distributed client-server architecture. WCF web services have published WSDL interface through which any WCF client can request a service independently of the platform of services.
- Silverlight [5] is an application framework for writing and running Rich Internet Applications (*RIA*). The run-time environment for Silverlight is available as a plug-in for web browsers running under Microsoft Windows and Mac OS X. Silverlight is focused on streaming media, support for multimedia, graphics and animation, and give developers support for CLI languages and development tools. Silverlight is also one of the two application development platforms for Windows Phone. A free software implementation named Moonlight, developed by Novell in cooperation with Microsoft, is available to bring Silverlight (versions 1 and 2) functionality to Linux, FreeBSD and other open source platforms. General concepts mentioned here sound very familiar, because similar approaches were applied in previous technologies, like Java, ActiveX, ShockWare and Adobe Flash. The most successful technology of all is Adobe Flash, to which Silverlight is most a like. But, to make a head start over well proven competition, Silverlight aims to join the power of flash with development abilities that .NET framework has to offer. The advantage of Silverlight over Flash is usage of strong .NET framework and ability of writing code in pure C# programming language. Flash offers better designer support and larger user community, but without comprehensive development framework. To point it out, for any enterprise project requiring heavy programming or data access that would benefit from Windows desktop integration, Silverlight is the way to go.
- Microsoft Visual Studio is an Integrated Development Environment developed from Microsoft. It can be used to develop console and graphical applications, web sites, web applications and web servers in a controlled an uncontrolled code (manage and native) for all platforms supported by Microsoft such as: Windows, Windows Mobile, Windows CE.
- Eclipse IDE is an Integrated Development Environment created by Open Source community. It that can be used for developing Java and Android application. It provides superior Java editing with validation, incremental compilation, cross-referencing, code assistants.

IV. IMPLEMENTATION

In this section we will try to present developed clients for PSA application, how system architecture is expanded in order to retain high system stability and usability, with remarks on base client – *PSAClient*.

A. Silverlight software solutions

The beauty of Silverlight is its easy joining with WCF, the technology of application server side. Developers are enabled to, in a few simple steps, setup a communication channel with WCF web services and almost all capabilities that this technology has to offer. Silverlight communicates with server side in asynchronous way, which means no application freeze on waiting for reply. In our Silverlight web client we will use a communication protocol that allows us bidirectional communication between client and server. For those reasons Silverlight was a simple choice as technology for browser client.

Since the current solution is still under development, and since neither *LofTee* nor *NeBu* web services do not allow changes that could potentially compromise the functionality of PSA Client, it is created a new web service middleware named *LightTube* (Fig 2.).

LightTube web service was introduced to bridge the gap between existing solutions and its expansion in the form of a new, Silverlight client. It serves as an intermediary between *LofTee* web services and Silverlight clients. *LightTube* communicates with the web service through a protocol not supported by Silverlight, and provides services to clients through protocol that is supported. That is why its introduction was necessary. Communication with web service is accomplished by simple adding a reference to the service, which generates a proxy object (the client), through which the transport takes place. All information to generate client are published through the WSDL view. Silverlight client is visible to users of the system. It is used to display the model, run certain processes, information on the progress of running progress, and for viewing results. The application supports the initiation of load flow calculations (*AC Load Flow* and *DC Load Flow*). Load flow calculation is the basic calculation in the auction process. Seventy percent of all other calculations refer to the results obtained from load flow calculations [6].

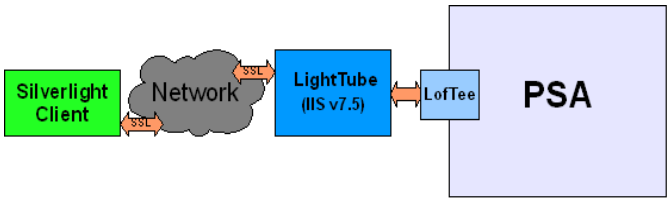


Fig. 2. Silverlight architecture

We have chosen Silverlight for developing client side application because of the following advantages:

- very good design capabilities,
- easy learning,
- good compatibility with existing system,
- no-cross browser issues,
- cross-platform options.

B. Android and Windows Mobile Software solutions

Choosing right technology for application in fast growing and fast changing mobile market can be a tough job to do. In our development process we chose to use two presently leading platforms in this area: Windows Mobile as natural choice considering technology of PSA server side, and Android which is an absolute market leader.

In order not to endanger the functionality of the auction server, and to successfully connect customers with mobile phones, it is developed PSA Adapter, an application presented in Fig 3.

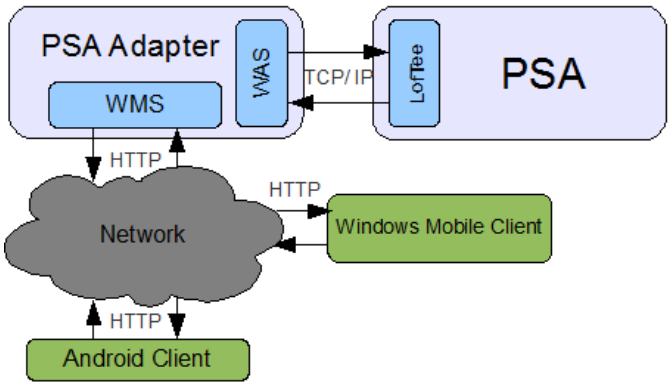


Fig. 3. PSA Adapter

Adapter establishes a connection with the auction server and presents certain information to clients who access it using mobile phones based on Android and WM operating systems. Android client application is written in Java programming language, while WM client application is written in Visual Studio and based on .NET framework environment – same as PSA Adapter. Communication between the auction server and PSA adapter is established via TCP/IP protocol. PSA Adapter is an application written in C# which consists of two web services. First, *WMS* web service, is used for communication between mobile phone and PSA Adapter, while second, WAS web service, is used for communication with the auction server via *LofTee* web service. The need to introduce an intermediary in communication between the auction server and mobile device is implemented for following reasons:

- auction server is implemented to support only connection with clients via TCP/IP, while mobile phones require the HTTP protocol,
- data structures that are used for transmission between the auction server and clients are complex and large mobile devices,
- there was no possibility of modification on the server side because of its functionality.

PSA Adapter takes complex data structures from the auction server, perform their filtering and forwards simplified and desired structures to clients that will be displayed on the mobile phones. This approach enables mobile applications to work faster because all data is prepared in the PSA Adapter so that the network transmits only minimal amount of information. The transfer of information between the PSA Adapter and the auction server is implemented in both directions using WCF duplex services. That is a kind of communication protocol between client and server side, where client can invoke communication with the server (by calling corresponding method), and after that, in similar way, server can invoke methods on the client side. Communication

between *WAS* and *Loftee* is established as user sends request to PSA Adapter, which is then forwarded to auction server. Server processes the request and creates an output structure that is returned to the adapter which receives data and performs repackaging into a format suitable for sending to the mobile phone.

V. Performance Results

The data needed to perform the auction process is given in the form of the following files:

- UCT (*Union of the Coordination of Transmission of Electricity*) file contains information about the status of a regional electricity network (nodes, transformers, branches),
- CBCO (*Critical Branches Critical Outages*) file contains information about the critical sectors of the system as well as critical network outages,
- ANTR (*Already Nominated Transactions*) file contains information about the amount of power that is exchanged between the operator through dealer,
- AATR (*Already Allocated Transactions*) file contains the electric transmission rights that are allocated in a previous stage, and the transfer of rights in case of cross – border exchange.

Table 1 shows the performance of applications described in this paper. Values obtained for the mobile phones have been calculated using simulators. The test was performed on a model that contains 7707 nodes, 6000 branches and 1500 transformers.

TABLE I
PERFORMANCE OF GIVEN APPLICATIONS

Operation		PSA Client	Silverlight	Mobile	
				Android	Windows
Making connection		0.8 [s]	1.5 [s]	2.7 [s]	3.1 [s]
Loading	UCT data	5.5 [s]	49.8 [s]	241.7 [s]	249.8 [s]
	CBCO data	0.6 [s]	2.1 [s]	2.7 [s]	6.7 [s]
	ANTR data	0.5 [s]	1.6 [s]	1.9 [s]	3.3 [s]
	AATR data	0.2 [s]	1.2 [s]	2.1 [s]	3.1 [s]
Starting calculation		0.2 [s]	0.97 [s]	1.3 [s]	1.7 [s]
Getting results of LF calculation		1.7 [s]	5.7 [s]	29.2 [s]	32.9 [s]
Searching	Node	0.3 [s]	1.2 [s]	1.6 [s]	2.2 [s]
	Branch	0.2 [s]	0.9 [s]	1.2 [s]	1.6 [s]
	Transformer	0.1 [s]	0.5 [s]	0.7 [s]	1.1 [s]

VI. Conclusion

We can conclude that each technology has its advances and drawbacks. Android and WM are oriented to fast growing mobile market, but their performance are noticeably lower than they are for Silverlight, which is designed primarily for personal computers.

Here, in mobile applications, we may notice slightly better results for android solution, but that should be taken with caution. We must keep in mind that developing Android application aimed to communicate with Windows services is not even nearly as easy as developing WM application aimed to communicate with Windows services. It is questionable that valuable time lost in development process is worthy slightly better performances. On the long term that is true, but in today world, if solutions are demanded for the next day, it's other way around.

Silverlight can also be used in developing mobile applications, but in environment with much less resources, it is questionable in what degree would that affect its performances. Main power of Silverlight is its capability to interact with machine it runs on, to use hardware to boost its performances. Silverlight is maid to be a strong client side application, which requires a sizable amount for resources just to run itself. Therefore, we should use it for developing web application that tends to be as powerful as desktop applications. To sum it all, main guides when choosing the right technology for any future application we plan to develop should be:

- In what environment will it be running?
- How much time we have to develop it?
- How familiar are we with desired technology?
- How that technology fits in our current system?
- What performances are expected? And of course:
- How big is user community for desired technology?

Today, many technologies in area of web development are rising and falling, and choosing the right one is just getting harder then it has ever been.

References

[1] Prof. dr. Branko Milosavljević, *XML i web servisi.* Fakultet tehničkih nauka, *Novi Sad, 2007/2008,* http://informatika.ftn.uns.ac.rs/XWS

[2] Andrew Troelsen, „Pro C# 2008 and .NET 3.5 Platform, Fourth Edition", Apress, 2008

[3] Prof. dr. Branko Milosavljević, *Bezbednost u sistemima elektronskog poslovanja.* Fakultet tehničkih nauka, *Novi Sad, 2007/2008,* http://informatika.ftn.uns.ac.rs/BSEP

[4] Chris Peiris, Shawn Cicoria, Amit Bahree, *Pro WCF: Practical Microsoft SOA Implementation.* Apress, januar 2007

[5] Matthew MacDonald, *Pro Silverlight 3 in C#.* Apress, novembar 2009., http://apress.com/book/view/1430223812

[6] Prof. dr. Nikola Rajković, Analiza elektroenergetskih sistema II. Akademska Misao, Beograd, 2008.

Rapid development of GUI Editor for Power grid CIM models

Sasa Devic[1], Lajos Martinovic[2], Branislav Atlagic[3], Zvonko Gorecan[4] and Dragan Tomic[5]

Abstract – **This paper presents a solution for designing a GUI editor for power grid CIM models, and a code generator that will ease the work on developing such editor and rapidly decreases time needed for development. The work contains basic description of CIM models and exchange procedure between clients participating in power trading process. Developed code generator relies on UML schema contained in a file created by a designer tool. The approach described here aims power grid CIM models, but it can be used for other CIM models as well. The general instructions for developing code generator are given. At the end, the overall CIM data handling process and resulting solution are presented.**

Keywords – **CIM model, GUI Editor, code generating, interoperability, ICEST 2012.**

I. INTRODUCTION

With constant growth of energy consumption and development of electrical power systems the need for connecting separate (national) transmission networks into one synchronous connection was recognized. In order to increase reliability and security, in 1951 *Union for the Co-ordination of Transmission of Electricity* (UCTE) was formed, which introduced UCTE standard for exchanging electrical network models between different TSOs (*Transmission System Operator*) operators. In order to increase reliability and security, in 1951 *Union for the Co-ordination of Transmission of Electricity* (UCTE) was formed, which introduced UCTE standard for exchanging electrical network models between different EMS (*Energy Managment System*) operators.[1]

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-E (*European Network of Transmission System Operators for Electricity*). ENTSO-E accepted CIM standard as preferred, and in 2009 first interoperability (*IOP*) tests were made, although CIM model is still in developing phase.

IOP tests are held yearly, during the second week of July in Brussels, Belgium. Main purposes of IOP tests are joint testing of developing CIM model, comparison of different software solutions, promotion and wider application of CIM as standard. This paper represents one important part of a product that took part in IOP 2011, where it was well noticed along side with products from companies like Siemens, DigSILENT, Cisco, CESI, GE Energy and 10 other companies. That was its first participation.

Here we will point out that this work is logical continuation of work presented in [4], and main concepts are derived from that project.

[1]Sasa Devic is with TelventDMS D.O.O, Narodnog Fronta 25A-B, 21000 Novi Sad, Serbia, E-mail: sasa.devic@telventdms.com.
[2]Lajos Martinovic is with TelventDMS D.O.O, Narodnog Fronta 25A-B, 21000 Novi Sad, Serbia,
E-mail: lajos.martinovic@telventdms.com.
[3]Branislav Atlagic is with TelventDMS D.O.O, Narodnog Fronta 25A-B, 21000 Novi Sad, Serbia,
E-mail: branislav.atlagic@telventdms.com.
[4]Zvonko Gorecan is with TelventDMS D.O.O, Narodnog Fronta 25A-B, 21000 Novi Sad, Serbia,
E-mail: zvonko.gorecan@telventdms.com.
[5]Dragan Tomic is with TelventDMS D.O.O, Narodnog Fronta 25A-B, 21000 Novi Sad, Serbia,
E-mail: dragan.tomic@telventdms.com.

II. CIM MODEL

A. Basics

CIM is a set of standards for system integration and information exchange based on a common information model. CIM is maintained by IEC *(The International Electrotechnical Commission)*. The part of CIM standards accepted by ENTSO-E is designed for energy market systems. With codename iec61970, development is run by IEC TC57 WG13 (Technical Committee 57, Work Group 13). Profile referred to in this work is for transmission networks – CPSM *(Common Power Systems Model)*. At the time of writing active version of the standard is 15.31. Hereinafter, when referring to CIM-iec61970, profile CPSM, version 15.31, we will say only CIM.

The purpose of CIM standard is to define how members of ENTSO-E, using software from different vendors, will exchange network models as required by the ENTSO-E business activities. The following basic operations are sufficient for TSOs to satisfy ENTSO-E network analysis requirements: ***export*** (export internal network model so it can be unambiguously combined with other TSOs internal models to make up complete model), ***import*** (to import exported models from other TSOs and combine it to make complete model), ***exchange*** (every sent model must carry the data who formed it, which data brings and for which use case is designed for).

B. File structure

ENTSO-E CIM models 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 logical groups and less rapidly changing information from those changing more frequently. Therefore, model information exchange is divided into eight files: *Equipment*, *Equipment Boundary*, *Topology*, *Topology Boundary*, *State Variables*, *Dynamics*, *Diagrams* and

Geographical data. Information from all six files can be combined into one "complete model", which joins all data. Changes on the model are exchanged by difference files that only contain information what are changed, new and deleted elements. When difference files are received, changes it carries are applied to the model.[2]

The CIM model itself is designed with *abstract* and *concrete* classes. Through those classes it maps physics of electrical power system, its states at the specific time, 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. In contrast 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. Data exchange involves only concrete classes. As an example *Voltage Level* class will be presented (see Fig. 1).

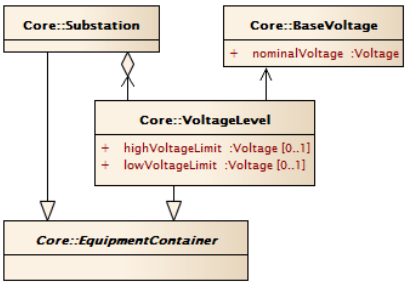


Fig. 1. Voltage Level (object-oriented model)

Voltage Level class inherits *Equipment Container*, as well as Substation class does. But, there are some attributes and associations in *Voltage Level* that do not exist in *Substation*, and vice versa. *Voltage Level* also has an association to *Base Voltage* and an aggregation to *Substation*. However, *Substation* doesn't have those connections.

This small part of CIM model is chosen to show the complexity of the model itself. Other elements of CIM model have more complicated associations and aggregations which would be hard to follow.

III. GUI EDITOR DESIGN

In this section we will try to analyze possible GUI design from designer point of view, with respect to existing solutions for this problem. As we know, a good GUI interface is an intuitive one; it reduces learning experience to plain visual search for commands that resemble verbs of spoken language. Work area should be well organized. We must keep in mind that interface is the one that attracts customers into buying a product, and background solution can only keep them.

Most solutions described in [3], that allow editing of all classes and fields in CIM model, tend to have display interface separated into two vertical sections, the left one – *navigation* panel, and the right one – *details* panel. Navigation panel usually consists of tree view, where root branches are elements that reference to no other element, their sub branches are elements referenced to them, and leafs are elements to which no other element references to. So, following those rules, elements that reference to multiple elements can be shown in tree view by their multiple representations (more than one occurrence). This can cause big confusion with users, especially considering the fact that elements are represented by their IDs, which mostly have no mnemonic meaning. We will also mention that to some elements many elements reference to, and that can recursively continue resulting in a tree with large hierarchical depth, which practically means – a lot of mouse clicking go get to the element that user needs. Search tool is usually added as additional tab in navigation panel and increases information density on one place. Details panel shows attribute details for one selected elements in navigation panels, where attribute name, type name and value have separate columns. Details panel is wider then navigation panel, and therefore it takes central place of application interface. But, unlike navigation panel, details panel offers much less information to the user.

We should use navigation panel to show with what types of elements we can work with, and actual list of elements move to the center of user's attention – to details panel. Attribute values of selected element from the list should be displayed right under the list, while attribute type should be shown only as a tool-tip. Navigation panel could be arranged as a tree that resembles hierarchy of abstract and concrete classes. By doing so, we will break the complexity of CIM model into logical groups. With this rearrangement we are able to set users focus to the center of our application, we have achieved a certain balance within the application commands, without forcing the user to specially learn how to use the new interface. Elements are organized into their logical groups, so looking for what user needs should not be a problem. A plan for this interface is presented in Fig. 2.

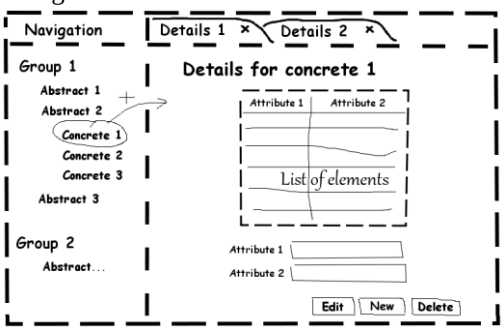


Fig. 2: Interface plan

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 present data in our GUI interface. The piece of the puzzle still missing in our software solution is *how*? We need to develop GUI that implements the idea that we want to realize.

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. We must keep in mind that for every field in every class comes around 100 lines of code, and for every class comes around 2.000 lines of code to correctly manipulate with classes and their

fields, to add, update and delete them all with respect to one another. Knowing that CIM, in version 15, has grown to a number of 337 classes, of which 222 are concrete and require a graphical representation, we get the picture how large this problem really is. If we consider the fact that for IOP 2011, CIM model was modified almost till the start of tests, our problem is even bigger. Since requests for changes on the model increases during last two months before the IOP, improper development plan could result in unwanted results. So, as presented in [4]:

The 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.

Situation now is significantly changed. Even though database is no longer the center of our application, still it is possible to write a code generator that will write a program that we need. CIM standard for power grid has been developing using UML description language, which are shared though a common file type. This is the key, if we can programmatically read the file containing CIM diagrams, we can build a code generator that will write a code that we need!

In this work we will first generate the code, do some additional changes (if needed), and then compile it with the rest of application written by hand. Before going into details of code generation, first let us analyze the existing application where product of our code generator needs to fit in, technology in which application is build and possible technologies for developing code generator.

A. Existing application

Existing application is a desktop client for a WCF web service, written in C# language, in .Net framework, with use of WPF graphical subsystem. CIM models are read and sent from server side to the client side though a web service. We will concentrate our attention only on client.

Since client side is written in WPF, we will try to explain its main principles. WPF is presentation system for building Windows applications with visually stunning user experiences. Good advantage of WPF is the ability to develop an application using both *markup* and *code-behind* specifications. Markup, here called *Extensible Application Markup Language (XAML)*, is used to implement the appearance of an application; while using managed programming languages (code-behind) is to implement behavior, in our case C#. In general this is done to separate design from behavior, to allow designers to take a larger part in development process [5]. Therefore, we can conclude that code generator has to support generating both code-behind (C#) files and markup (XAML) files. Let us examine possible technologies for development of such code generator.

B. Possible technologies

For generating code-behind choosing possible technology was very straight forward. We can write required library by ourselves, like in [4], or we can use CodeDOM [6], a build in .Net API for programmatically constructing and compiling C# source code . General approach is to first construct source code, save it to a file, use built-in compiler to compile it and run it. Here, source code will be construct and saved into a file, and later compiled with the rest of developing application. This is recommended and fully satisfactory technology for needs in this project.

For generating markup, XAML code, things are not so clear, even though XAML is actually one more extension of XML. All built-in elements that appear in XAML can be specified though XML. Creating classes that resemble WPF classes and serializing them to XAML file could really ease the work for us. But, if we want to design reusable components that can be used just like system controls, we should create *user controls*. User controls can contain other (user) controls, resources, and animations, just like a WPF application. For each element from CIM model we need to create a few user controls that have references to other user controls, like specified in CIM model.

User control has root XML element with the same name – *UserControl*, where XML attribute *Class* specifies code-behind class. If user control needs to reference to other user controls, it needs to have an XML element that represents that user control, with the name from its *Class* attribute. This is the real problem with this approach.

In general we would write new user control by simple extending base user control class, and adding references to other already written user controls. But while code generator is running, we can't declare, initialize and reference to the new class from some other also new class, whose names we just read from UML diagram. We could try with constructing a class for serialization with needed references to some default user-control-reference class, and override its name before serialization to the name of a class it actually represents. But this also won't fit our needs, because we always have more than one reference to other classes. Overriding one element type will override its name wherever it appears. We will end up with always referencing to the last element that was overrided.

One more possible solution is to try inventing a new approach, to write our own, new library that will help us in generating wanted XAML files...

C. XAML Generator

By analyzing XML structure, which is also XAML structure, and also WPF structure in whole, we can notice that it reflects composite design pattern. If we *copy* WPF design pattern by using plain classes that have identical hierarchy and attributes, with simple overriding *ToString* method and saving the output to a file, we can get the result we need. Each class will print XAML code that corresponds to WPF class it represents. Printing all attribute values that differ from their default values, will also be included. Of course, like in WPF, all attributes are inherited from base to derived classes. Plain class that represent user control will have additional method called *GenerateReferenceCall*, for simple solution of problem we had with setting references in standard XML approach, explained in previous subsection.

The beauty of this solution is its simplicity and ease of use. Library that implements ideas here presented will allow us to

generate code for our application like if we wrote it by hand. That library we will call *XAML Generator*.

Since code-behind and markup are basically inseparable in WPF, to join the two libraries, CodeDOM and XAML Generator, we will generate both simultaneously. Functions that handle user events are defined in code-behind and called from markup. In that way view and controls are linked. In listing 1 we can see an example class from XAML Generator.

```
public class XButton : XButtonBase{
  public XButton(){}
  public override string ToString(){
    // print XAML element
    String atts = GetAttributeValues();
    String startElement = "<Button"+atts+">\n";
    String body = generateContent();
    String endElement = "</Button>\n";
    return startElement + body + endElement;  }
}
```

Listing 1: Example class from XAML Generator

The possible drawback is the fact that WPF is rather large subsystem of .NET, and writing generator to allow us full support for it would be very demanding. But this is not a real problem, at first, we will have support to generate only a small part of WPF, and gradually, as we progress in building our application, we will expand XAML Generator to fit our needs.

D. Implementation

After defining what we want user to see, navigation and details view, how to develop our code generator, CodeDOM and XAML Generator, we have left *easy* part to do – to generate code architecture to suite our needs.

By using UMLReader component, we will read the meta-data from the UML file, and use it in GUI code generator. This component was already developed as part of another project, and it won't be analyzed here. We will only mention that, like in [4], meat-data is stored into class model from which is easier to read classes names, attributes and relations among them. When meta-data is read into the memory, that will be meta-meta-data.

Navigator component we have already discussed in section III. Here we will only mention that this component is realized by knowing all the classes in CIM model, their names and their base classes.

Details View components will be separated into tree logical sections: tables, selectors and views components.

Tables are used to display a list of one type of CIM elements, where their attributes are shown in columns. Those are the simplest components. Search by criteria is also available for every attribute of string type.

Selectors relay on tables. They are used to select one element from a list of elements of certain type. This will be very useful for setting associations and aggregations among different CIM elements.

Views relay on both tables and selectors. They allow creating new element for the model, modifying and deleting existing elements selected in tables list. For setting associations and aggregations, selector components are used. View components are the most complex of all, because they take care of new, modified and deleted elements, storing them into three separated list for sending to the server side.

This code generator is designed to cover some of specific needs of ENTSO-E CIM model, but it can be applied for other CIM and UML models as well. Appearance of produced CIM GUI editor is given in Fig. 3. Sum of all generated lines of code, for 220 concrete tables, is 355,997 lines.

Development process is presented in Fig. 4.

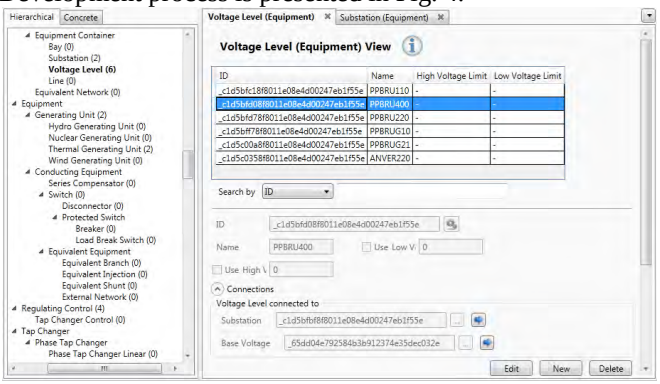


Fig. 3: Produced CIM GUI Editor

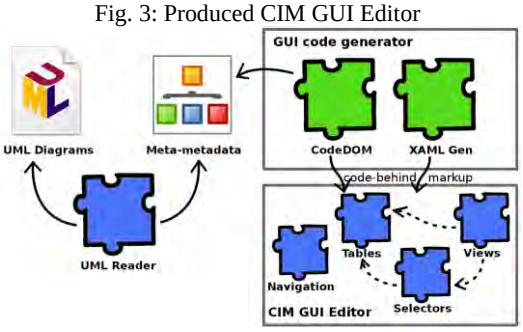


Fig. 4: Development process

V. CONCLUSION

In this work we have explained what CIM models are, and how they are exchanged. We designed more user friendly graphical interface. Code generator is described, for both code-behind and markup specifications. Solving problems with code generators allows us to develop applications much faster, with fewer errors and with less people on the project. The most important advantage of code generating is that it allows us to easily adopt our application on future (frequent) change requests, which is becoming a major obstacle for any large application.

REFERENCES

[1] D. Kirschen, G. Štrbac "Fundamentals of Power System Economics", 2005.
[2] ENTSO-E, "UCTE CIM Model Exchange", component interface exchange, revision 1.0, version 14, 2009.
[3] ENTSO-E, "Interoperability test: CIM for System Development and Operations – Final Report", 2010.
[4] S. Devic, B. Atlagic, Z. Gorecan, "Database modelling and development of code generator for handling power grid CIM models", ICEST, 2011.
[5] Ted Hu, "WPF Series Getting Started", Microsoft MSDN, 2010.
[6] Luke Hoban, "C# 3.0 and CodeDOM", Microsoft MSDN, 2007.

Building an 8085 Microprocessor Module for the HADES Simulation Framework

Goce Dokoski[1], Dimitar Bojchev[2] and Aristotel Tentov[3]

Abstract – **The Hamburg Design System (HADES) is a popular framework for running interactive digital logic simulations. It has a modular architecture written in the Java programming language that makes it ideal for modelling complex digital logic circuits and therefore processor architectures. Given that there are very few HADES processor modules publicly available today, we have started developing our owns.**

This paper focuses on the process of building a HADES processor module, by using the classic Intel's 8085 processor as an example. Some emphasizes on the use of Java threads and its synchronization mechanisms is made, in order to show how the concurrent circuitry of the 8085 internal architecture is modelled.

Finally, it presents the prospects for flexible simulations and analyses of processor systems in HADES.

Keywords–**microprocessor, 8085, module, HADES, simulation.**

I. INTRODUCTION

HADES is a popular framework for running interactive simulations of digital logic circuits[1]. It has a graphical editor that provides a convenient method for modelling digital logic systems. For example, it allows a simulation to be run at the same time its design is implemented as well as show all signal waveforms (wires, pins etc.) on timing diagrams. This makes it suitable for flexible and precise design analysis.

A HADES design usually consists of several digital components interconnected by wires. Its basic building blocks are the boolean logic gates (AND, OR, NOT, etc.), but a large number of more complex components is also available. This includes various latches, flip-flops, registers, timers, multiplexers, decoders, FSA as well as many other models of ROM and RAM memories, bus controllers, peripherals etc. Few processor cores are also available: Intel i4004, Microjava 2 and PIC16C84.

One design can be easily saved as a module and reused in another design as a "sub-design". This is especially important for the hierarchical nature of digital designs. However, there is another alternative when it comes to modelling very complex digital systems, such as complete processor architectures. HADES provides a JAVA programming API that can be used to define custom made logic blocks. Having in mind the vast number of Java libraries available today, one gets practically unlimited number of possibilities for simulation.

When it comes to building a HADES model, the programmer only needs to define and present the input/output pins to the rest of the design. Afterwards he can read or write their values, and have the rest of the system modelled in any way. For example, in the case of a processor architecture, it can be defined either instruction- or cycle- accurately.

The module is a separate Java class that can use any additional software, including Java wrappers for other programming languages and libraries. It can also communicate with hardware components, so it becomes easy to interface the module to devices outside the HADES design in a fully co-simulation environment.

This paper will present the modelling of the classic Intel's 8085 processor as a HADES model. There are several 8085 simulators currently available, most of which are free and open-source [4,5,6]. However, all of these simulators are stand-alone applications that lack the possibility of interconnecting with other components and devices. This is a major disadvantage as it limits them to assembler testing only.

The HADES model presented in this paper is built on top of one of the existing 8085 simulators – the J8085 simulator written in Java. As the source code was available for non-commercial purposes, we used it as a base and added the necessary interface so that it can function as a HADES component. This way a complete working model of the processor is accomplished.

The rest of this paper is organized as follows: the second section gives an overview of the HADES simulation framework and API. In section 3 a short introduction to the 8085 instruction set architecture is given. Section 4 deals with the interfacing of the 8085 HADES module with the J8085 simulator. It emphasises the Java threading mechanisms used to model the concurrent behaviour of the processor. Finally the fifth section gives a conclusion and summarizes the pros and cons of the model.

II. THE HADES SIMULATION FRAMEWORK

A. Overview

As shown in figure 1, a simulation consists of a hierarchy of design objects [1]. Only one of them is a top-level design object, and it manages the other design objects.

[1]Goce Dokoski, Faculty of Electrical Engineering and Information Technologies at the University "Ss. Cyril and Methodious" of Skopje, bul. Krste Misirkov bb, Skopje 1000, R.Macedonia, E-mail: gocedoko@feit.ukim.edu.mk.

[2]Dimitar Bojchev, Faculty of Electrical Engineering and Information Technologies at the University "Ss. Cyril and Methodious" of Skopje, bul. Krste Misirkov bb, Skopje 1000, R.Macedonia, E-mail: dime@feit.ukim.edu.mk.

[3]Aristotel Tentov, Faculty of Electrical Engineering and Information Technologies at the University "Ss. Cyril and Methodious" of Skopje, bul. Krste Misirkov bb, Skopje 1000, R.Macedonia, E-mail: toto@feit.ukim.edu.mk.

The communication among the components is made by a simulation kernel that transfers SimEvent objects among the design objects.

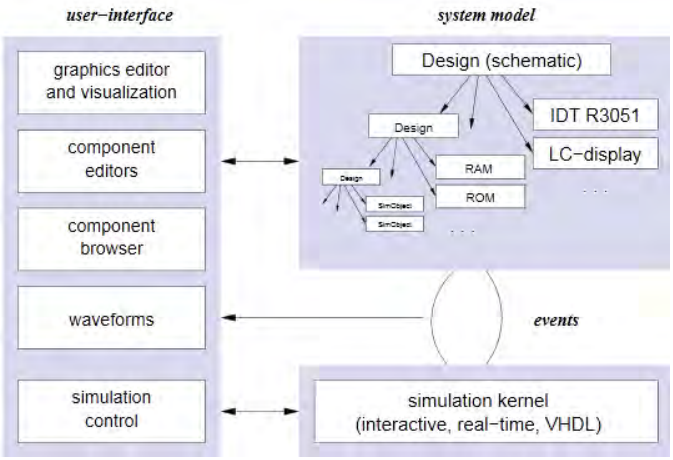


Fig. 1. Overview of the HADES simulation framework as described in the HADES tutorial [1]

The SimEvents carry information on how the wire values chage over time – one SimEvent object carries the new value of a wire that should take place at a given time instant. The wire values can be of the IEEE1164 standard types [2].

B. Component Symbol Description

A component in HADES is a subclass of the SimObject class. In order to design a custom component model, the SimObject class needs to be sub-classed to the custom component class and afterwards only two methods must be overriden:

- **elaborate()** - does the initialization of the component. It is called at the loading and preparation of the HADES design for simulation;
- **evaluate()** - contains the full behavioral model of the component. It is called every time there is a pending SimEvent object, for ex. one or more of its input pins have changed their value.

From here it is easy to recognize the discrete nature of the behavioral model and the simulation: everything that needs to be done in order to model a discrete dynamic system is to program the state changes of the component as a result of its input pins' changes.

C. Component Symbol Description

In order to use a component in the graphical user interface of HADES, a graphical representation needs to be defined as well. This is done in a textual (.sym) file where the number of pins, their names and their position in the graphical representation is specified.

There is also a possibility to define a dynamic graphical representation, for example to maintain visual representation of the internal state and registers in the graphical symbol. However, this is not currently covered.

III. The 8085 Instruction Set Architecture

The 8085 microprocessor is one of the predecessors of the famous Intel's x86 architecture. It is a CISC architecture, that can execute a set of instructions grouped in 5 functional subsets, as described in its reference manual[3]:

- **Data Transfer** – each instruction moves a byte or a (2 byte) word between a register, register pair, memory or I/O location;
- **Arithmetic Operations** – contains instructions that perform addition, subtraction, incrementation and decrementation on data in register, register pair or memory location;
- **Logic Operations** – contains instructions that perform the boolean operations on data in register, register pair or memory location;
- **Branch Group** – performing conditional or unconditional branching in the execution flow;
- **Stack, I/O and Machine Control Group** – instructions that maintain the stack, communicate with I/O devices, and work with interrupt masks and flags.

All instructions execute for at least one and up to five machine cycles, where each machine cycle lasts from three to six clock cycles.

A machine cycle resembles one READ or WRITE operation on a memory or I/O location. It can also serve as an INTR interrupt acknowledge cycle, or leave the bus idle, in case it is in a HOLD state, or if it executes an instruction that doesn't need communication with peripheral devices or memory.

Every instruction may be 1, 2 or 3 bytes long, so in order to fetch it, its execution will last at least 1, 2 or 3 memory READ machine cycles, correspondingly. Afterwards, depending on its functionality, it will activate additional READ/WRITE machine cycles. This happens in case it needs communication with a memory or I/O device, (for example an instruction of the Data Transfer group).

.

IV. The 8085 HADES Module

A. Definition and Initialization

First, the 8085 component symbol definition is specified. All component pins are entered as they are specified in the reference manual[3]. Figure 2 shows the graphical representation as the component will appear in the HADES simulation.

The elaborate() method is the most appropriate place to instantiate the J8085 simulator's main class, and to call its main function. This sets its internal state to the initial values, as when the simulator is started or the simulation reset. A reference to the simulator object is kept as a member variable in the SimObject component class, for convinient access to its functionalities.

Some of the component's pins are also set to their intial state: The AD_{0-7}, A_{8-15} and all input pins are set to high-impedance state so that they can be driven by other devices.

An enumerated variable is added to the class, to keep the state of the processor, and it's initialized to the first state of a READ operation, so that the model is ready to fetch the first instruction.

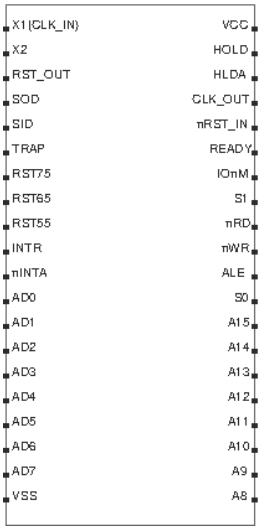


Fig. 2. The graphical representation of the 8085 HADES module

B. Implementation of the Evaluate() Method

The rest of the functionality is implemented in the evaluate() method. As mentioned previously, this method is called whenever **any** input pin changes value.

Currently the progress of the instruction execution is conducted by the CLK_IN signal. It is sufficient to provide sequence of pulses to the CLK_IN pin, because the program execution inside the evaluate() function, only checks for a falling edge on the CLK_IN pin. The X1 and X2 pins' original purpose is to provide a clock sequence to the processor by using signals from a crystal oscillator, but this is not currently supported. At the beginning of the function it is safe to set the CLK_OUT pin to follow the CLK_IN.

The nRST_IN signal is also checked on every evaluate() call. If a low (0) level signal is detected, all of the processor's output pins are reset to their initial values, and the simulator object's execution is reset. The simulator object already included public methods for pausing, resuming and resetting the processor execution, so the reset method is called as well.

Another important step in the processor reset is to cancel the current READ/WRITE operation if any is currently executed. This is explained later in the text.

The next step is to check for the falling edge of the CLK_IN input and progress the instruction execution such as the current READ/WRITE operation.

If the machine cycle is the last one of the current instruction, the interrupt pins are also checked. The J8085 simulator object provides public procedures that trigger handling of the hardware interrupts, so these functions are called in this case. This procedures were originally used with the GUI buttons of the J8085 simulator, but now they are reused to handle the events of the HADES interrupt pins.

C. Interfacing with the J8085 Simulator Object

The J8085 Simulator is a stand-alone Java application with its own GUI that displays the processor's registers, flags and memory state. The whole project consists of several classes, the most important of which are shown in Table 1, together with their functions.

TABLE 1
THE MOST IMPORTANT J8085 SIMULATOR CLASSES

MainFrame	Contains the GUI and all the interfacing with the Executer class
Executer	The behavioural model of the processor. It fetches op-codes, manages the execution flow and READ/WRITE operations
Memory	This class simulates the 8085's 64 KB memory space as an array of 64K Strings that keep the byte values in hex format. It contains methods for memory read/write operations that use this array.
Ports	Class that provides methods for READ/WRITE operations over the 255 I/O locations (ports). The port values are alse kept as a String array.

The most important part of the 8085 model execution was to replace the methods of the Memory and Ports classes so that they use the HADES memory and I/O space instead of the internally simulated ones.

For this purpose, we added special methods to the HADES 8085 model, that perform the READ/WRITE cycles on the multiplexed address/data bus. So instead of using the strings array, the read/write methods of the Memory class, are changed to call the appropriate methods of the HADES 8085 model. This is also the case with the Ports class.

The methods that perform the memory and I/O READ/WRITE operations, have access to all the necessary pins (ALE, nRD, nWR, AD_{0-7}, A_{8-15}, etc). They receive the memory or I/O address as arguments, and receive or return the data that needs to be accordingly read or written.

The only problem that arises is that the functions need to wait for several clock signals in order to drive the pins properly. These methods will be called from the J8085 simulator, and executed in a separate thread. They should execute one state, and wait for a falling edge on the CLK_IN before proceeding with the next one.

In order to efficiently wait for the next falling edge of the CLK_IN signal, without wasting processor time in a loop, the function executes the Java sleep() procedure, which puts it to idle state, until someone calls notify().

In Java a process may call a sleep() function but it needs to specify an object on which it will sleep. Afterwards, any other method that has access to that object, may call notify() on it, which in turn will resume any methods that have called sleep(). The object itself may be ay subclass of the Java Object type.

Whenever the evaluate() method detects the falling edge it will notify() the object on which the READ/WRITE function sleeps, so it will be able continue executing the next machine cycle state. The low level of RST_IN may also trigger a notify(), because when resetting the simulator, the function must not stay in the sleep() state.

Some precaution is needed, because it may happen that several threads notify() a same object, which may result in unpredictable situations. For example, the RST_IN signal may notify the sleeping object simultaneously with the CLK_IN edge. This is easy to resolve - all notify() calls to the object need to be locked on that object, to prevent them from executing concurrently.

D. Compilation and Transfer of the Compiled Code into External Memory

The J8085 simulator has an internal parser and compiler that write the compiled code to its internal memory. In order to transfer the compiled code to external HADES memory component, we added additional option in the GUI menu that exports the compiled machine code to a .rom file.

This file is in a format that allows it to be imported to the standard HADES memory components (ROM and RAM).

As the simulator currently reads the first instruction from the first address of its memory space (0), the memory component that contains the compiled code, should have its chip-select activated on this address.

V. CONCLUSION

In this paper we showed how to build a HADES simulation object that models the behavior of the Intel 8085 processor architecture. We used an existing open source 8085 simulator written in Java, and adjusted its interface to the HADES simulation model.

Instead of letting the GUI maintain the interrupts, memory and I/O ports, the model uses the HADES component pins to communicate with the rest of the HADES design, thus providing a powerful and realistic simulation environment for testing and analyzes of microprocessor systems.

The HADES framework and its Simulation API provide almost unlimited possibilities when it comes to component design, and having in mind their open source availability, their use should be increased in both academic as well as professional system development areas.

This discrete simulation model is sufficient to model and simulate any kind of digital logic on the gate level, or above.

The only limit may be seen in the continuous-space problems, although custom-made components that work with continuous signals may also be easily written in Java.

REFERENCES

[1] Norman Handrich, “HADES tutorial”, University of Hamburg, 2006
[2] "IEEE Standard Multivalue Logic System for VHDL Model Interoperability (Stdlogic1164)," *IEEE Std 1164-1993*, 1993
[3] MCS-85 User's Manual, Intel Corporation, 1976
[4] 8085Simulator, http://8085simulator.codeplex.com/
[5] GnuSim8085, http://gnusim8085.org/
[6] J8085 Simulator, http://sourceforge.net/projects/j8085sim/ S.Haykin, *Neural Networks*, New York, IEEE Press, 1994.
[7] Karola Krönert, Ullrich Dallmann: JavaFSM - Animation und Simulation von Mealy- und Moore-Schaltwerken. Studienarbeit (ps.gz) (Mai 1997)
[8] Norman Hendrich: From CMOS-Gates to Computer Architecture: Lessons Learned from Five Years of Java-Applets. Proceedings of the 4th European Workshop on Microelectronics Education, EWME 2002, Baiona, 23-24 May 2002
[9] Norman Hendrich: A Java-based Framework for Simulation and Teaching, Proceedings of the 3rd European Workshop on Microelectronics Education, EWME 2000, Aix en Provence, France, 18-19 May 2000, Kluwer Academic Publishers
[10] Norman Hendrich: HADES: The Hamburg Design System, ASA'98, European Academic Software Award/Alt-C Conference, Oxford, 19-21 Sep. 1998

Algorithms for scheduling of resource-constrained jobs

Ivaylo Penev[1], Milena Karova[2]

Abstract – **The paper presents two heuristic algorithms for scheduling resource constrained jobs in an environment, consisting of identical connected computers. The first algorithm uses sorted list of jobs and is applicable to the particular case when the number of jobs is not greater than the number of available computers. For the general case of the problem with arbitrary number of jobs and computers a genetic algorithm is proposed. The experimental results present comparison of solutions, obtained by the algorithms, performed on test data sets with the exact solutions, produced by a classical branch and bound algorithm.**

Keywords – **job-scheduling, resource constraints, heuristics, genetic algorithm.**

I. INTRODUCTION

The goal of the resource-constrained job-scheduling problem is constructing a schedule with starting moments for each job on an available computer from the environment, such that preliminary defined limiting criteria are satisfied and extreme value (minimum or maximum) of objective function is achieved.

The following common approaches and algorithms for solving the resource-constrained job-scheduling problem are known.

- Algorithms for producing exact solution;

The most frequently used exact method is the branch and bound algorithm. The algorithm performs exhaustive check of all branches of a tree with possible solutions (i.e. all possible schedules). For reducing the space of solutions the tree is usually pruned by the usage of a bound value for the examined branch.

The results of many realizations of the branch and bound algorithm are published, for example for solving the resource constrained project scheduling problem [1].

- Algorithms for producing good (near the exact) solution;

Most scheduling problems are NP-hard [2]. For finding good solution, which is close to the exact one, heuristic approaches and algorithms are applied. Multiple heuristic rules and algorithms for solving scheduling problems are suggested, summarized for example in [3]. Genetic algorithms are among the most often algorithms used. A genetic algorithm, solving common resource-constrained scheduling problem is presented for example in [5].

[1]Ivaylo Penev is with the Department of Computer Sciences and Technologies at Technical University - Varna, 1 Studentska str., Varna 9000, Bulgaria, E-mail: ivailopenev@yahoo.com.

[2]Milena Karova is with the Department of Computer Sciences and Technologies at Technical University - Varna, Bulgaria.

The paper presents two heuristic algorithms for scheduling of resource-constrained jobs. The first algorithm is based on sorted lists of jobs. The second one is genetic algorithm. The solutions, obtained by these algorithms are compared to the exact solution, produced by classical branch and bound algorithm. The authors demonstrate, that the proposed heuristic algorithms produce near the exact solutions for significantly reduced time.

II. FORMAL DESCRIPTION OF THE RESOURCE-CONSTRAINED JOB-SCHEDULING PROBLEM

The resource-constrained job-scheduling problem could be formally defined by the following sets.

1. Jobs

$J = \{j_i \mid i \leq n\}$ - set of jobs

$j_i = \{v_i, q_i, w_i\}$ – set of sub jobs for each job

v_i – reading data from common resource

q_i – calculation operations

w_i – storing results into the common resource

$t(v_i), t(q_i), t(w_i)$ – completion time for each sub job of job j_i

C_i – completion time of job j_i

2. Resources

Sub jobs v_i and w_i uses common resource. Sub job v_i requires $r(v_i)$ units and sub job w_i requires $r(w_i)$ units of the common resource.

3. Temporal constraints

$S = \{S_i \mid i \leq n\}$ – set of starting moments along the time axis for each job

$S(v_i) = \{S(v_i) \mid i \leq n\}$ – set of starting moments for sub job v_i

$S(q_i) = \{S(q_i) \mid i \leq n\}$ – set of starting moments for sub job q_i

$S(w_i) = \{S(w_i) \mid i \leq n\}$ - set of starting moments for sub job w_i.

The sub jobs of each job are performed in the following order – reading data, calculation operations, storing data, i.e.

$$\forall_{j_i} \in J \; \exists \; S_i \rightarrow (S_i + t(v_i) \leq S(q_i)) \cap (S(q_i) + t(q_i) \leq S(w_i))$$

The completion time of each job (which means completion of all sub jobs) in a schedule is defined as $C_i = S(w_i) + t(w_i)$. The completion time of the last job in a schedule is $C_{max} = max(C_i)$ and is called ***makespan***.

4. Mathematical model of the resource-constrained job-scheduling problem

Assuming the above notations the problem for scheduling **n** jobs in an environment of **m** identical connected computers, using constrained common resource could be formally described:

$min(Cmax)$, $i = 1, \ldots, n$ – minimum makespan - objective function (1)

$(S_i + t(v_i) \le S(q_i)) \cap (S(q_i) + t(q_i) \le S(w_i))$ - temporal constraints (2)

$$\sum_{i=1}^{n} (r(v_i) \cup r(w_i)) \le 1$$ – resource constraints (3)

A feasible solution of the scheduling problem is assignment of starting moments to all sub jobs, i.e. $(S(v_1), S(q_1), S(w_1), \ldots, S(v_n), S(q_n), S(w_n))$, which satisfies the temporal constraints (2) and the resource constraints (3). The **exact solution** is the minimum value of all possible values of the completion time of the last job in the schedule, i.e. the exact solution guarantees minimum makespan *Cmax* in the schedule.

III. HEURISTIC ALGORITHMS FOR SOLVING THE RESOURCE-CONSTRAINED JOB-SCHEDULING PROBLEM

A. Algorithm with sorted lists of jobs for the particular case $n \le m$ of the resource-constrained scheduling problem

The algorithm uses two lists with jobs. It is presented by the following pseudo code:

List1 ← j_i, for $\forall i<j \rightarrow t(v_i) + t(q_i) \ge t(v_j) + t(q_j)$
$S_1 = 0$
for each job in List1 do
    $S_i = S_{i-1} + t(v_{i-1})$
endfor
List2 ← List1, for $\forall i<j \rightarrow S_i + t(v_i) + t(q_i) \le S_j + t(v_j) + t(q_j)$
$S_1 = 0$
j_1 ← List2 – Get first job from List2
$t_1 = S_1 + t(v_1) + t(q_1) + t(w_1)$
for each job in List2 do
    if $S_i+t(v_i)+t(q_i)>=S_{i-1}+t(v_{i-1})+t(q_{i-1})+t(w_{i-1})$
        $t_i = S_i+t(v_i)+t(q_i)+t(w_i)$
    else
        begin
            $\Delta t(w_i) = t_{i-1} – (S_i + t(v_i) + t(q_i))$
            $t_i = S_i + t(v_i) + t(q_i) + \Delta t(w_i) + t(w_i)$
        end
    endif
endfor
j ← Get last job from List2
$C_{max} = j$

B. Genetic algorithm for the general case $n \le m$ of the resource-constrained scheduling problem

The genetic algorithm iteratively chooses the best candidates (individuals) of a set of possible solutions. The choice is based on the calculation of an objective function of the genetic algorithm, which presents the number of time units, in which two or more jobs access the common resource concurrently. At each iteration the algorithm shifts an arbitrary job on the time axis and calculates the objective function. If the shifting decreases the objective function's value, the solution is saved in the population.

B. 1. Distribution of the tasks between the computers

To assign jobs to a computer a list with the jobs, sorted by execution time is used (i.e. t(vi) + t(qi) + t(wi)). Each job from the list is afterwards assigned to a computer. After the first m jobs are assigned to m computers, the assignment of jobs continues in reverse order.

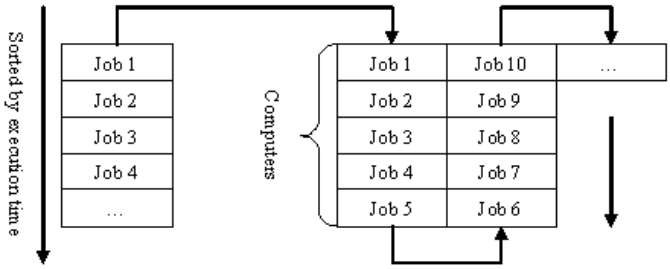


Fig. 1. Distribution of jobs between the computers

B.2. Coding of the chromosomes

Each chromosome presents a list with all available computers. Each computer is assigned to a job with starting moment and known times for reading data, calculation operations and storing results.

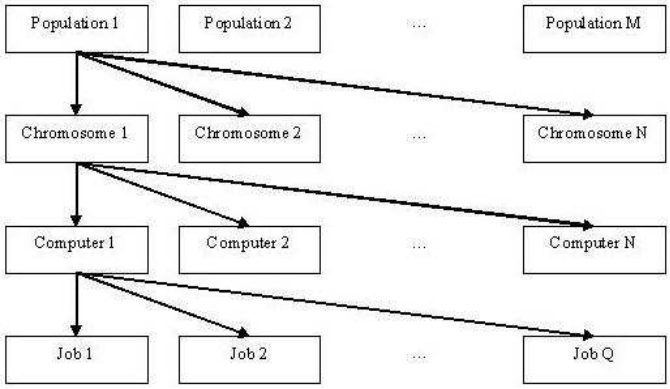


Fig. 2.Structure of the genetic algorithm

Each chromosome is presented as a string. Each element of the string is coded with one of two possible values – 0 or 1. The time units, in which the job uses the common resource is presented with 1, the others are presented with 0.

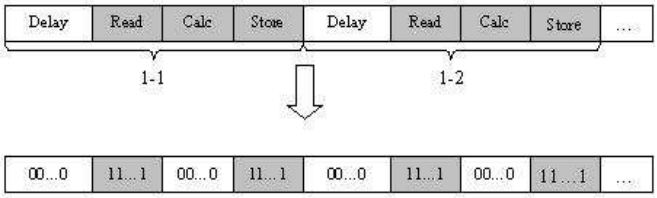


Fig. 3. Structure of a chromosome

B.3. Objective function of the genetic algorithm

The objective function of the chromosome is calculated by intersection of the strings of each pair of computers from the chromosome. The number of "1"-s is added to the sum of time units with concurrent access to the constrained resource.

Algorithm for calculating the objective function's value of the genetic algorithm:

foreach chromosome
sum_of_time_units_with_concurrent_access = 0
for i = 1 to m-1
for j = i + 1 to m
result_string = $string_i \cap string_j$
sum_of_time_units_with_concurrent_access = sum_of_time_units_with_concurrent_access + number of "1"-s in result_string
endfor
endfor
end foreach

B.4. Genetic operators

Mutation is performed by choosing an arbitrary job from an arbitrary computer of the chromosome. The job is then shifted on the time axis. The objective function is calculated. If the number of time units with concurrent access to the common resource is decreased, the shifted job is saved in the population.

IV. Experimental results

The branch and bound algorithm and the two presented algorithms are tested for constructing schedules with different number of jobs. Times for reading data, calculation operations and storing results for each job are generated. For each number of jobs 100 attempts are performed. The minimum makespan of the schedule and the necessary time for constructing the schedule are measured.

Table I shows the deviation of the makespans, produced by the algorithm with sorting and the genetic algorithm from the corresponding exact makespan, obtained by the classical branch and bound algorithm. The deviations are measured in percents.

Table I

Deviations of the heuristic algorithms' makespans from the branch and bound's makespan

Number of jobs	Sorting (%)	Genetic algorithm (%)
2	0	0
3	0,74	0,95
4	1,92	3,07
5	1,19	0,89

Table II shows comparison of the average time, necessary to the algorithms to produce the makespan. The following notations are used:

BB – Branch and bound algorithm
Sorting – Algorithm with sorting
GA – Genetic algorithm

Table II

Comparison of times for producing the makespan

Jobs	BB (ms)	Sorting (ms)	GA (ms)
3	3	3,29	23,10
4	63,85	2,68	288,67
5	305,67	2,85	275,56
7	11437	2,15	6645
8	110554	6,69	880
9	6459285,35	65545	10880

The next figures demonstrate the trend of the time, necessary to all the algorithms for constructing the schedule with the minimum makespan for different number of jobs.

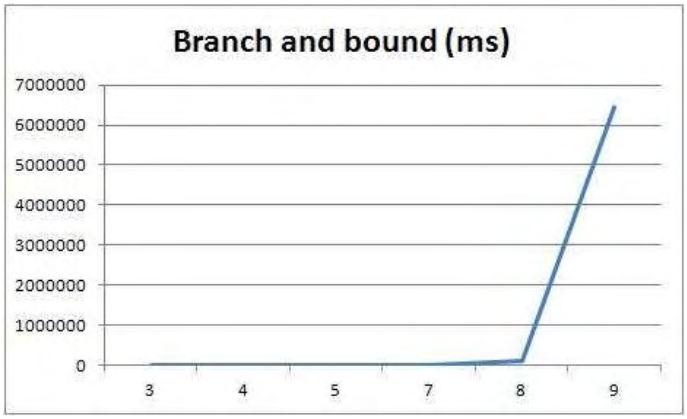


Fig. 4. Trend of time, necessary to the branch and bound algorithm for producing exact minimum makespan

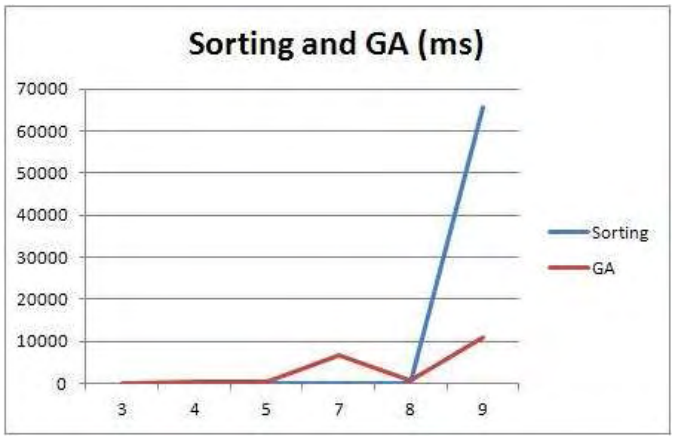


Fig. 5. Trend of time, necessary to the algorithm with sorting and the genetic algorithm for producing near the exact minimum makespan

V. Conclusions

The experimental results prove, that both the proposed algorithms produce near the exact minimum makespan of the constructed schedule. The algorithm with sorted lists of jobs is very easy to implement and produce makespan for significantly reduced time in comparison with the branch and

bound algorithm. The genetic algorithm constructs schedule for least time in comparison with the other algorithms and is applicable to the general case of the scheduling problem with arbitrary number of jobs and identical connected computers.

The presented algorithms will be used in a real portfolio management system for evaluation of various financial objects' risk. Simulation analyses must be performed with historical data of portfolios. Typically there are no data dependencies between the data of the portfolios, which makes possible parallel execution of the estimations. The historical data are stored into common resource. The concurrent access to the common resource decreases the efficiency of the parallel execution [4]. The proposed heuristic algorithms will be used to construct schedule with jobs, estimating portfolios, which ensures completing of all jobs for polynomial time.

REFERENCES

[1] Dorndorf, U., W. Pesch, T. Phan-Huy. A Time-Oriented Branch-and-Bound Algorithm for Resource-Constrained Project Scheduling with Generalised Precedence Constraints.

[2] Garey, M., D. Johnson. Computers and intractability: A Guide to the Theory of NP-Completeness. Freeman. 1979.

[3] Hartmann, S., D. Briskorn. A survey of variants and extensions of the resource-constrained project scheduling problem. European Journal of Operational Research, vol. 207, iss. 1, pp. 1 – 14, 2010.

[4] Penev, I. Realization of Portfolio Management System in a Distributed Computing Environment. Proceeings of Fifth International Scientific Conference Computer Science'2009, ISBN: 978-954-438-853-9, pp. 291 ÷ 296.

[5] Wall, M., B., A Genetic Algorithm for Resource-Constrained Scheduling, partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mechanical Engineering, Massachusetts Institute of Technology, 1996.

Hybrid Automatic Repeat Request (HARQ) Overview

G. Marinova[1], S. Yordanova[2], N. Kostov[3]

Abstract:- **This article mainly investigates the combining schemes for hybrid automatic retransmission request (HARQ) protocols in communication systems. Based on of HARQ combining, we classify the HARQ combining schemes into three types. We are discussed advantages and disadvantages of HARQ systems.**

Key words: **ARQ, HARQ I, HARQ II, HARQ III, RB-HARQ.**

I. INTRODUCTION:

The combination between ARQ and FEC schemes is known as a hybrid automatic repeat request (HARQ).

HARQ systems are used to enhance system efficacy and efficiency and are employed in modern data communication systems.

In FEC schemes, by means of the attempts to correct, the behavior of the corrected word is evaluated and if the code word is found valid, then it is assumed as such. If errors have been found in the corrected word, then most likely there has been a greater number of errors in it than a FEC is capable of correcting. ARQ is used for verification of retransmission. HARQ systems provide higher reliability than FEC systems and feature higher efficiency than ARQ systems [1]. Depending on the schematic realization, the following types of HARQ systems are available – HARQ Type I, HARQ Type II, HARQ Type III and RBHARQ.

II. TYPE I HARQ SYSTEM

Fig. 1 shows Type I HARQ system. It is used for simultaneous error detection and correction [2, 3].

[1]G. K. Marinova is with the Technical University of Varna, Department of Computer Science and Engineering, Studentska Street 1, Varna 9010, Bulgaria, E-mail: gin_kaleva@abv.bg
[2]S. M. Yordanova is with the Technical University of Varna, Department of Computer Science and Engineering, Studentska Street 1, Varna 9010, Bulgaria, E-mail: slava_y@abv.bg
[3]N. T. Kostov is with the Technical University of Varna, Department of Communication Engineering and Technologies, Studentska Street 1, Varna 9010, Bulgaria, E-mail: n_kostov@mail.bg

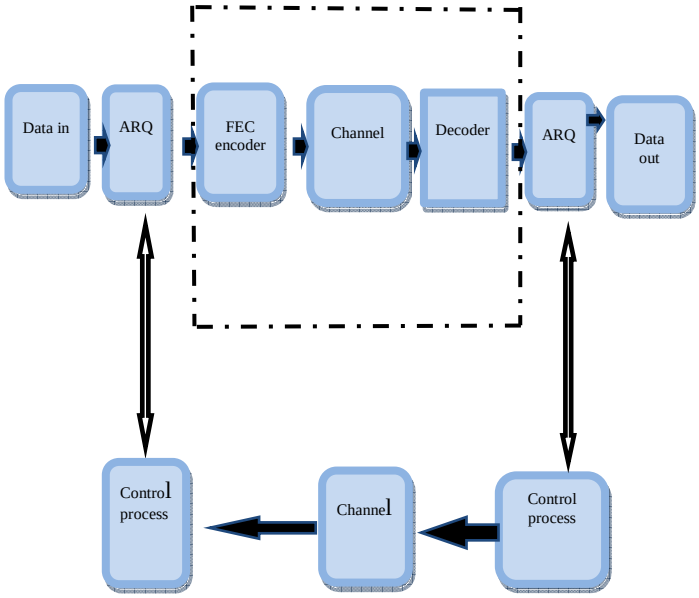


Fig. 1 Block scheme Type I HARQ system.

A code is used for error detection and a different one for correction. When receiving a packet containing errors, the receiver first tries to correct them. If correction is unsuccessful, the packet is rejected and a retransmission is required. The code speed of this system is fixed. This shortcoming may be overcome by optimizing channel conditions.

Type I HARQ is appropriate for systems with constant noise level and with interferences present in the channel.

Advantages:

- Type I HARQ provides higher throughput than an ARQ scheme.

Disadvantages:

- With noise in the channel the correcting ability of the code may appear to be insufficient.
- The number of retransmissions increases but the HARQ system throughput decreases.

III. TYPE II HARQ SYSTEM

Fig. 2 illustrates a scheme of Type II HARQ system. In this case a buffer with memory is required. Two basic systems of Type II are known [3,4]:

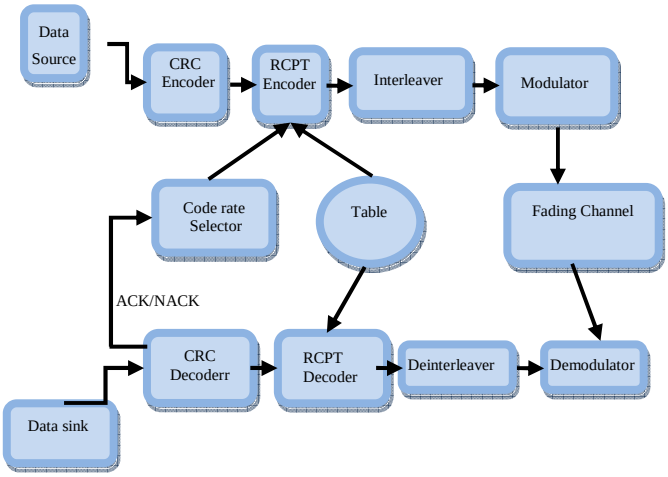


Fig. 2 Block scheme Type II HARQ system.

- Chase combining
- A system with excess adding

The main idea of the Chase combining is in sending numbers of a certain packet with the quotations in each packet of encoded data. It is possible for the decoder to accomplish decoding by measuring the signal/noise proportion from a previous decoding.

Type II HARQ system with excess adding is characterized by the following:

If there is an error, the word received is stored in a buffer and NACK is sent to the transmitter. The latter sends a block of bits for initial message verification and makes an attempt for error correction.

Advantages:

- The system adapts itself to the channel characteristics so that jamming can be overcome.
- At high speed, the system with excess adding is better than the one with Chase.

It enhances permeability in comparison with Type I.

Disadvantages:

- The amount of information increases due to the added code symbols sent along with the packet.
- The format retransmission depends on the applied strategy and on the error recovery code.
- A buffer with greater memory is required resulting in increase in the price of the system.
- Greater complexity of decoding in comparison with Type I HARQ.

IV. Type III HARQ system

Fig. 3 shows a scheme of a Type III HARQ system. It is based on independent decoding. The system is of adaptable structure and determines the amount of the additional information. Under good channel conditions FEC code is used [6].

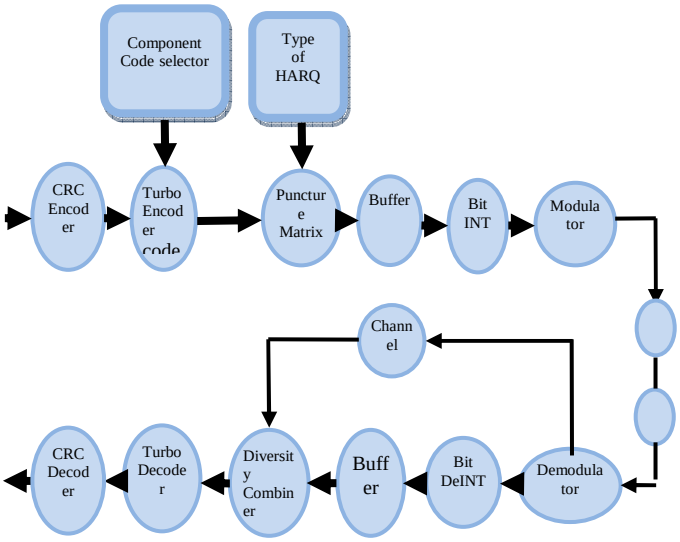


Fig. 3 Block scheme Type III HARQ system.

According to the type of excess these schemes are divided into two groups:

- with one variant of excess;
- with several variants of excess.

Advantages:

- In comparison with Type I permeability is improved;
- In comparison with Type II efficiency is enhanced [7];
- Type III HARQ system is of adaptable structure, i.e. the number of abbreviations is decreased to the minimum.

Disadvantages:

- with noisy channels the amount of excess information increases;
- complex algorithms for encoding and decoding are applied;
- under good channel conditions Type III HARQ system has got lower permeability in comparison with Type II HARQ system.

V. Reliability Based Hybrid ARQ (RBHARQ)

With the RBHARQ, bits are retranslated arbitrary to the receiver by using the calculated bit of reliability [7].

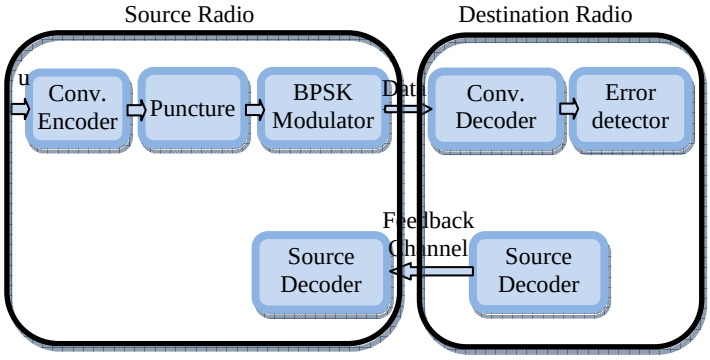


Fig. 4 Block scheme RBHARQ system.

One of the possible algorithms is the MAP algorithm [8] where for each bit of information the logarithm of a posteriori probability is calculated by formula (1) in which y is the received code word with noise [6].

$$L(u_k) = \log\left(\frac{P(u_k = +1 \mid y)}{P(u_k = -1 \mid y)}\right), \qquad (1)$$

There are three techniques for reliability assessment [8,9]:

- logarithmic likelihood ratio;
- determination of erroneous bit probability;
- error assessment by using reliable information.

Advantages

- The employment of reliable information enhances network permeability;
- Network general usefulness for choosing functions is increased.

Disadvantages:

- This type of systems requires greater memory necessary to support the combining of the transferred packets
- Under bad channel conditions and with the use of convolution codes the time for analysis increases [7].

VI. CONCLUSION

With digital signal receiving and decoding it is possible that an error in certain bits or groups of bits may occur. Each error distorts the message; thus the greater the number of errors, the more unusable the received information becomes.

Error likelihood increases with the decrease of the signal/ noise proportion at the input of the input device. There are many factors leading to errors in digital signal decoding. To decrease the possibility of errors in the digital signal, protection against errors is applied by using certain algorithms. The protection against errors is realized by forming a digital flow by the multiplexer. So far various solutions to this problem have been suggested but with the advances of computer technologies this issue is becoming more and more serious.

REFERENCE

[1] J. Moreira, P. Farrell, "Essentials of error control coding" John Wiley & Sons Ltd 2006

[2] S. Lin and P. S. Yu, "A Hybrid ARQ Scheme with Parity Retransmission for Error Control of Satellite Channels," IEEE Transactions on Communications, Vol. COM-30, No. 7, pp. 1701-1719, July 1982.

[3] S. Lin, D. J. Costello, Jr., Error Control Coding: Fundamentals and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[4] S. Lin, D. J. Costello and M. Miller, "Automatic-Repeat-Request error-control schemes," IEEE Commun. Mag., vol. 22, no. 12, pp. 5–17, Dec. 1984.

[5] S. Kallel, "Complementary Puncture Convolutional Codes (CPC) and Their Applications," IEEE Transactions on Communications, Vol. 43, No. 6, pp. 2005-2009, June 1995

[6] J. J. Metzner, "Improvements in Block-Retransmission Schemes", IEEE Transactions on Communications, Vol. COM-27, No. 2, pp. 524-532, February 1979.

[7] Roongta and J. M. Shea, "Reliability-based hybrid ARQ using convolutional codes," in Proc. IEEE International Conference on Communications (ICC 2003), Anchorage, Alaska, May 2003, pp. 2889–2893.

[8] J. C. Fricke, H. Schoeneich, and P. Hoeher "Reliability-Based HARQ using Word Error Probabilities", in Proc. NEWCOM-ACoRN Joint Workshop, Vienna, Austria, Sep. 2006.

[9] 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, January 2008*

Digital information transfer systems an overview

G. Marinova[1], S. Yordanova[2], N. Kostov[3], B. Nikolov [4]

Abstract:- **Digital information transfer systems are used techniques based either on Automatic Repeat Request (ARQ) or Forward Error Correction (FEC) method. ARQ systems constitute a simple and efficient method for providing highly reliable transfer of messages from the source to the user over a variety of transmission channels. ARQ systems are therefore widely used in data communication systems that are highly sensitive to errors.**

Key words: **FEC, ARQ, HARQ .**

I. INTRODUCTION

Apart from high speed, the modern requirements to digital information transfer systems include high reliability of the information transferred. Reliability is of vital importance in the theory of encoding, particularly in terms of using codes capable of detecting and correcting errors which have occurred in the process of information transfer along the communication channel.

II. A CODE AND ENCODING

Nowadays the problem of enhancing information reliability is of particular relevance and its solution is accomplished by means of diverse methods and means. Theory of encoding is part of the engineering sciences and mathematics which study the reliability of data transfer as well as data storage. Encoding is a method for converting information by implementing an algorithm in such a way that it has nothing to do with the original and no one else can read it. In the theory of encoding attention is focused on finding optimal codes for a certain purpose. For fast information transfer codes of less length are required. For the transfer of a greater number of messages the required codes should have a greater number of words. The encoding algorithms should provide the opportunity for retrieving the initial information required. By implementing codes additional information is added to the initial data in such a way that the receiving party alone finds and corrects the likely errors.

[1]G. K. Marinova is with the Technical University of Varna, Department of Computer Science and Engineering, Studentska Street 1, Varna 9010, Bulgaria, E-mail: gin_kaleva@abv.bg
[2]S. M. Yordanova is with the Technical University of Varna, Department of Computer Science and Engineering, Studentska Street 1, Varna 9010, Bulgaria, E-mail: slava_y@abv.bg
[3]N. T. Kostov is with the Technical University of Varna, Department of Communication Engineering and Technologies Studentska Street 1, Varna 9010, Bulgaria, E-mail: n_kostov@mail.bg
[4] B. Nikolov is with the Technical University of Varna, Department of Department of Communication Engineering and Technologies Studentska Street 1, Varna 9010, Bulgaria, E-mail: b.nikolov@tu-varna.bg

III. METHODS OF ERROR CONTROL

Several methods for control of the errors in data transfer systems are known.

Forward Error Correction (FEC)

The methods for channel coding are also called methods for error correction in forward direction FEC. All FEC methods are based on the principle of data encoding prior to their transfer by introducing a certain excess of information. Based on this excess the receiver, finding out this excessive information, manages to decode, and in many cases, to correct the errors introduced to the data during their transfer along the channel. The purpose of the channel coding is to process the information sequence in such a way, that the receiving part is capable of detecting the transfer errors and possibly of correcting some of them thus improving the quality of the received signal.[1]

The simplest FEC method is data retransmission upon error detection, which in turn unnecessarily prolongs the transfer time and therefore channel work load. Modern mobile radio networks use FEC methods based on the implementation of channel codes. The latter being capable of only detecting errors are referred to as detecting codes, whereas codes which can not only detect errors but also correct them are known as correcting codes.[1]

There are two basic groups of detecting and correcting codes – block codes and convolution codes.

- Block codes are FEC codes which provide the opportunity to detect and correct a limited number of errors without data retransmission. Special bits are introduced into them (called bits for odd-even check) to the blocks of data bits by means of which the so called code words, or code blocks, are created.[2]
- In the block encoder is done encoding *k* information bits into *n* code bits. Fig. 1.

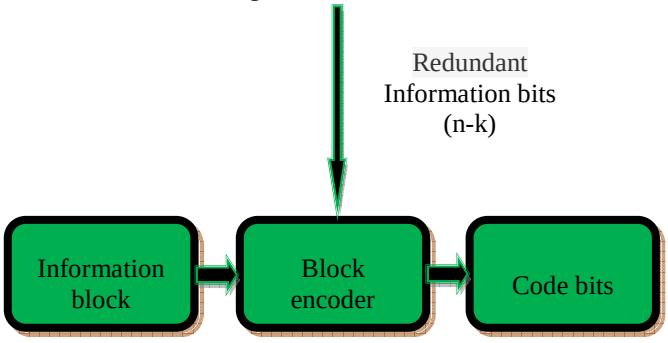


Fig.1 Block coding

➢ With convolution codes the information sequence is immediately converted coordinately into a certain code sequence without being divided into blocks or being encoded. The convolution encoder receives at its input *k* bits, the output is *n* bits, and the proportion *k/n* is called an encoding level. Convolution encoding is non-linear. Memory for data processing is required. Each output bit c_n depends not only on the input bit b_n, but also on several previous bits. The number of the input bits required for obtaining the encoded output is called code restriction. Encoding is accomplished by using the following equations (1,2):

$$c_{2k} = b_k \oplus b_{k-3} \oplus b_{k-4} \quad (1)$$

$$c_{2k+1} = b_k \oplus b_{k-1} \oplus b_{k-3} \oplus b_{k-4} \quad (2)$$

Where ⊕ is the sum module 2, k ∈{0, 1, 2, ……, 189} and b_k=0 for -∞ ≤κ <0.

By using a shifting register the scheme has the structure shown in fig. 2.

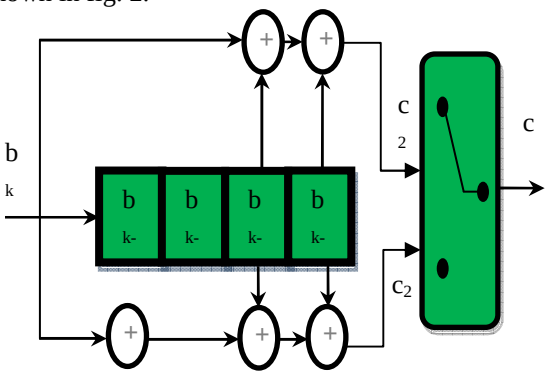


Fig 2. Scheme of coding

It is worth mentioning that the frequency of the output sequence is twice as high as that of the input. Convolution codes are used mainly in digital networks as part of the so called Concatenated Coding.

Channel coding introduces excess to the information transferred. Increase in the speed of transfer along the channel and consequent increase in the bandwidth are required.

Automatic Requirement Request (ARQ)

With errors present in the vector received, the functions of the receiver are connected with location and correction of the former. The improvement of the communication channel characteristics of FEC systems is achieved by means of additional information to the data transferred. The application of ARQ is connected with enhancement of the efficiency of permeability of the physical channels. The purpose is to decrease the losses from official information transfer.[3]

Hybrid Automatic Repeat Request (HARQ)

The combination of both FEC and ARQ techniques is known as Hybrid Automatic Repeat Request (HARQ) which is used in modern data communication systems. Depending on the conditions of operation there are three basic types of HARQ systems:

- "Stop and Wait"
- "Go back – N"
- "Selective Repeat"

"Stop and Wait" HARQ system

It is mandatory for this mode of operation that the system requires a positive acknowledge (ACK) or a negative acknowledge (NACK) when transmitting a block or a codeword. No word is transmitted unless the previous one is received correctly. In this case packet buffering is required.

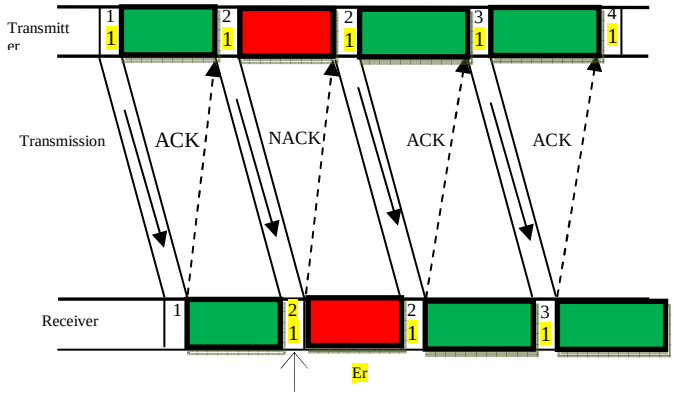


Fig.3 Stop and Wait scheme

"Go back – N" HARQ system

The system is connected with the buffering of more than one packet. In the cases of NACK reply all packets that follow are ignored. The process continues until a correct packet is received. Fig.4

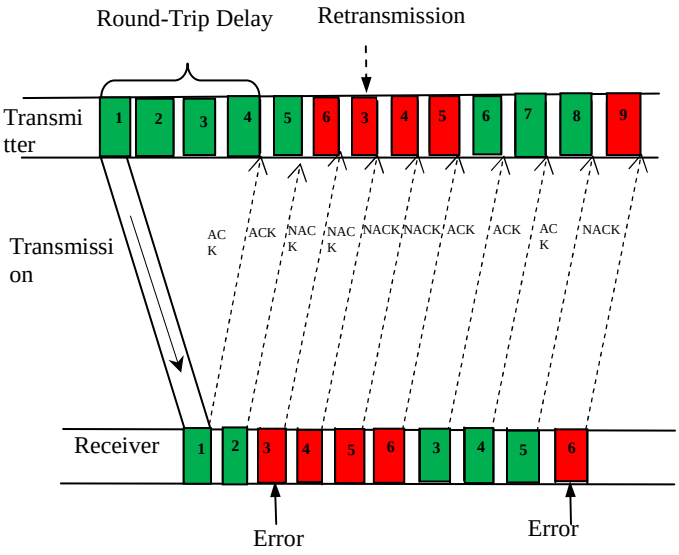


Fig.4 Go back –N scheme

"Selective Repeat" HARQ system

This system features transmission of the missing packets only. Fig. 5

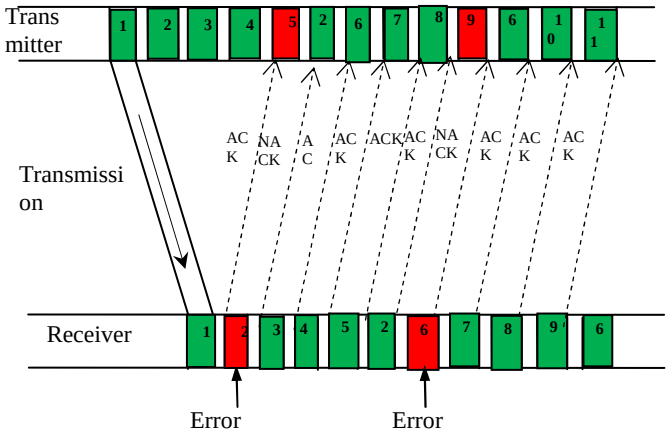


Fig. 5 Selective repeat scheme

To increase the efficacy and the efficiency of the system HARQ systems are used. They provide higher reliability in comparison with a FEC system and feature higher efficiency than an ARQ one.

IV. Conclusion

A great part of the theory of binary codes is based on the assumption that each of the symbols is distorted by the noise independently and therefore the errors of a particular combination depend on the number of errors in the latter. There is a certain limit for the efficiency of the systems in which only error detection is used. However, in practice, there are always accidental errors. The purpose of the codes is to make errors detectable and to correct them if possible.

Reference

[1] S. Lin, D. Costelo,'Error Control Coding – Fundamentals and Applications' Prentice-Hall, 1983

[2]J. Moreira, P. Farrell,"Essentials of error control coding" John Wiley & Sons Ltd 2006

[2] S. Lin and P. S. Yu, "A Hybrid ARQ Scheme with Parity Retransmission for Error Control of Satellite Channels," IEEE Transactions on Communications, Vol. COM-30, No. 7, pp. 1701-1719, July 1982.

[3] S. Lin, D. J. Costello and M. Miller, "Automatic-Repeat-Request error-control schemes," IEEE Commun. Mag., vol. 22, no. 12, pp. 5–17, Dec. 1984.

[4] S. Kallel, "Complementary Puncture Convolutional Codes (CPC) and Their Applications,"IEEE Transactions on Communications, Vol. 43, No. 6, pp. 2005-2009, June 1995

[5] Roongta and J. M. Shea, "Reliability-based hybrid ARQ using convolutional codes," in Proc. IEEE International Conference on Communications (ICC 2003), Anchorage, Alaska, May 2003, pp. 2889–2893.

[6] J. C. Fricke, H. Schoeneich, and P. Hoeher "Reliability-Based HARQ using Word Error Probabilities", in Proc. NEWCOM-ACoRN Joint Workshop, Vienna, Austria, Sep. 2006.

[7] 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, January *2008*

Towards applicability of agile software development methodologies

Aleksandar Dimov[1], Stavros Stavru[1], Dessislava Petrova-Antonova[1] and Iva Krasteva[2]

Abstract – ***Agile software development has been an active research area for more than a decade. Nevertheless, straightforward application of agile methods in real practice is hampered by some challenges which need to be resolved both by academia and industry. One of the biggest issues in this respect concerns the lack of structured guidelines for application and deployment of agile methods and techniques. Usually this is performed in an ad-hoc manner, depending on the particular project that is being executed. In this paper we present the initial framework of our ongoing research work on empirical research of applicability of agile software development methods and how to develop a methodology for their deployment.***

Keywords – ***Agile software development, Organizational values, Software processes, Software engineering.***

I. INTRODUCTION

Recently in software engineering emerged a trend in research of how to organize the development efforts (in terms of design, programming, testing, etc.) in order to receive predictable results within time and budget limits of the organization. In this respect, the software processes define different kinds of activities, artifacts, roles, etc. needed for successful execution of a software development process.

A classical example for software development process is given in [6]. In this case, activities that software developers should undertake when building software intensive systems, are rigorously divided into five core workflows: requirements capture, analysis, design, implementation and testing. However, real practice has shown that the multitude of strict and formal procedures of classical process, in many cases put an extra burden and actually impedes software development. An alternative of such strict and formal processes are agile processes and methods for software development [7]. The latter do not insist on strict policies or planning of the software development process from long term point of view. Instead, they focus on collaborative work between members of the development team and on adaptability of process to the specific needs of the project during life-cycle. Moreover, agile methods put focus on availability of executable software and interaction with stakeholders.

Generally speaking, in order to define a software development method as agile, it should conform to the main values and principles of the so-called agile manifesto [8]. This has been an active research area for more than a decade and it is still a hot topic to work on [1], [2]. The reason for that might be found not only in its potential to overcome the shortcomings of the traditional methods of software development, but also in the wide acceptance of its principles and techniques from the industry [1], [2], [3], [4]. Referring to some recent industrial reports [3], [4], the tendency towards the increasing number of organizations, interested in or already adopting agile methods and practices, is strong and is likely to stay for the coming years. Therefore, the research in this field has both the potential to impact significantly the software development industry and the perspectives for a promising future.

There are many challenges to be addressed in regard to agile software development [1], [2]. Some of them address development of new methodologies, practices and tools to support application of the methods. Others, which are the main interest to us, touch the topic of successful adoption and assessment of agile methods and practices [2].

In this paper we are going to present the initial framework of our ongoing study on empirical research of applicability of agile software development methods. This study should end up with development of a methodology for deployment of agile methods in particular software development companies.

The remainder of the paper is organized as follows: Section 2 describes the motivation that backs our research efforts; Section 3 presents our ongoing research work in applicability of agile software development methods; Section 4 presents the research that directly relates to ours and finally, Section 5 concludes the paper and states the main directions for further research in the area.

II. MOTIVATION

The main motivation of our work is to raise the opportunities for application of agile methods, specifically in Bulgarian companies.

Main principles and values that serve as a basis for agile methods are focused on software development in a concurrent

[1]Aleksandar Dimov is with the Department of Software Engineering in Faculty of Mathematics and Informatics, University of Sofia "St. Kl. Ohridski", Tsarigradsko Shose 125 blvd, bl. 2, 1113 Sofia, Bulgaria, E-mail: aldi@fmi.uni-sofia.bg

[2]Stavros Stavru is with the Department of Software Engineering in Faculty of Mathematics and Informatics, University of Sofia "St. Kl. Ohridski", Tsarigradsko Shose 125 blvd, bl. 2, 1113 Sofia, Bulgaria, E-mail: stavross@fmi.uni-sofia.bg

[2]Dessislava Petrova-Antonova is with the Department of Software Engineering in Faculty of Mathematics and Informatics, University of Sofia "St. Kl. Ohridski", Tsarigradsko Shose 125 blvd, bl. 2, 1113 Sofia, Bulgaria, E-mail: d.petrova@fmi.uni-sofia.bg

[4]Iva Krasteva is with Rila Solutions, acad G. Bontchev Str., bl. 27, Sofia 1113, Bulgaria. E-mail: ivak@rila.bg

business environment. Projects in such environment usually have very fast-changing requirements and limited time for implementation and deployment of the final software system. This way, the final goal of agile methods is to improve organizational processes and to maximize both the profit and the quality of the product. One of the last industrial surveys [9] about applicability of agile methods, made by VersionOne in 2010, points out that software development time is decreased in 66% of the cases, when agile methods are applied and only 5% of the projects suffer from increased development time. Moreover, 87% of the participants in the same survey claim that agile methods has improved management of the dynamic and changing user requirements.

Despite the obvious advantages of agile methods, the problem of low level of reuse in their deployment hampers the broad applicability of this approach [10]. It is very common to adapt the development method (or process), according to the specifics of each software project, even within the same organization. Besides this, the adaptation is made in an ad-hoc manner without well-defined approach and rules. To summarize – deployment of agile methods is a tedious task and is determined by factors as: previous experience, skills and knowledge of individuals. This way many authors in the community identify the need of structural approach towards the knowledge, used in agile methods deployment as a challenge that requires a lot of research efforts [11].

In next sections we present our approach towards solving the problem specified above.

III. A Proposed Method for Research in Applicability of Agile Software Development

In our research work we are focusing on examination of current state of the art and improvement in applicability of agile software development methods in Bulgarian companies. For this purpose we focus on and make the so-called notion of *organizational value* a main driving force for adoption of agile methods and techniques (fig. 1). Organizational values are usually defined as *latent constructs that refer to the way in which people evaluate activities or outcomes and in the same time* they *drive and regulate both means and ends of the organization* [12], [13].

Organizational values, as a core cognitive element of the organizational culture [7], have been studied in the context of agile software development. As stated in [14], a main driving force towards successful adoption and deployment of agile methods in a wide number of companies is to focus on understanding and appreciating the relation between organizational culture (including organizational values) and agile methods.

In order to achieve this, we plan to undertake the following steps:

I. Develop a new organizational classification technique, and adapt it, in order to reflect the power of organizational values in driving and justifying organizational behavior;

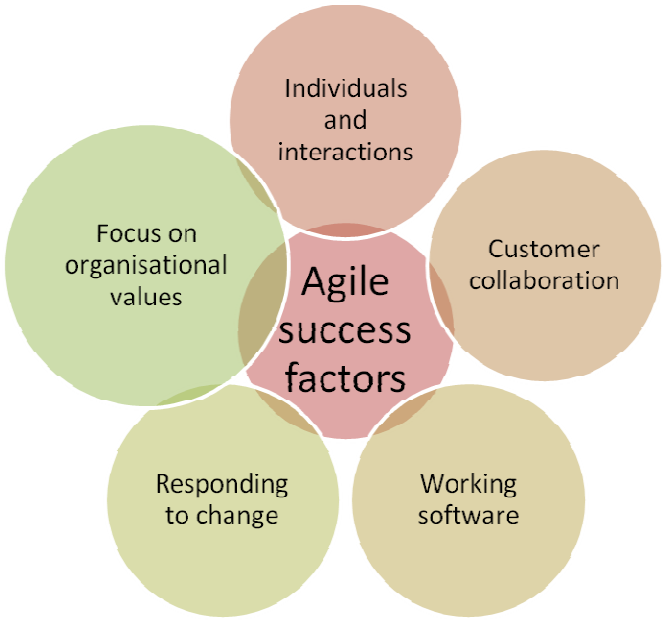


Fig 1: Importance of organisational values in succesful deployment of agile methods

II. Develop a new method for building organizational culture, based on strong organizational values, which includes:
   - A predefined set of formalized and widely applicable organizational values.
   - A repository of organizational metrics, which is applicable in the context of software engineering and associated with concrete organizational values from the aforesaid set.
   - New techniques for the selection, prioritization, propagation and monitoring of organizational values.

III. Develop a new structured method for the selection, introduction and evaluation of agile methods and practices, based on organizational values.

To be able to apply our approach to the Bulgarian software development sector we plan to carry empirical research as suggested by [15], [16]. This way we should inspect the real situation in software development industry, by executing the following activities:

- Develop of a structured questionnaire to assess the extent of applicability of agile methods in each organization.
- Develop a questionnaire to assess the organizational values of each organization.
- Conduct the survey into appropriate number of Bulgarian software development companies.

Results of the surveys will be used to define metrics for organizational values and their correlation with deployment of agile methods in particular companies. This way, software development organizations will become able to measure the effectiveness and agility of their development methodologies. It will also become become possible to improve methods within the organization and to assess the benefit and the extent of this improvement.

Together with this, the impact of organizational values over the behavior of a given organization should also be assessed. In order to do so, appropriate indicators for organizational values should be defined. Here we define an initial list of such of organizational indicators, which is shown in Table I.

TABLE 1
ASSESSING THE POWER OF ORGANIZATIONAL VALUES

	Organizational indicator
I1	Number of organizational values, explicitly stated;
I2	Variety of organizational values, explicitly stated;
I3	Number of organizational metrics continuously measured and associated with concrete organizational values;
I4	The acceptance of organizational values by stakeholders;
I5	The percentage of organizational decisions aligned and consistent with the organizational values;
I6	Continous organizational assessment of organizational behavior based on the level of adherence to organizational values;

IV. RELATED WORK

The next research direction that relates directly to ours concerns successful adoption and assessment of agile methods and practices. There exist a number of frameworks that address this particular challenge and some of them are briefly described in [5]. They include Agile Adoption Framework, Agile Software Solutions Framework, Objectives, Principles and Practices Framework, and some industrial instruments as Comparative Agility and Thoughtworks Agile Assessment Survey. These frameworks and tools have defined organizational agility as the degree to which an organization adheres to a specific agile method or a set of practices. As a result, their approach to agile software development adoption and assessment is driven by the motivation to apply as much as possible in terms of existing agile methods and practices, leaving organizational values aside.

Another related research direction concerns Adapting and tailoring agile methods and practices to best fit the organizational and project context. For example, in [17], the authors try to answer the question of *how are agile software development methodologies adapted for use in different contexts* and provide prescriptions for adapting agile development methodologies? This work, describes various sources of agile structures and examine how different structures affect organizations, project outcomes and etc., together with how individuals influence agile methods and practices. In [18], an approach for situational engineering of agile methods is proposed. It is based on current organizational knowledge in adopting agile practices in both internal and external projects. The approach also introduces a knowledge-base that supports the selection of agile practices that are suitable for a particular project. Further, the paper also presents how automated generation of appropriate software development process could be made. However, both these works do not take into account how organizational values affect agile software development.

In our research we are going to focus on typical organizational values as a driving force for application of any kind of process, including agile methodologies.

V. CONCLUSION

Agile software development has proved to be an appealing alternative to classical software development processes. However, the lack of structured guidelines for application and deployment of agile methodologies appears as a serious obstacle in their adoption in industry. Usually this is performed in an ad-hoc manner, depending on the particular project that is being executed. In this paper we set the background directions of ongoing efforts in measuring, improving and streamlining the applicability of agile software development methods. The research will be based on a massive empirical survey of agile adoption in Bulgarian software development companies. The survey will use appropriate indicators in order to assess the impact of organizational values over the behavior of Bulgarian companies.

VI. ACKNOWLEDGEMENT

The research presented in this paper was partially supported by the National research fund in Bulgaria, under contract No. DMU 03-40/2011.

VII. REFERENCES

[1] Dingsoyr, T., Dyba, T., Moe, N.B., Agile Softawre Development: Current Research and Future Directions, Springer-Verlag Berlin Heidelberg, 2010.
[2] Abrahamsson, P., Conboy, K., Wang, X. 'Lots done, more to do': the current state of agile systems development research, European Journal of Information Systems, vol.18, 2009, pp. 281–284.
[3] Softwate Rally, Agile Adoption Trends 2010, Available: http://www.rallydev.com/learn_agile/agile_for_executives/
[4] VersionOne, State of Agile Development Survey 2010, Available at: http://www.versionone.com/state_of_agile_development_survey/10
[5] Soundararajan, S., Arthur, J. D., A Structured Framework for Assessing the "Goodness" of Agile Methods, 18th IEEE International Conference and Workshops on Engineering of Computer-Based Systems, 2011, pp. 14-23.
[6] Jacobson I., G. Booch, J. Rumbaugh, - Unified Software development Process, Addison-Wesley, 1999.
[7] Abrahamsson, P., Salo, O., Ronkainen, J., Warsta, J.: Agile software development methods: Review and analysis. Technical Research Centre of Finland (2002).
[8] Manifesto for Agile Software Development. Available at: http://agilemanifesto.org/ (last visited April 2012).
[9] 2010 State of Agile Development Survey. Available at: http://www.versionone.com/pdf/2010_State_of_Agile_Development_Survey_Results.pdf (last visited: April 2012).

[10] Abrahamsson, P., Oza, N., Siponen, M.: Agile Software Development Methods: A Comparative Review. In : Agile Softawre Development: Current Research and Future Directions. Springer-Verlag Berlin Heidelberg (2010).
[11] Conboy, K., Morgan, L.: Future Research in Agile Systems Development: Applying Open Innovation Principles Within Agile Organisation. In : Agile Softawre Development: Current Research and Future Directions. Springer-Verlag Berlin Heidelberg (2010).
[12] Roe, R.A., Ester, P., Values and work: empirical findings and theoretical perspectives, Applied Psychology: An International Review, vol. 48 (1), 1999, pp. 1-21.
[13] Jaakson, K., Management by values: are some values better than others?, Journal of Management Development, 2010, pp.795-806
[14] Iivari, J., Iivari, N., The relationship between organizational culture and the deployment of agile methods, Information and Software Technology, vol. 53, 2011, pp. 509–520.
[15] D. Sjøberg, T. Dyba, M. Jørgensen, The Future of Empirical Methods in Software Engineering Research, in: Future of Software Engineering (FOSE'07), IEEE, 2007, pp. 358–378.
[16] D. Perry, A. Porter, and L. Votta, Empirical studies of software engineering: a roadmap, In Proceedings of the Conference on The Future of Software Engineering 2000 (ICSE), pp. 345-355.
[17] L. Cao, K. Mohan, P. Xu, B. Ramesh, A framework for adapting agile development methodologies, European Journal of Information Systems, vol. 18 , 2009, pp. 332–343.
[18] I. Krasteva, S. Ilieva, and A. Dimov, Experience-Based Approach for Adoption of Agile Practices in Software Development Projects, Advanced Information Systems Engineering, Springer Berlin / Heidelberg, LNCS vol. 6051, 2010, pp. 266-280.

FPGA (Field Programmable Gate Arrays) – Based Systems-On-a-Programmable-Chip (SOPC) Development for Educational Purposes

Valentina Rankovska

Abstract – **An approach to studying Systems-On-a-Programmable-Chip (SOPC) based on Field-Programmable Gate Arrays and microprocessor cores is presented, considering the complexity of the circuits and design software. Altera's soft microprocessor core – Nios II, design software Quartus II and SOPC Builder and development board DE2-70 are used.**

Keywords – **Systems-On-a-Programmable-Chip (SOPC) design, Field Programmable Gate Arrays - FPGA, Microprocessor cores, Nios II, Education.**

I. INTRODUCTION

The innovative component base in the area of the digital and microprocessor circuits - Field Programmable Gate Arrays (FPGA), intelligent peripherals (controllers, sensors, actuators, etc.), modern interface standards, and the development technology and tools, affect the embedded systems development. Their application areas continuously expand which leads to more complex architecture and increasing requirements to their features and parameters.

At the same time the design approaches, stages and tools develop – they renovate and become more complex, upgrading the traditional ones. On the other hand the incomplete compatibility of hardware and software development tools sometimes leads to errors which are difficult to be found, especially when using several software tools simultaneously.

The FPGA datasheets and development tools literature extends from several hundreds to more than 1000 pages.

All mentioned above makes difficult studying them by the students at the bachelor's degree at the Technical University.

In the present paper an approach for studying development stages of Systems-On-a-Programmable-Chip (SOPC)/ embedded microprocessor systems based on FPGA and microprocessor cores is suggested.

II. SYSTEMS-ON-A-PROGRAMMABLE-CHIP (SOPC)/ EMBEDDED SYSTEMS IN FPGA. MICROPROCESSOR CORES.

Using microprocessor (Central Processor Unit -CPU) cores in FPGA makes possible the implementation of embedded systems with considerable advantages in the cases when they include a large amount of digital circuits or when it is necessary to use many peripheral modules, some of which missing in the traditional microcontrollers.

There are two types of SOPCs according to the architecture of the FPGA – FPGA consisting only of an array of configurable logic blocks and FPGA including configurable logic blocks and hardware implemented blocks – microprocessor(s) and peripherals. Except Altera there are many partners supplying Intellectual Property (IP) cores for software implementation of peripherals – memory controllers, transceivers, arithmetic blocks, signal processing, protocol interfaces, etc.

CPU cores can be classified as follows [1]: according to implementation – *hardware and software*; according to architecture – *based on traditional architecture and unique*; according to the possibilities for reconfiguring in real time – *with hard implemented and reconfigurable architecture*; *commercial and open-cores,* etc.

The study of architecture, features, resources and application of the FPGA in the Microprocessor Circuits and Embedded Systems laboratory in the Department of Electronics of the Technical University – Gabrovo is based on Altera's FPGAs. That is why Altera's soft-microprocessor Nios II is used in the initial study of embedded systems. After achieving some experience the designed systems are expanded with additional peripheral modules and the next stage have to be design of a unique CPU core.

There are many CPU cores for Altera's FPGAs. The features of some of them (recommended on their site) are presented in Table I. The core which occupies the least logical area is Nios II/e. Another advantage is that it is free with software Quartus II v. 9.1 and the following and that is why it is used in the design in section IV.

Nios II is a general purpose RISC processor with 32-bit instructions, address and data busses; 32 general purpose registers; 32 interrupt sources, etc.

III. TECHNOLOGY SYSTEM-ON-A-PROGRAMMABLE-CHIP (SOPC) IN FPGA. DEVELOPMENT TOOLS

Design technology for SOPC with CPU cores follows in general the development stages for arbitrary devices and systems in FPGA [2], [3]. Altera's Quartus II supplies all the stages of the hardware design of the project – creating a project, entering the design, simulation and configuring the FPGA. There are two variants of Quartus II – free and subscription edition. The free edition is used in the laboratory

Valentina Rankovska is with the Faculty of Electrotechnics and Electronics at Technical University of Gabrovo, 4 Hadji Dimitar str., Gabrovo 5300, Bulgaria, E-mail: rankovska@tugab.bg.

TABLE I
FEATURES OF CPU CORES FOR FPGA OF ALTERA

Processor Category / Features	Cost- and Power-Sensitive Processors			Real-Time Processors		Applications Processors	
	ARM® CortexTM- M1	V1 ColdFire	Nios® II Economy	Nios II Standard	Nios II Fast	MP32	ARM Cortex-A9 MPCore
Maximum performance efficiency (MIPS per MHz)	-	0.93	0.15	0.64	1.13	1.15	2.5
Maximum performance (MIPS*(1)* at MHz) Cyclone® SoC	80 at 100	84 at 90	30 at 175	90 at 145	195 at 175	140 at 145	4,000 at 800
16-/32-bit instruction set support	16 and 32	16, 32, and 48	32	32	32	32	32, 16-bit Thumb/ Thumb2
Level 1 instruction cache	-	-	-	Configurable	Configurable	Configurable	32 KB
Level 1 data cache	-	-	-	-	Configurable	64 KB	32 KB
Level 2 cache	-	-	-	-	-	-	512 KB
Memory management unit	-	-	-	-	Yes	Yes	Yes
Floating-point unit	-	-	-	Floating-point custom instruction	Floating-point custom instruction	-	Double precision
Vector interrupt controller	Yes	-	-	Yes	Yes	-	-
Tightly coupled memory	Up to 64K	-	-	Configurable	Configurable	-	-
Custom instruction interface	-	-	Up to 256	Up to 256	Up to 256	-	-
Equivalent LE	2,500	6,800	600	1,200	1,800 – 3,200	5,500	HPS

because it has nearly full functionality for the supplied devices and is completely suitable for learning purposes.

Depending on the type of the CPU core there are some more stages in the design technology (Fig. 1).

When using a software CPU core it can be entered into the project in several ways:

- When there is *an available HDL model* (designed or ready) it is added to the rest part of the project using a text editor.
- When *a firmware core* (such as Nios II) is used together with the hardware design the CPU core is configured by the means of special software – for instance *Nios II Eclipse Platform* for Nios II.

There is a particular module in Quartus II for generating SOPCs - **SOPC Builder**, which makes easy embedding and configuring a processor core in the design [5]. It is a powerful design tool allowing defining and generating a whole system in a single chip (SOPC) for quite shorter time in comparison to the traditional design approaches. SOPC Builder generates many files defining the system hardware, memory map, simulation model, additional data, etc.

SOPC Builder can be used for creating systems including Nios II processor, some other processor or without any processor.

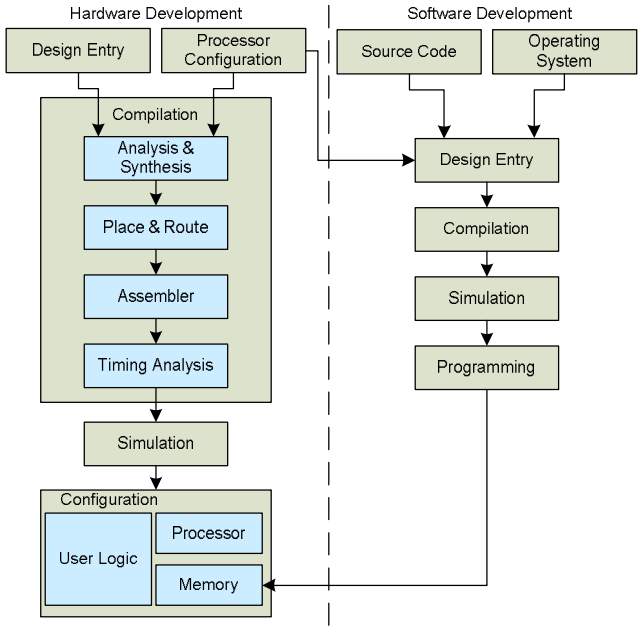


Fig. 1. Design flow for SOPC with FPGA

SOPC Builder includes a component library with processor cores, peripherals (timers, etc.), serial interfaces, and various controllers (off-chip memory, etc.). It is possible to use components (IP cores) of Altera's IP partners and also user components.

CPU core configuring means to define capabilities and parameters such as instruction set (the architecture of the processor), data width, address space, required peripherals, and their features and so on. In general the differences in the variants reduce to the following [4]: *inclusion or exclusion of a feature* – for example hardware multiplier; *more or less of a feature* – for example the volume of the instruction cache memory; *hardware implementation or software emulation* – for example interrupt controller.

Together with the hardware design it is necessary to **design the software** for the CPU core. Further more – in contrast to the traditional microprocessors, when a unique CPU cores will be used it is necessary to make development tools for them – compilers, simulators, in-circuit debuggers, etc.

Altera supports the University education by a special program including discounts for development tools, special software for creating, simulation, loading and real time test of the software for Nios II – *Altera Monitor Program*, many tutorials, example applications, on-line courses, etc.

Downloading the program/ data memory

Downloading the binary file (and data, if any) is possible in several ways depending on the type of the memory used – if on-chip memory is used its initialization is a part of the FPGA configuring. However it is not a large amount and commonly off-chip memory is used. In that case the code and data downloading is made using a special interface and the on-chip memory can be used for a bootloader.

IV. Design of Embedded Microprocessor System with Minimum Features for Educational Purposes

SOPC Builder and the development board DE2-70 of Terasic Technologies Inc. are used for the designed SOPC system. DE2-70 includes FPGA Cyclone II various peripherals, LEDs, switches, LCD display, various types of off-chip memory, IrDA transceiver, 10-bit video-DAC, 10/100 Ethernet PHY/MAC, etc. Therefore DE2-70 has enough resources to be used in the engineering education at an initial learning stage without necessity of additional peripheral modules and devices.

An example system with Nios II implemented in DE2-70 is shown in Fig. 2. The processor Nios II and the necessary interfaces for connection with the other components on the development board are implemented in FPGA Cyclone II. These blocks are interfaced by interconnections and logic, called *Avalon Switch Fabric*. The on-chip memory in FPGA Cyclone could be used as a program memory for short programs for Nios II processor. The access to the SRAM, SSRAM, SDRAM and Flash memories on DE2-70 can be implemented by the means of appropriate interfaces. The parallel and serial interfaces allow implementing the desired input-output ports. A special JTAG UART interface is used to connect by USB to a personal computer (PC). Together with appropriate software it is called USB-Blaster. Another block, called JTAG debug module, allows the PC to control the Nios II system.

It is possible to download program memory in that way, to test and debug the software by various operations - running and stopping the execution, using breakpoints, collecting trace

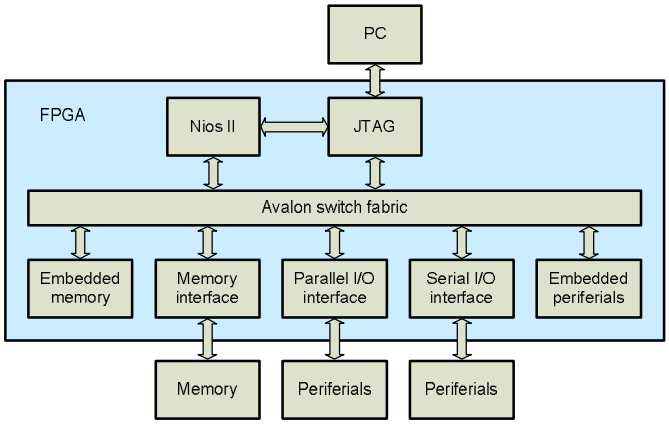


Fig. 2. SOPC with Nios II and DE2-70

data, etc.

All the components in the Nios II systems are modeled in hardware description languages (HDLs). That is why it is necessary for the designer to know some HDL (which needs time and efforts) or to use SOPC Builder for its implementation, by simply choosing and configuring the components needed according to the application requirements.

The design flow in brief is the following:

- **Creating project in Quartus II Web Edition**

The project creation is a stage, which the students had been acquired in advance. At this stage the name of top-level file and entity, the FPGA family and device, and other data are entered.

- **Configuring and generation of the Nios II system in SOPC Builder**

SOPC Builder library components are used in our case: processor Nios II/e, on-chip memory, two parallel interfaces – input and output and JTAG interface for connection with the PC to configure the FPGA and design debugging. (Fig. 3)

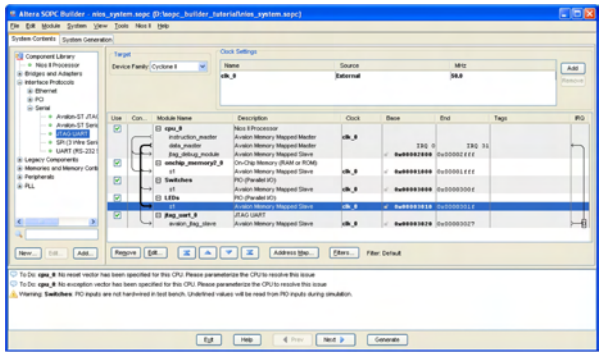


Fig. 3. Nios II system in the SOPC Builder

- **Including the SOPC system into a Quartus II project**

SOPC Builder generates corresponding .vhd files for the Quartus II project. They have to be integrated into the top-level .vhd file created using the text editor in Quirtus II.

- **Assigning pins, compilation the design and configuring the FPGA**
- **CPU core software design and testing the whole system**

The software for the CPU can be written either in Assembler, or in a high level language. Assembler is used in our case to test the functionality of the designed system. Its operation is simple as the aim at this stage of learning is not to study the language – the state of eight switches (SW7-SW0) on DE2-70 is read and it is output on eight LEDs (LEDG7-LED0). *Altera Monitor Program* is used to develop the software, to download it into the memory assigned to the processor and to run the program (Fig. 4 and Fig. 5).

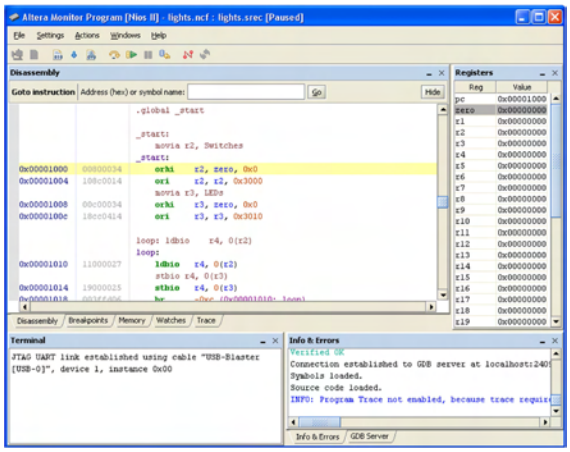


Fig. 4. Nios II software in Altera Monitor Program

V. CONCLUSION

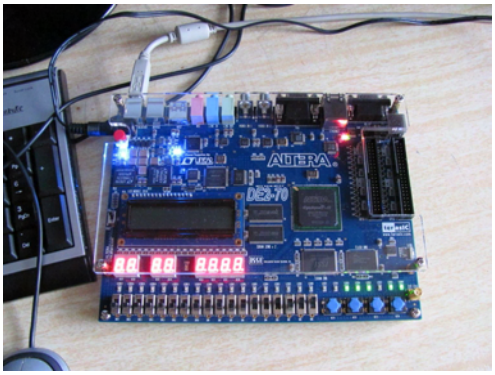


Fig. 5. Testing the Nios II system with the application software

An approach to learning the initial steps in designing FPGA-based SOPCs/ Embedded systems is presented in the paper. These steps are an obligatory stage in studying the complex software development environments, prior to expand the features and architecture of the embedded systems.

ACKNOWLEDGEMENT

The present work is supported by the Science Research Fund at the Ministry of Education, Youth and Research.

REFERENCES

[1] V. Rankovska, „Microprocessor cores for complex programmable logic arrays", Unitech'11, Conference Proceedings, vol. I, pp. I-186 – I-191, Nov. 2011. (in Bulgarian)

[2] H. Karailiev, V. Rankovska, „Synthesizing Sine Wave Signals Based on Direct Digital Synthesis Using Field Programmable Gate Arrays", ICEST 2007, Conference Proceedings, pp. 637-640, Ohrid, Macedonia, 2007.

[3] J. Hamblin, T. Hall, "Using System-on-a-Programmable-Chip Technology to Design Embedded Systems", 33rd International Symposium on Computer Architecture, vol. 13, no. 3, pp. 2395-2503, Sept. 2006.

[4] Nios II Processor Reference Handbook, Altera Corp., May, 2011.

[5] Quartus II Handbook Version 9.1 Altera Corp., May, 2009.

Electronic Simulator of Sound (Noise) Effects for Electric Vehicles in Urban Areas

Georgi Pavlov[1], Galina Cherneva[2], Radoslav Katsov[3], Ivaylo Nenov[4], Ilko Tarpov[5]

Abstract - **The report proposes a technical solution based on a microprocessor electronic system for reproducing real noise effects with the silent movement of electric vehicles in the urban areas. By its mounting on an appropriate place in the electric vehicles and parameterization depending on the specific conditions, problems, related to their disability (silence), will be solved.**

Keywords – **Electric Vehicles, Electronic Simulator of Sound**

I. INTRODUCTION

Electric drives powered by a rechargeable battery were used in mobile equipment as early as in the late 19th century. What followed was a rapid progress of the so-called electric vehicle that even before the beginning of the 20th century overcame the dreamt limit of 100 km/h. However, the rapid rise was followed by a failure related to its relatively high price and limited mileage. Vehicles powered by internal combustion engine won the battle with electric vehicles thanks to their great advantage, virtually unlimited mileage.

For the recent years the use of electric vehicles has come ahead in searching for a new generation of power sources and vehicles. The reasons for their appearance on the market are related to the crisis and the rising price of liquid fuels, global environmental problems, global warming and all other consequences related to it. This has united efforts to reducing car emissions in cities. [1]

In general, all car manufacturers have specified the production of electric cars in their programs that can be classified into the following groups: electric vehicles designed to move only in cities; specially designed electric vehicles and electric cars based on the previously existing classic cars; sports EVs.

From the surveys accomplished, it has been established that now mainly EVs of the first group are available on the market: mini cars or the so-called urban electric vehicles. The EVs of other two groups are under development and experimental studies. The problem is what the cost of technology will be and what their consumer ratings will be from an economical point of view. If sophisticated environmental technologies are used, they may be too costly in terms of energy saving and efficiency and it could be a bad solution. Such an expensive solution can be used for sports cars but it could not be a solution for electric vehicles of mass production. [2].

It means that the use of expensive technology will be inappropriate at this stage and probably in future. Perhaps this is the main reason now for some manufacturers to solve the problem based on classic cars.

Recently the **NHTSA (National Highway Traffic Safety Administration)** in the USA has issued a statement about the process of drafting a law from 2011, **Pedestrian Safety Enhancement Act**. The Act requires **NHTSA** to investigate whether quiet cars are more dangerous, whether to introduce a requirement for electric cars to imitate some noise and formulate a regulation if such a noise is necessary. Others reject the idea of adding artificial warning to electric cars. [3]

In many countries it is considered that electric cars are more dangerous because they are quieter. The considerations are that people face to a greater risk because they can not rely on their hearing to detect and avoid approaching vehicles.

On the other hand, the problem of noise pollution must be considered. Cities are noisy places and the additional noise causes various health problems. Electric cars offer a solution to this serious problem of noise.

These studies are aimed to establish a standard requiring the equipment of electric and hybrid cars with an appropriate sound generating range for warning pedestrians if necessary by 2013. The idea is to help "visually impaired" and other pedestrians to detect the presence, direction, location and operation of such vehicles.

Based on the mentioned above, this paper further presents a device developed by us, which is based on a microprocessor electronic system for reproducing realistic noise effects with electric vehicles movement in urban areas. The electronic system does not require a large power supply and allows setting parameters depending on specific conditions.

II. MAIN FUNCTIONS AND FEATURES OF A SIMULATOR FOR GENERATING SOUND EFFECTS OF ELECTRIC VEHICLES

The electronic device (Fig. 1) used to generate appropriate sound, noise effects intended for a remote detection of moving

[1]assoc. prof. dr. Georgi Pavlov - Todor Kableshkov University of Transport, Bulgaria, Sofia, 158 Geo Milev Str.

[2]assoc. prof. dr. Galina Cherneva - Todor Kableshkov University of Transport, Bulgaria, Sofia, 158 Geo Milev Str.

[3]Radoslav Katzov – student at the Todor Kableshkov University of Transport, Bulgaria, Sofia, 158 Geo Milev Str.

[4]Ivaylo Nenov – student at the Todor Kableshkov University of Transport, Bulgaria, Sofia, 158 Geo Milev Str.

[5]Ilko Tarpov – PhD at the Todor Kableshkov University of Transport, Bulgaria, Sofia, 158 Geo Milev Str.

electric units is a complex system composed of electronic blocks with specific function and operating by a predefined algorithm.

The hardware consists of two separate duplicate schemes that operate in a particular specified line. The main function of the first scheme is to simulate the sound of the engine when accelerating speed and the second circuit simulates the sound with speed decreasing.

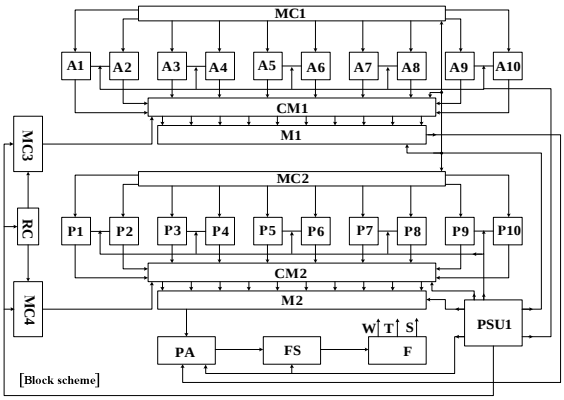


Fig.1. Block scheme of an electronic simulator of sound (noise) effects for electric vehicles in urban areas

Description of main blocks in the scheme structure:

MC1 - microcontroller of ATMEGA 8 (48) series. It generates electrical impulses that are pre-programmed in a way to input beat to memories in a specified time interval. This in turn leads to a cyclic repetition of the simulated sound.

Blocks **(A1-A10)** and **(P1-P10)** are electronic devices (memories), which mainly store and generate the simulation.

CM1 –controlled module of a significant application. This module is filled with a different set of electronic circuits that allow a smooth transition from one to another state of simulation.

M – mixer, which serves to unite ten channels into two stereo channels.

RC – resistive circuit, which changes the resistance of microcontrollers **MC3, MC4** of ATMEGA 826 series. Their function is to regulate the voltage of the controlled module.

PA – preamplifier. Its function is to partially amplify the signal and then transmit it to the final stage **FS** where the signal is increased to optimal levels, then filtered by filter **F.**

PSU is a power supply unit, which ensures the normal mode of system operation.

III. Algorithm of System Operation

The sound simulated by hardware is recorded by an external source in advance. The record is installed on a computer and is converted from analog into digital with the help of appropriate software. The transformed signal is made discrete into a series of pulses with duration of one second. The signal treated in this way is recorded from the computer onto separate memories connected to the controlling microcontroller. It is programmed in a way to repeat the integrated signal continuously in the individual memories thus forming a cycle.

The outputs of the individual memories are connected to a control module. The main function of this module is to receive the signal from memories and ensure a smooth transition from one signal to another. The control of this module is performed by another microcontroller that, with changing its input resistance within 0 and 47 kΩ, regulates simultaneously both the strength of the signal from the control module and the smooth transition of the signal from one state to another. As a result simulation is accelerated. The signal that comes out from the controlled module is a multi-channel one. Therefore each signal is connected to a mixer, which reduces the number of outputs from ten to two.

The second part of the scheme operates by an opposite principle. The microcontroller potentiometer or the input controller impedance varies within the same range. The difference in the operation algorithm of duplicate schemes is that, with reducing the resistance of the first microcontroller, the memory switch with the help of the control device is done by successively merging the signals of memories from A1 to A10 and, with an increase of the resistance, the signal of memories from A1 to A10 is ignored. Thus, with increasing the input resistance, the microcontroller starts gradually switching signals gearing down from P1 to P10.

The resulting sound simulation passes through a number of filters and finally the signal is enhanced multiple times through the preamplifier and output stage.

IV. Conclusion

The limited use of electric vehicles in the country at this stage is only temporary and as slowly as things move, this type of cars sooner or later will start to enter our market. What we can do until then is to trace all possible trends, the solutions of various problems related to electric vehicle technology and the development of regulatory regimes for electric mobility across countries.

The main function of the designed electronic simulator upgraded in an electric car is to create the necessary noise effects identifying it as a vehicle. It will create conditions for safe operation in urban areas in terms of pedestrians and especially of the blind and people with low-vision.

The advantages of the proposed electronic system are related to high reliability, low cost, lower power consumption and ability to adapt to specific requirements and operation modes of various electric vehicles.

References

1. http://www.bg-history.info/882/
2. http://cars.actualno.com/-actualno.com
3. http://ecobusiness.dnevnik.bg/show/?storyid=530001

Spray Deposition of PVDF Layers with Application in MEMS Pressure Sensors

Georgi Kolev[1], Mariya Aleksandrova[1] and Krassimir Denishev[1]

Abstract – **Layers of piezoelectrical polymer polyviniliden fluoride (PVDF) were prepared by spray deposition technique on silicon and glass substrates. The temperature of substrates and concentration of the solution were investigated and their influence on the adhesion to the surface and the film's morphology were explored. The pre- and post temperature treatment of the sample was conducted and changes in the layer's behaviour was detected.**

Keywords **– PVDF, Pressure sensors, MEMS, Piezoelectric effect.**

I. Introduction

With the microelectronic progress it has been observed decreasing of topology's dimensions of the integrated circuits (respectively the entire electronic device) and integration of microsensors and actuators together with the processing electronics modules [4]. Thus, it could be produced completed and individual as functions complex systems. From another side the researches in the microelectronics field are connected with using of new materials, having properties and parameters beyond of already known, which are standard used until this moment.

Nowadays, many scientists work in the area of new material synthesis. The developed new organic polymers expand their application in many sensors and actuators [2]. Recently data about poly(vinyliden fluoride) (PVDF) with piezo- and pyroelectrical properties appear in the literature [3]. The wide application of PVDF as piezoelectric layers in pressure and force sensors and actuators is conditioned from its ability for low cost and simply methods of deposition onto standard substrates. The PVDF possess forward and reverse piezoelectric effect, which make it very suitable material in MEMS structures. It is resistant to many organic solvent and acids, which is big advantages in technologically aspect.

The layers deposited by spray technology are characterized with higher uniformity in comparison with the other popular techniques for solution deposition (for example spin-coating, dip coating and casting) [7]. There are no restrictions about the coverage quality over any size of the covered substrates, because of its operational principle based on fine aerosol stream with controllable spot. The thickness of the produced layers also can be easy controlled and can vary in wide range – from several nanometers to several hundreds of nanometers.

In this paper we aim to investigate the behavior of PVDF material dissolved in MEK solvent in different ratios and to deposit the solution by spray method at different substrate's temperatures. We expect different adhesion of the layers as well as different surface morphology. Our purpose is to find the optimal deposition, pre- and post deposition conditions for obtaining of highly adhesive and uniform films on any type of surfaces used typically in the microsystems (silicon, glass).

[1]Georgi Kolev, Mariya Aleksandrova and Krassimir Denishev are with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: georgi_kolev1@abv.bg

II. Experimental Section

Poly(vinyliden fluoride) (PVDF) is semicrystalline polymer with typical melting point about 160-170 ^{o}C. Because it contains only a single type of repeat unit, in pure form it is homopolymer. When crystalline and amorphous PVDF phase is combined, layers with properties like high plasticity and chemical resistance to acids and bases are created. The thickness of the layers is variable from x.100 nm to 10μm, mono- and bi-orientated, where the first has isotropic piezoelectric properties, and the second one - anisotropic properties (more information could be found in [6]).

TABLE I
Parameters of the PVDF material

Symbol	Parameter	PVDF	Copolymer	Units
t	Thickness	9, 28, 52, 110	<1 to 1200	μm (micron, 10^{-6})
d_{31}	Piezo Strain Constant	23	11	10^{-12} $\frac{m/m}{V/m}$ or $\frac{C/m^2}{N/m^2}$
d_{33}		-33	-38	
g_{31}	Piezo Stress constant	216	162	10^{-3} $\frac{V/m}{N/m^2}$ or $\frac{m/m}{C/m^2}$
g_{33}		-330	-542	
k_{31}	Electromechanical Coupling Factor	12%	20%	
k_t		14%	25-29%	
C	Capacitance	380 for 28μm	68 for 100μm	pF/cm^2 @ 1KHz
Y	Young's Modulus	2-4	3-5	10^9 N/m^2
V_0	Speed of Sound stretch:	1.5	2.3	10^3 m/s
	thickness:	2.2	2.4	
p	Pyroelectric Coefficient	30	40	10^{-6} C/m^2 ^{o}K
ε	Permittivity	106-113	65-75	10^{-12} F/m
$\varepsilon/\varepsilon_0$	Relative Permittivity	12-13	7-8	
ρ_m	Mass Density	1.78	1.82	10^3kg/m
ρ_e	Volume Resistivity	$>10^{13}$	$>10^{14}$	Ohm meters
$R_{\square}$	Surface Metallization Resistivity	<3.0	<3.0	Ohms/square for NiAl
$R_{\square}$		0.1	0.1	Ohms/square for Ag Ink
tan δ_ε	Loss Tangent	0.02	0.015	@ 1KHz
	Yield Strength	45-55	20-30	10^6 N/m^2 (stretch axis)

For deposition of PVDF layers by spray method, laboratory stand was designed (Fig.1), which consists of adjustable flat heater with power 350 W, placed on dialectical ceramic substrate. The set temperature is controlled by electronic regulator with pulse width modulation (PWM) principle. The temperature range which can be set is 30-120 ^{o}C and the accuracy in this range is 2 ^{o}C.

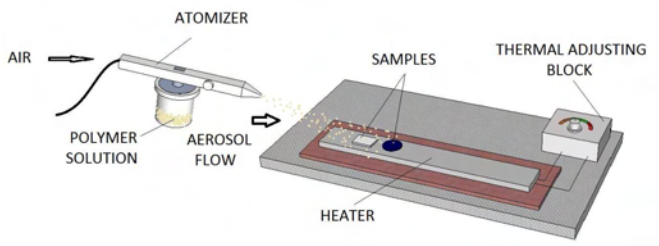


Fig. 1 Experimental setup for spray deposition of PVDF thin films.

According to previously reported experiments this accuracy is high enough to guarantee repeatable parameters of layers, because it is already know that temperature difference more than 10 ^{o}C causes difference in morphology. The deposition is conducted by atomizer with possibility to regulate the diameter of the nozzle, which enables formation of aerosol flow with different diameters of the droplets. The working air pressure is 3.8 bars. For obtaining of PVDF layers 5 mm diameter grains were used, purchased from Goodfellow. The used solvent for crystallization of PVDF in β-phase is metil-etil-cetone (MEK).

As the deposition principle is connected with generation of aerosol flow, the uniformity is determined by the density of this flow and the angle under which it falls on the substrate's surface. The kinetic energy of the pieces is defined by the pressure of the transporting (carrying) air. This pressure can be in certain range, whose optimal value depends on the distance nozzle-substrate. The final layer quality is related with the substrate temperature and concentration of the solution. The dependence between initial aerosol diameter and final diameter of the pieces, forming the layer on substrate is given by Eq. 1:

$$\left(\frac{d^*}{d}\right) = 0{,}32T^{0{,}11}P^{0{,}003}C^{0{,}31} \quad (1)$$

By experimental investigation of the optimal pressure, it has to be taken into account that the temperature of evaporation of the solvent (for MEK 70 ^{o}C). If the pressure is lower than the optimal, the pieces don't possess the necessary kinetic energy to reach the substrate. Their energy is taken away from the pieces of the surrounding environment. When the energy is higher than the optimal, respectively the pressure is also higher and the aerosols penetrate inside the already deposited monolayers, worsening the layer uniformity and planarity.

III. Results and Discussion

Before deposition the substrates were preliminary cleaned in standard detergent solution consists of hydrogen peroxide, ammonia and distilled water. That prepared substrates were sprayed at room temperature of 20^{o}C with cycles of 3 times and 6 times. The deposited layers were observed under microscope(Fig.2). They are characterized with uniformity and density of the structure. The layers show strong adhesion to the substrates.

Fig. 2 PVDF layers sprayed at 20°C on glass substrate.

Fig. 3 PVDF layers after thermal treatment at 180°C.

After thermal treatment of the samples at 180^{o}C for 30s are not seen vastly change in the morphology, but adhesive properties are worsen. Because of the production of thin films the numbers of mono layers were increased at 10 spraying cycles. In this way denser and more uniform layers were produced(Fig.3). To decrease the monolayers splitting, heating of the substrates to 180^{o}C was necessary (around the material melting point).

During pulverization, the substrates are gradually cooled with the time, because of the solvent evaporation. In this way, however, the adhesion of the monolayers to each other is decreased. The possible reason for this behavior is the straying of the temperature away from the melting point of PVDF(Fig.4). After the adhesion test, the upper monolayers, which are deposited at the lower temperatures, drop off.

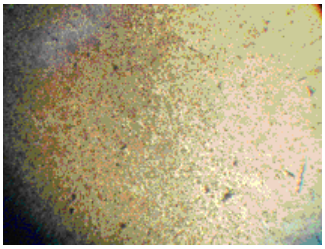


Fig. 4 Adhesion test of PVDF layers

For this reason it is necessary to maintain and compensate the substrate temperature all the time during the deposition process for every spraying cycle. For comparison on the figures below are shown layers deposited at different temperatures in the range 70-100 ^{o}C for experimental determination of the most suitable heating temperatures for the substrates. At increased solution concentration (Fig.5),

layers pulverized at 80 °C are smooth and uniform, but the adhesion is considerably worsened.

Fig. 5 PVDF layers sprayed at 70 °C on glass substrate (in left) and silicon substrate (in right), having lower adhesive strength.

Fig. 6 PVDF layers sprayed at 80 °C on glass substrate (in left) and silicon substrate (in right), having higher adhesive strength.

Fig. 7 PVDF layers sprayed at 100 °C on glass substrate, showing good adhesion, but low planarity and uniformity.

Fig. 8 PVDF layers sprayed at 80 °C and increased concentration on glass substrate, showing good adhesion, but low planarity and uniformity.

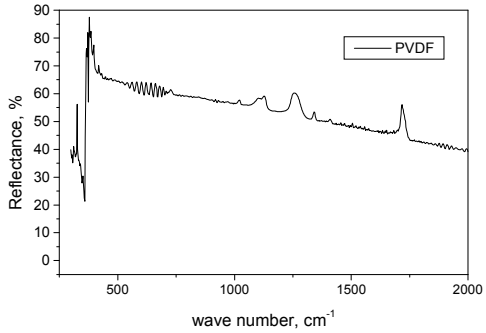


Fig.9 FTIR spectrum of PVDF deposited at optimal spraying conditions.

For chemical identification of the obtained material's phase after deposition and for ensuring that the polymer is not with damaged chemical bonds, a Fourier Transform Infrared analysis (FTIR) was performed (Fig.9). The samples are treated in reflectance mode, because the polymer is not transparent for the infrared wavelength. According to the measured spectrum and the peak positions it can be conferment that the temperature treatment doesn't influence on the changing of the polymer's physico-chemical properties [1], but only on the layer's morphology.

IV. Conclusion

From the conducted experiments, it was established the optimal spraying temperature, at which PVDF layers for potential sensing applications are successfully fabricated. The observed adhesive behavior to different surfaces (the most used in microsystem technologies – glass and silicon) depends on the substrate temperature and the solution concentration. It was found that 80 °C is the most suitable temperature. When the temperature gradient occurs, the adhesive strength is worsened. The future work will be connected with investigation of the piezosensitivity of pressure sensing structures, consisting of pulverized layers. It will be determined the connection between the deposition conditions and the sensor properties of the structures.

Acknowledgement

The authors are thankful to Prof. Kostadinka Gesheva and PhD Student Georgy Bodurov (Central Laboratory of Solar Energy and New Energy Sources – Bulgarian Academy of Science) for the measured FTIR spectra.

References

[1] S. Lanceros-Me´ndez, J. F. Mano,A. M. Costa, and V. H. Schmidt, "FTIR AND DSC STUDIES OF MECHANICALLY DEFORMED β-PVDF FILMS", Sci.—Physics, B40(3&4), 517–527 (2001)

[2] H. Guckel, Surface micromachined pressure transducers, Sensors and Actuators A: Physical, Vol. 28, Issue 2, pp. 133-146, 1991.

[3] R. P Vijayakumar, D. V. Khakhar, A. Misra, STUDIES ON A TO B PHASE TRANSFORMATIONS IN MECHANICALLY DEFORMED PVDF FILMS, Journal of Applied Polymer Science, Vol. 117, Issue 6, pp. 3491–3497, 2010.
[4] H. Guckel, SURFACE MICROMACHINED PRESSURE TRANSDUCERS, SENSORS AND ACTUATORS A: Physical, Vol. 28, Issue 2, pp. 133-146, 1991.
[5] Georgi D. Kolev, Mariya P. Aleksandrova, Krassimir H. Denishev, "INFLUENCE OF THE DEPOSITION CONDITIONS ON THE MORPHOLOGY OF PVDF LAYERS WITH POSSIBLE APPLICATION IN MEMS PRESSURE SENSORS", International Journal of Advances in Engineering, Science and Technology (IJAEST), Vol. 1,pp193-198, No. 2 Dec 2011-Feb 2012
[6] http://www.msiusa.com"PIEZO FILM SENSORS"
[7] S. Ju Kang, Y. Jung Park, J. Sung, P. Sung Jo, Ch. Park, K. Jin Kim, and B. Ok Cho, SPIN CAST FERROELECTRIC BETA POLY(VINYLIDENE FLUORIDE) THIN FILMS VIA RAPID THERMAL ANNEALING, Appl. Phys. Lett. Vol. 92, 012921, 2008.

Different Technological Methods for Offset Compensation in Si Hall Effect Sensors

Ivelina Cholakova[1]

Abstract – **This paper presents technological methods for offset compensation of symmetrical Hall sensors. The offset compensation is one of the most important tasks to solve for accurate measurements by such type of magnetic sensors. The presented methods could be a good solution because the main conductive region is deeply "buried" in the epitaxial layer. In this way it will be away from the surface and well isolated from the substrate modulation.**

Keywords – **Magnetic sensors, Hall effect sensors, offset compensation, CMOS 0.18 µm technology.**

I. Introduction

A Hall element is used for contactless measurement, for example as linear and angular positions, electrical current and power, etc. The Hall element, fabricated by means of CMOS technology gives a weak output signal (of the order of few millivolts). This signal is corrupted of offset and noise [1]. The offset is the output signal of the modulating type of sensor (in most cases with a differential output) in the absebse of an external magnetic field (B = 0). Essentially, this kind of error is a static value or a very slow variation with time of voltage, current or frequency. If there is no additional information concerning the magnetic field or the sensor itself, the offset cannot be distinguished from the useful output signal. If the offset is time invariant, it causes a parallel shift of the whole calibration curve [2]. The offset occurrence is due to several external and internal reasons, as most of them are related mainly to the fabrication process as tolerance of device geometry (for example misalignment of Hall terminals), non-uniform distribution of doping impurities, uneven thickness, crystal damage and dislocations, mechanical stress and strain. Great numbers of methods are utilized in order to reduce or compensate the offset: applying improved fabrication technologies or particular sensor constructions; additional technological treatment of each device; formation of additional control electrodes in the active sensor area, various offset compensation circuits, sensitivity-variation offset reduction method, etc [3].

II. Methods for Offset Reduction of a Hall Effect Device

In many applications magnetic sensors are used, e.g. for contactless measurements. Silicon Hall sensors are the prime candidates for such applications due to their cost-effective integration potential. But they have a stress voltage offset which cannot be easily compensated in the production line, so that the offset reduction methods have to be applied [4].

There is a spinning current method which averages the results of several consecutive Hall measurements with different orientations in the crystal plane. Very low residual offsets in the microtesla range can be achieved but the circuitry required may introduce problems [4].

The so called Anti-Hall (AH) method for Hall plates described in [5] also utilizes different orientations in the current injection. The currents are injected at different points but at the same time on the outer and inner boundaries of the sample. Besides the magnetic field sensitive signal, a separate stress sensitive signal can be measured simultaneously at another pair of voltage contacts.

Of course, there are many schematic solutions for offset reduction. The simplest method of offset adjustment is shown in Fig. 1, and uses a manual potentiometer to null out the offset of the Hall effect transducer. The potentiometer is used to set a voltage either positive or negative with respect to the output sense terminal, and a high value resistor sets an offset current into or out of the transducer. It is possible to null out either positive or negative offsets with this scheme.

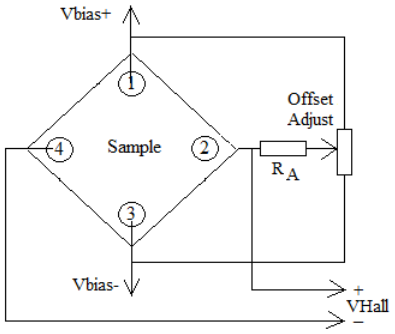


Fig. 1. Manual offset adjustment using potentiometer

A feature of this method is that it can be used regardless of whether the transducer is biased with constant current or a constant voltage source [6].

More complicated method, for example is the dynamic compensation of the offset of an integrated Hall sensor with an instrumentation amplifier. As a result of the interchange of the bias and Hall contacts, Hall voltage remains with the same value and sign, while the offset retains its magnitude, but its sign is reversed. By using the sample & hold technique the signals from each of the two consecutive commutations are added in the signal processing circuit which in essential is a instrumentation amplifier. This approach enables building

[1]Ivelina Cholakova is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: inch@ecad.tu-sofia.bg.

electronic systems for accurate contactless measurements by magnetosensitive device [3].

Another solution, for example, is proposed by Melexis in their Programmable Linear Hall Effect Sensor. CMOS Programmable, Ratiometric Linear Hall Effect sensor IC. The linear output voltage is proportional to the magnetic flux density. The ratiometric output voltage is proportional to the supply voltage. The MLX90251 possesses active error correction circuitry, which virtually eliminates the offset errors normally associated with analog Hall Effect devices. Integrated on the MLX90251 is a temperature-compensated quad switched Hall plate, chopper stabilized amplifiers, adjustable output filter, output driver, voltage protection circuitry and a programmable EEPROM with security and redundancy. Programming the EEPROM allows each device to be calibrated in the application [7].

An alternative to all mentioned methods above, are the technological methods for offset compensation, some of them will be described in the publication with experimental results.

The offset originating from the mask misalignment can be minimized by designing an appropriate layout of the sensor. Also the sensor should be symmetrical for better results. The sensor geometry could also be optimized [1]. One of the proposed methods is the cell to be rotated to 45 degrees in order to orient it along the crystallographic direction (110). It is experimentally proved that in this direction the piezo-resistive coefficient takes the minimum for an N-well Hall device.

A 40um Hall plate was implemented as a basic cell. The layers used in the design are: 3.3V N-well, n-implantation for the contacts, diffusion layer, p-implantation, which is a shallow p+ layer which increases the average resistance and decreases the thickness of the Hall plate. It also acts as an electrostatic shield. The main purpose is to avoid the surface malformations which lead to offset; there is also a metal layer on the top of the Hall sensor which acts as a shield layer. The basic cell is illustrated on Fig. 2.

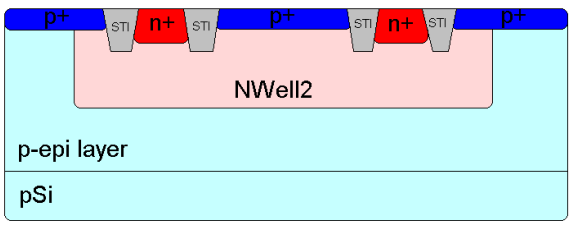


Fig. 2. Hall plate

III. Experimental results

A test chip was designed with Hall sensors on the XFAB XH018 process. The purpose of this test chip is to test different Hall sensors on the XFAB XH018 technology and to find the optimum Hall sensors for future projects on this technology.

The sensors were tested in order the residual offset to be established. A four-phase spinning method is used with the aim the offset to be cancelled. The test equipment is shown in Fig. 3.

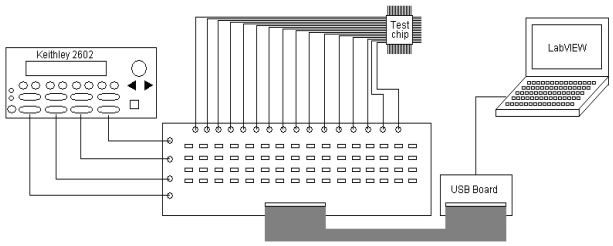


Fig. 3. Test equipment

The equipment is made in order four Hall plates to be tested at once. A LabVIEW program was created for more automated test process. For chip measurements a Keithley 2602 was used which is duo channel source meter with 10,000 readings/s and 5.500 source-measure points/s to memory. One of the channels is used as a source to supply the test chip and the other channel is used to measure the output signal. The plates are tested in six different supply voltages (from 0.5V to 3.00V with step of 0.5V). We also use a Switch Matrix Board with 64 relays which provide the 4-phase spinning method, which in our case is used to cancel the offset (Fig. 4).

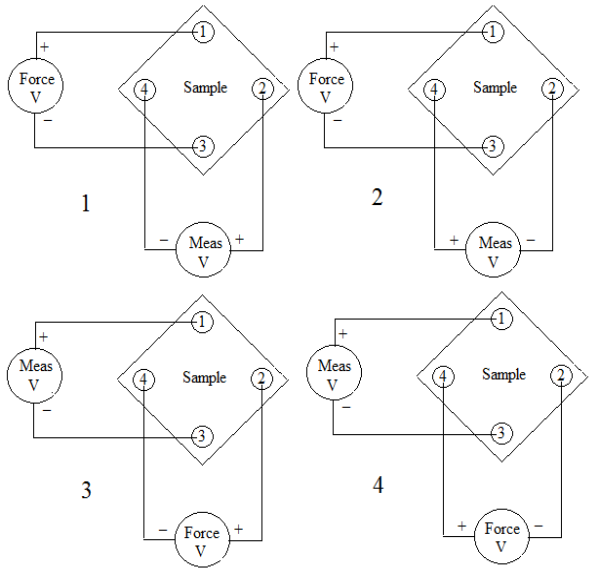


Fig. 4. 4-phase spinning method

The experimental results are shown at Table 1, where the residual offset is in μV and the supply voltage is at Volts. The structure which is examined is designed with shallow trench isolation layer (STI) instead of active area and is shown in Fig. 5.

TABLE I
RESIDUAL OFFSET

Supply voltage, V	Residual Offset, uV	Residual Offset, uV	Residual Offset, uV
0.5	1.8675	1.881	1.147
1	0.7185	0.553	0.054
1.5	0.8275	1.0125	0.7125
2	0.501	1.2865	0.693
2.5	1.9515	1.46	0.926
3	1.9255	1.9745	1.494

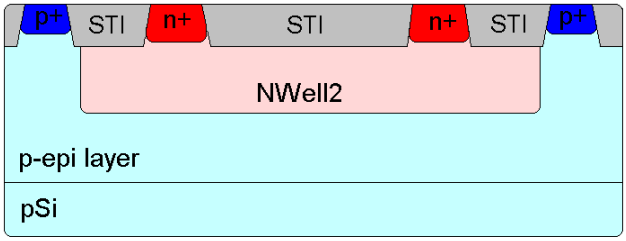


Fig. 5. Hall plate with STI layer

The results present that the worst residual offset is at 0.5V and it is ~ 0.9% of the output signal.

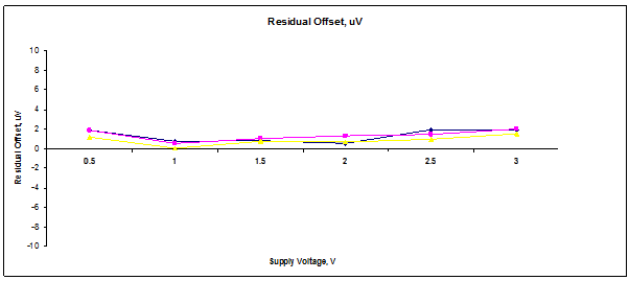


Fig. 4 illustrates three measurements at the same Hall plates under certain conditions. This result is in accordance with the maximum limits for the residual offset and the sensor is appropriate for the applications it is designed, because a high accuracy is required in the automotive industry.

IV. CONCLUSION

In this paper a technological solutions of symmetrical 0.18 CMOS Hall effect sensors for effectively offset compensation are proposed and explained. These solutions effectively improve the stability of the Hall output voltage, because the residual offset, which is inevitable, is optimally reduced.

ACKNOWLEDGEMENT

The research in this paper was carried out within the framework of Contract No. 122ПД0060-03.

REFERENCES

[1] Zoran B. Randjelovic, Maher Kayal, Radevoje Popovic and Hubert Blanchard, *Highly Sensitive Hall Magnetic Sensor Microsystem in CMOS TechnologyHighly Sensitive Hall Magnetic Sensor Microsystem in CMOS Technolog*, IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 37, NO. 2, pp. 151-159, 2002.

[2] C. S. Roumenin, *Solid State Magnetic Sensor*, HANDBOOK OF SENSORS AND ACTUATORS, Volume 2, Elsevier, 1994.

[3] Kamen Fillyov, Tihomir Takov, Tzvetelina Tzeneva, Slaveiko Neytchev, Slavka Tzanova, Dynamic Compensation of the Offset of an Integrated Hall Sensor with an Instrumentation Amplifier, 6th IMAPS International Symposium for Design and Technology of Electronic Modules, SIITME 2000, Bucharest, 21 -24 September, 2000, pp. 119-123

[4] C. Muller-Schwanneke, F. Jost, K. Marx, S. Lindenkreuz, K. von Klitzing, *Offset reduction in silicon Hall sensors*, Sensors and Actuators 81 (2000), Elsevier, pp. 18-22.

[5] R.G. Mani, K. von Klitzing, F. Jost, K. Marx, S. Lindenkreuz, H.P. Trah, Method for simultaneously reducing the misalignment offset and separating the Hall voltage from the off-diagonal piezoresistive voltage in Hall effect and piezoresistive devices based on silicon, Appl. Phys. Lett. 67 1994 2223.

[6] Ed Ramsden, *Hall Effect Sensors Theory Application*, Advanstar Communications, Cleveland, Ohio, 2001.

[7] Melexis Microelectronic Integrated Circuits, Datasheet, Programmable Linear Hall Effect Sensor MLX90251.

Multipoint Video Control System Applicable in Assistance of Elderly and People with Disabilities

Ivo Iliev[1], Serafim Tabakov[2], Velislava Spasova[3]

Abstract – **Ambient Assisted Living (AAL) is a new research area which promises to address the needs of the elderly and people with disabilities. AAL has a lot of barriers and challenges that impede its development and more universal deployment. These could be summarized as: technological, ethical, legal and psychological challenges. An approach which allows overcoming part of these barriers, mostly related to the integration of a video control system in the assistive process, is presented by the authors in this paper.**

***Keywords* – AAL, Fall detection, Assistance for the elderly, ICEST 2012.**

I. INTRODUCTION

An on-going tendency in developed countries is the ever increasing percentage of the elderly population, a process also known as gentrification. The situation in Europe is very serious – according to Eurostat 2008 in 2015 the demographic growth in the EU will become negative and by 2060 up to one third of the EU population will be aged 65 or above [1]. Similar demographic processes are also evident in other developed countries such the US, Canada, Japan, etc. These tendencies would put additional stress to the social and pension systems in the affected countries as the ratio of working to non-working people will drop. Also, there wouldn't be sufficient resources in the caring sector as the demand for personal assistants, nursing homes, hospices, rehabilitation centers and hospitals will greatly surpass the availability.

This is the reason why in recent years a lot of efforts have been concentrated in the search of alternative caring solutions. Ambient Assisted Living (AAL) is a relatively new area of research whose objective is to provide a solution for independent living of the elderly and people with disabilities. AAL could be defined as the distributed collection of intelligent systems (either implanted, body-worn, or embedded in the surrounding environment) that assist and monitor the elderly in their daily lives, thus assuring a better quality of life [2]. Under quality of life we understand the physical, mental, emotional and social well-being of a person.

One of the main objectives of AAL systems is to provide personal safety. Falling is a major risk faced by older people with one third of the people over 65 falling each year. Currently fall detection is a well researched topic and sensor based solutions are available and commercialized. The majority of such solutions are based on accelerometers, but one of the drawbacks of these systems is that users always need to wear the sensor. The system will not work if the user forgets to wear the device.

Recently a number of researchers have been looking at developing solutions based on video cameras and computer vision approaches. A monitoring system based on video cameras has potential advantages. The video signal is semantically rich, thus a single camera in a room could pick up most of the activities performed in the room, and consequently could replace a large number of sensors.

Video and image processing could be applied in different areas and for different purposes in an AAL system [3]:

- fall detection – cameras could be used in order to extend or replace other more traditional fall detection algorithms based on accelerometers, motion sensors, smart floors, etc.;
- activity monitoring – cameras could be used in simple actions tracking or more complicated daily life activities monitoring or classification;
- security – cameras are used in solutions that ensure that the home is protected, most often this means video surveillance;
- safety – cameras could be used in solutions related to personal safety, for example tracking people with dementia or other mental illnesses;
- telemonitoring of physiological parameters such as pulse.

The implementation of computer vision in AAL systems is different and depends on the aim of assistance (in relation with the assisted person condition) and the type of used cameras. Simple 2D digital cameras are cheap and easy to find but often a single 2D camera isn't sufficient for the purposes of AAL and multiple cameras have to be used on the price of processing algorithms complication [5]. 3D cameras are more expensive but provide more information about the position of the user's body in space and one 3D camera could be sufficient for an AAL application [6]. Omnicameras provide 360° view of the environment and could be used in more specific scenarios such as location tracking [3]. Stereocameras use two lenses in order to obtain stereo video signal which could be beneficial, especially for fall detection [7], although the processing algorithms are more complicated. There are also several applications of wearable cameras such as Microsoft's SenseCam in AAL – they are used in home risk zones detection or detection of significant events [8]. Finally, there are a lot of multimodal applications – i.e. applications in which cameras are used in conjunction with other sensors to obtain better precision of the selected algorithms.

Video or image capturing systems and their applications in AAL are a relatively recent research field and there are still a lot of challenges to be overcome [3].

[1]I. Iliev 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: izi@tu-sofia.bg

[2]S. Tabakov 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: tabakovsd@gmail.com

[3]V. Spasova 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: vgs@tu-plovdiv.bg

Video and image processing on its own is not a new research area but AAL is, and as a consequence there are specific issues related to the application of computer vision in AAL which haven't been fully resolved: the choice of video or image processing algorithms; the problem with occlusions; data fusion from multiple cameras; multi occupancy of the environment; presence of a pet such as a cat or dog, etc;

This integration challenge is related to the integration of the video processing subsystem with the other components of an AAL system. Problematic issues are: the interfaces to the other parts of the system; the choice of where to process the sensitive information – locally or centrally; ensuring good protection of the sensitive information.

A major and very serious problem in AAL systems is the acceptance by the users – the AAL system would be useless if the user refuses to accept and use it. Factors that might get in the way of acceptance are fear, reluctance or inability to use technology, or unclear evidence of the real benefits of the AAL system. Video monitoring systems present an additional acceptance challenge - the users are often very concerned that they will loose their privacy and as a consequence are reluctant to accept camera-based solutions in their homes or rooms. There are various approaches to handle this problem: to process locally the video or images (i.e. to use smart cameras) and if possible to forward only alarms to other parts of the system which are accessible by careers; to use silhouettes and edges instead of full pictures [4]; and most importantly to educate the prospective users, to explain that their privacy wouldn't be sacrificed.

An approach allowing overcoming a part of the acceptance barriers is presented by the authors in [9], where special attention is paid to the privacy. The problem was solved by interactive (voice) communication between the assisted person and the system, immediately before the beginning of video registration. The proposed two step approach for reliable fall detection was further improved and adapted in such a manner as to be able to embed it in multipoint video assistive system.

II. Method and Devices for Video Control

The paper presents the multipoint video control component of an AAL system which detects critical situations such as fall of the assisted person. The architecture of the AAL system is presented at Fig. 1. The main part of the system is a small computer (Mini PC), used as a gateway, on which a specially developed control application is running. All of the body-worn and environmental sensors, as well as the supplementary devices, are connected to the gateway via a standardized (Bluetooth, ZigBee, WiFi) or proprietary interfaces. If an emergency situation is recognized, and there is no consecutive person's reaction in a predetermined period, an alert is sent to the authorized person (or relative) via a GSM module embedded in the Mini PC.

The multipoint video control system is introduced as a means of improving the efficiency of the system. Sometimes an alarm would be triggered when there is no critical situation – e.g. the assisted person has slipped but has not fallen or has fallen but is not injured, and without the video monitoring component the AAL system will generate a false alarm.

Fig. 1. Architecture of the AAL system

The idea behind the introduction of video monitoring component is to verify that a generated alarm is indeed valid and whether further action should be initiated in order to provide help to the assisted person. Thus, the purpose of the video component is to increase the reliability of the AAL system and to increase the efficient use of the available monitoring and human caring resources.

The video monitoring is organized in a network of IP cameras situated at different places in the assisted person's home, in a manner allowing coverage of the whole area. The used cameras are model HS-691B-M186I (manufactured by EasyN), and have features appropriate for the concrete task.

It can be seen form the table that the cameras are equipped with audio input/output, which gives possibility to establish voice contact with the assisted person, for example immediately before starting the video monitoring.

The program window of the software application running on the gateway is presented on Fig. 2. As can be seen, the operator has the possibility to enter the architectural plan of the assisted person's home and to choose appropriate locations for the cameras in order to ensure coverage of the whole home area. The camera icons are fixed on the e-plan during system installation.

Fig. 2. Software toolbox for system configuration

When a potentially critical state is recognized by the system and there is no consecutive adequate reaction by the assisted

person, the AAL system sends a SMS (Short Message Service) message to a previously specified cell phone number or numbers.

If the contact person has access to a telecommunication network (such as WiFi, 3G, LAN), he or she could establish connection to the Mini PC and could remotely start video monitoring mode. This way the contact person would be able to assess whether the situation is indeed critical, could accept or reject the alarm, and could initiate any further action if it is required.

In order to minimize the traffic between the contact person and the Mini PC instead of video monitoring mode, the system could generate and send one or several MMS (Multimedia Messaging Service) messages showing the state of the assisted person and the reason for the alarm triggering. Of course, specific image processing and analysis algorithms are needed in order to recognize critical situations and to choose the suitable images to send. These algorithms are currently under research and development.

All triggered events as well as the corresponding reactions are stored in a local Mini PC database and could be accessed at any time.

III. Conclusion

The paper presents a multipoint video control system applicable in AAL systems for elderly and people with disabilities. The presented approach permits flexible realization of such a system, in accordance with the concrete needs and requirements of the assisted person, the architectural specifics of his/her home, the furniture, etc. The developed software application for control of IP cameras network allows the realization of a comprehensive AAL system not only in a single home, but also in specialized nursing homes, hospices, rehabilitation centers, etc. Additional plus of the system is the possibility for integration of additional network architectures, e.g. a network of detectors to monitor environmental parameters, controllers to turn on and off electrical devices, to measure their power consumption, security and alarm systems, etc.

Acknowledgement

This study is supported by the Bulgarian National Scientific and Research Fund, project ДДВУ02/18.

References

[1] K. Giannakouris, Aging characterizes the demographic perspectives of the European societies, Eurostat 72, 2008.

[2] G. Broek, F. Cavallo, and C. Wehrmann, AALIANCE Ambient Assisted Living Roadmap, IOS Press BV, Netherlands, ISBN 978-1-60750-498-6B, 2010.

[3] K. Arning, and M. Ziefle, “Different Perspectives on Technology Acceptance: The Role of Technology Type and Age”, Proceedings of USAB 2009 -5th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, Austria, pp.20-41, 2009.

[4] A.M. Tabar, A. Keshavarz, and H. Aghajan, “Smart Home Care Network using Sensor Fusion and Distributed Vision-based Reasoning”, Proceedings of VSSN'06 - The 4th ACM international workshop on Video surveillance and sensor networks, USA, pp.145-154, 2006.

[5] M. Grassi, et al., “A Multisensor System for High Reliability People Fall Detection in Home Environment”, Sensors and Microsystems, vol. 54, Part 4, pp.391-394, 2010.

[6] F. Cardinaux, D. Bhowmik, C. Abhayaratne, and M.S. Hawley, “Video based technology for ambient assisted living: A review of the literature”, Journal of Ambient Intelligence and Smart Environments , vol. 3, pp.253-269, 2011.

[7] R. Cucchiara, A. Prati, and R. Vezzani, “A Multi-Camera Vision System for Fall Detection and Alarm Generation”, Expert Systems, vol. 24, Issue 5, pp.334-345, 2007.

[8] A. Williams, D. Ganesan, and A. Hanson, “Aging in Place: Fall Detection and Localization in a Distributed Smart Camera Network”, Proceedings of the MULTIMEDIA'07 – The 15th international conference on Multimedia, Germany 2007, pp.892-901

[9] I. Iliev, S. Tabakov, and I. Dotsinsky, “Two steps approach for falls detection in the elderly”, Annual Journal of Electronics, vol. 5, No 2, 2011, pp. 46-48

Computer Modeling of RF MEMS Inductors Using SPICE

Elissaveta Gadjeva

Abstract – **Computer models are developed for RF MEMS inductors. They are implemented in the general purpose circuit simulator Cadence PSpice. Parameterized inductor macromodels are developed taking into account the frequency dependence of the series resistance due to the skin-effect. An automated procedure for parameter extraction of MEMS inductor model is developed. The procedure is realized in the environment of Cadence Capture and Cadence PSpice, using macrodefinitions in the graphical analyzer Probe.**

Keywords **– RF MEMS inductors, PSpice simulator, Computer models.**

I. INTRODUCTION

With the recent development of wireless communications, there is a growing demand for integrated RF circuits with high performance and low cost. Various micromachining technologies, implemented in microelectromechanical systems (MEMS), have been applied to RF applications. Micro inductors are important passive components applied in LC tank, VCO and DC-DC converters [1,2]. The MEMS technology allows to enhance the *Q*-factor of micro inductors, to increase the selfresonant frequency and to decrease the energy dissipation in inductors [1-5].

In the present paper, computer models are developed for RF MEMS inductors. They are implemented in the general purpose circuit simulator *Cadence PSpice* [6]. An automated procedure for parameter extraction of MEMS inductor model is developed. The procedure is realized in the environment of *Cadence Capture* and *Cadence PSpice*, using macrodefinitions in the graphical analyzer *Probe*.

II. RF MEMS INDUCTOR MODELS

The Π – RF physical inductor model shown in Fig. 1 [7,8] describes the performance of a MEMS inductor, where L_s is the inductance, C_s is the capacitance between the windings of the inductor, C_1 is the capacitance in the oxide (or polyamide) layer between the coil and the silicon substrate, C_p is the capacitance between the coil and the ground through the silicon substrate, and R_p is the eddy current losses in the substrate [1]. The series resistance R_s is assumed constant up to frequency f_o and then increases as $\sqrt{f}$ to model the skin-effect (Fig. 2), where f is in GHz:

$$R_s(f) = \begin{cases} R_{so} & \text{for } f < f_o \\ A\sqrt{f} & \text{for } f \geq f_o \end{cases} \quad (1)$$

A simplified variant of the model shown in Fig. 1 is presented in Fig. 3.

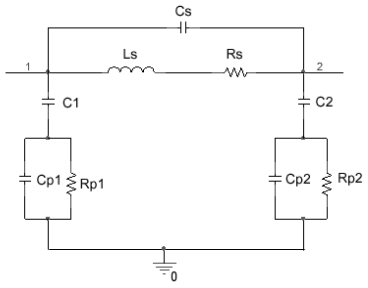


Fig. 1. Π model of MEMS inductor

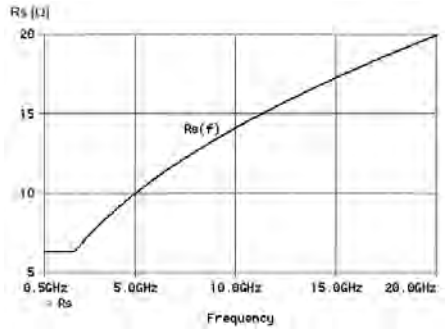


Fig. 2. Frequency dependence of the series resistance R_s

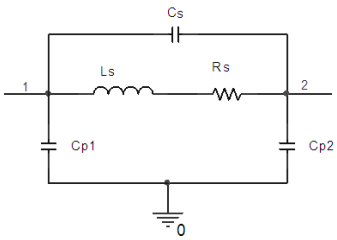


Fig. 3. Simplified Π model of MEMS inductor

The high performance solenoid-type MEMS inductor is modeled by the equivalent circuit shown in Fig. 3 [3], where the frequency dependence of R_s due to the skin-effect is represented by the dependence:

$$R_s(f) = A\sqrt{f} \ . \quad (2)$$

[1]Elissaveta Gadjeva is with the Faculty of Electronic Engineering and Technologies, Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail:egadjeva@tu-sofia.bg.

III. COMPUTER RELIZATION OF MEMS INDUCTOR MODELS

The *PSpice* model corresponding to the equivalent circuit shown in Fig. 3 is presented in Fig. 4.

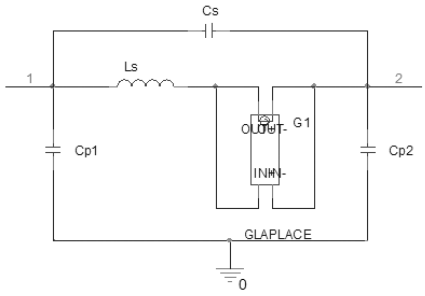


Fig. 4. *PSpice* model corresponding to the equivalent circuit in Fig. 3

The frequency dependence of R_s is defined by the voltage controlled current source (VCCS) G1 of GLAPLACE type using the expression (1). It is realized in the form:

$$R_s(f) = KR_{so} + (1-K)A\sqrt{f}\ , \tag{3}$$

where

$$K = \begin{cases} 1 \text{ for } f < f_o \\ 0 \text{ for } f \geq f_o \end{cases} . \tag{4}$$

As the frequency f is not directly accessible for the user in the input language of the *PSpice* simulator, the LAPLACE variable s= j ω is used for obtaining the frequency in (3). The coefficient K is calculated in the parameterized macromodel, defined using block, according to the input language of the PSpice simulator in the form:

0.5*(sgn(@Fo-M(s)/6.283165)+1)

where M(s) is used to obtain the angular frequency ω.

The admittance $1/R_s$ calculated using (3) and (4), is defined in the property XFORM of the VCCS as follows:

{@Rso*0.5*(sgn(@Fo-M(s)/6.283165)+1)+
0.5*(sgn(M(s)/6.283165-@Fo)+1)*@Rso/sqrt(1E-9*@Fo)*
sqrt(1E-9*M(s)/6.283165)}

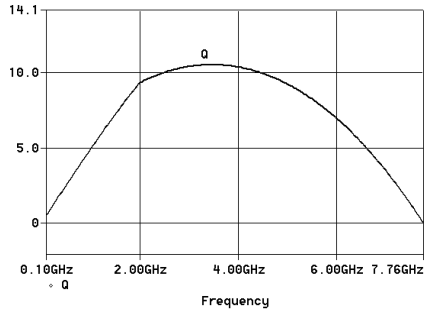


Fig. 5. Simulated Q factor

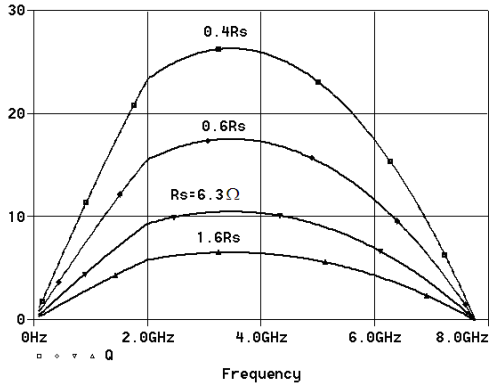


Fig. 6. Dependence of Q on R_s

The simulated Q factor of the inductor with L_s=5 nH, $C_{p1} = C_{p2}$ = 75 fF, C_s = 9 fF, R_{so} = 6.3 Ω and fo = 2 GHz is presented in Fig. 5. Using parametric sweep, the simulation results for the dependence of Q on Rs are shown in Fig. 6.

Using *File/Append* option of the graphical analyzer *Probe*, the results for the dependence of Q on C_s and C_p variation are presented together on the same plot as shown in Fig. 7. The curve 1 corresponds to $0.2C_p$ and $0.4C_s$, curve 2 – to $0.4C_p$ and $0.6C_s$, curve 3 – to $0.6C_p$ and $0.8C_s$, curve 4 – to C_p and C_s, curve 5 – to $1.6C_p$ and $1.4C_s$. The simulated results are in agreement with the results given in [1].

The *PSpice* realization of the solenoid-type RF MEMS inductor model shown in Fig. 3 [3] is presented in Fig. 8.

The frequency dependence of R_s is defined by equation (2), where A = 1.503. The admittance $1/R_s$ is defined in the property XFORM of VCCS:

{1/(1.503*sqrt(1E-9*M(s)/6.283185))}

The simulation results for the quality factor Q and inductance L are presented in Fig. 9. They are in a good agreement with the measured results given in [3].

The calculated impedance Zeq of the air suspended RF MEMS inductor represented by the model shown in Fig. 1 [4] is shown in Fig. 10. The model parameters are: Cs=1.14 fF, L_s=6.76 nH, R_1= 1 Ω, C_1=11.6 fF, C_2=90.5 fF, C_{p1}=1 fF , C_{p2}=10.2 fF, R_{p1}=275 Ω, R_{p2}=332 Ω.

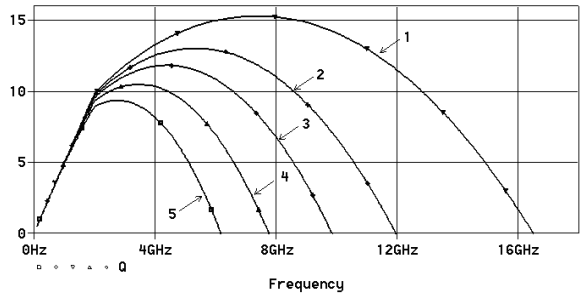


Fig. 7. Dependence of Q on C_s and C_p variation

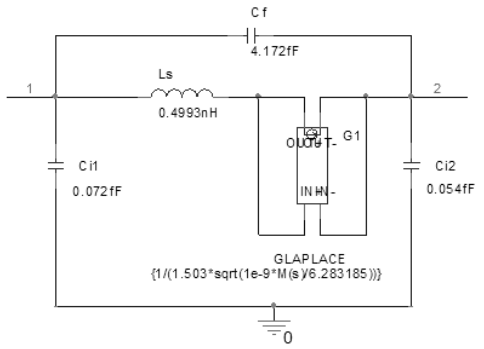


Fig. 8. PSpice realization of the solenoid-type RF MEMS inductor model

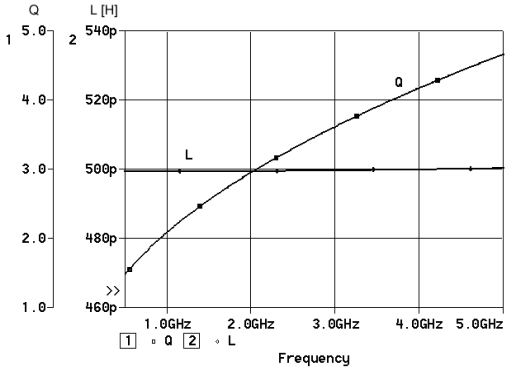


Fig. 9. Simulation results for the quality factor *Q* and inductance *L*

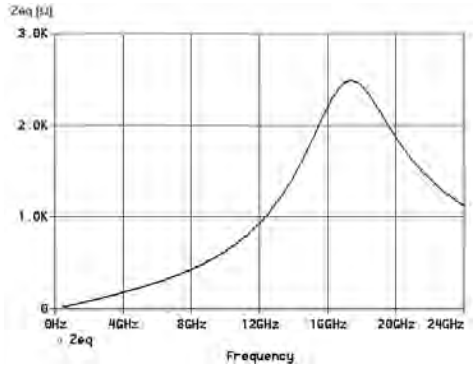


Fig. 10. The impedance Z_{eq} of the air suspended RF MEMS inductor

IV. Parameter Extraction of RF MEMS Model

An automated procedure for parameter extraction of MEMS inductor model is developed based on two-port measured *S*-parameters. The procedure is realized in the environment of *Cadence Capture* and *PSpice*, using macrodefinitions in the graphical analyzer *Probe*.

The equivalent circuit of a fully compatible, highly suspended spiral inductor is presented in Fig. 11 [5].

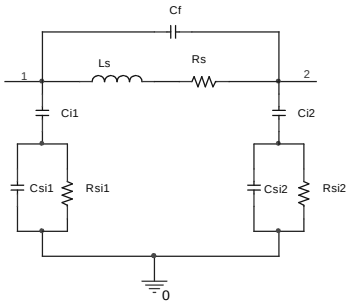


Fig. 11. Equivalent circuit of a fully compatible, highly suspended spiral inductor

The two-port measured S-parameters are used as input data. They are defined using voltage controlled current sources VCCS of GFREQ type (Fig. 12). The table is in the form:

<frequency,magnitude [DB], phase [DEG]>:
(freq1,mag1,phase1), (freq2,mag2,phase2) …

The measured S-parameters are obtained in the form of corresponding node voltages (Fig. 12). They are defined using the following macrodefinitions in the graphical analyzer *Probe*:

S11=V(S_11) S21=V(S_21)
S12=V(S_12) S22=V(S_22)

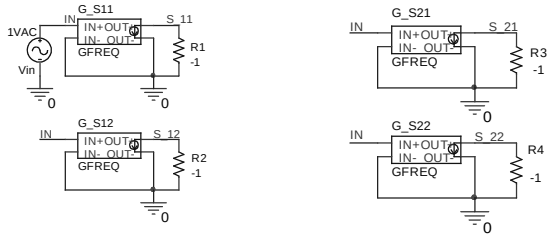


Fig. 12. Introducing the measured S-parameters in PSpice

The extraction procedure is realized in the following steps:

1. Conversion of two-port *S*-parameters S_{ij} in two-port Y-parameters

The measured two-port *S*-parameters S_{ij} are converted to the two-port *Y*-parameters Y_{ij}, i,j = 1,2, needed for the parameter extraction procedure.

$$Y_{11} = \frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{\Delta}, \quad (5)$$

$$Y_{22} = \frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{\Delta}, \quad (6)$$

$$Y_{12} = \frac{-2S_{12}}{\Delta} \quad ; \quad Y_{21} = \frac{-2S_{21}}{\Delta}, \quad (7)$$

where

$$\Delta = ((1+S_{11})(1+S_{22})-S_{12}S_{21})R_o \ .$$

R_o =50 Ω - the characteristic impedance.

2. Determination of L_s

L_s is obtained from the parameter Y_{12} at low frequency f_l. :

$$L_{sf}(f)=\frac{\left|\mathrm{Im}\left[1/Y_{12}\right]\right|}{\omega}\ ;\ L_s=\min\left(L_{sf}(f)\right) \qquad (8)$$

3. Determination of the coefficient *A*

The coefficient *A* is calculated from (2) using the resistanve value $R_s(f_o)$ for the frequency f_o

$$R_{sf}(f)=\left|\mathrm{Re}\left[1/Y_{12}\right]\right|\ ;\ R_s=R_{sf}(f_o)\ ;\ A=R_s/\sqrt{f_o} \qquad (9)$$

4. Determination of C_{i1}, C_{i2} at low frequency f_l.

$$C_{i1}=-\frac{1}{\omega\mathrm{Im}\left[1/\left(Y_{11}+Y_{12}\right)\right]},\ C_{i2}=-\frac{1}{\omega\mathrm{Im}\left[1/\left(Y_{22}+Y_{12}\right)\right]} \qquad (10)$$

5. Determination of C_{si1} and C_{si2} at high frequency f_h

$$C_{si1}=\frac{1}{\omega}\frac{1}{\frac{1}{Y_{11}+Y_{12}}-\frac{1}{\mathrm{j}\omega C_{i1}}},\ C_{si2}=\frac{1}{\omega}\frac{1}{\frac{1}{Y_{22}+Y_{12}}-\frac{1}{\mathrm{j}\omega C_{i2}}} \qquad (11)$$

6. Determination of R_{si1} and R_{si2} at high frequency f_h

$$R_{si1}=\mathrm{Re}\left[\frac{1}{Y_{11}+Y_{12}}-\frac{1}{\mathrm{j}\omega C_{i1}}\right]^{-1},\ R_{si2}=\mathrm{Re}\left[\frac{1}{Y_{22}+Y_{12}}-\frac{1}{\mathrm{j}\omega C_{i2}}\right]^{-1} \qquad (12)$$

7. Determination of C_f at high frequency f_h

$$Y_f=-Y_{12}-\frac{1}{R_s(f)+\mathrm{j}\omega L_s};\quad C_f=\frac{1}{\omega}\mathrm{Im}\left[Y_f\right]. \qquad (13)$$

The extracted parameter values for the MEMS inductor developed in [5] are presented in Table I. The parameter values obtained in [5] are also given. The calculated parameters are in a very good agreement with to the model parameter values and the measured results given in [5].

TABLE I. EXTRACTED PARAMETER VALUES

Model parameter	***Extracted parameter value***	***Parameter value obtained in [5]***
Ls	1.34nH	1.34nH
A	0.27	0.27
*Cs*1	11.6fF	11.6fF
*Cs*2	90.5fF	90.41fF
*Csi*1	1.012fF	1fF
*Csi*2	10.2fF	10.2fF
*Rsi*1	275 Ω	275.01 Ω
*Rsi*2	331.92 Ω	332 Ω
Cf	1.139fF	Cf=1.14fF

V. CONCLUSIONS

Computer *PSpice* models of RF MEMS inductors have been developed. The frequency dependence of the series resistance due to the skin-effect is taken into account. An automated parameter extraction procedure for MEMS inductor model is developed.

ACKNOWLEDGEMENT

The investigations are supported by the project DDVU 02/6/17.12.2010

REFERENCES

[1] Rebeiz, G., "RF MEMS: Theory, Design, and Technology", Published by John Wiley & Sons, Inc., Hoboken, New Jersey 2003.

[2] Tseng, S.-H., Y.-L. Hung, Y.-Z. Juang, M. S.-C. Lu, „A 5.8-GHz VCO with CMOS-compatible MEMS inductors", Sensors and Actuators A 139, 2007, pp.187-193.

[3] Seok, S., Nam, C., Choi, W., Chun, K. "A High Performance Solenoid-type MEMS Inductor", Journal of Semiconductor Technology and Science, vol. 1., n. 3, June 2001.

[4] Merkin, T. B., S. Jung, S. Tjuatja, Y. Joo, D. S. Park, and J. B. Lee,"An Ultra-Wideband Low Noise Amplifier with Air-suspended RF MEMS Inductors," in Ultra-Wideband, The 2006 IEEE 2006 International Conference on, 2006, pp. 459-464.

[5] Yoon, J., Y. Choi, B. Kim, T. Eo nd E. Yoon, " CMOS Compatible Surface-Micromachined Suspended-Spiral Inductors for Multi-GHz Silicon RF Ics", IEEE Electron Device Letters, vol. 23, n. 10, Oct. 2002, pp. 591-593.

[6] PSpice User's Guide, Cadence PCB Systems Division, USA, 2000.

[7] Yue, C. P.; Ryu, C.; Lau, J.; Lee, T. H. & Wong, S. S. "A Physical model for planar spiral inductors on silicon", Proc. IEEE Int. Electron Devices Meeting Tech. Dig., San Francisco, Dec. 1996, pp. 155-158.

[8] Gadjeva, E., V. Durev and M.Hristov, "Matlab - Modelling, Programming and Simulations", Chapter 14: Analysis, model parameter extraction and optimization of planar inductors using MATLAB, Published by SCIYO, Copyright © 2010 Sciyo, - ISBN 978-953-307-125-1, www.sciyo.com, http://www.intechweb.org/

Deposition of Transparent Electrodes for the Future Generation of Flexible Displays

Mariya Aleksandrova[1], Georgy Dobrikov[1], Kostadinka Gesheva[2], Georgy Bodurov[2], Ivelina Cholakova[1] and Georgy Kolev[1]

Abstract – **Indium tin oxide transparent films with thickness 58 nm were deposited by modified method for radiofrequency vacuum sputtering on flexible substrate for application as display electrode. Amorphous and colourless polyethyleneterephtalate foil (APETF), having melting point of 80 °C, was chosen as flexible substrate. The oxygen partial pressure, varying during the sputtering with one order of magnitude (1.10^{-3}-2.10^{-4} Torr), strongly affects the film structure. Another important parameter is the post deposition exposure with UV light. This treatment leads to decreasing of the sheet resistance from 359 to 68,6 Ω/square. The changes in these conditions also result in changes of the transmission for the visible light - the maximal transparency is in the range 71% to 89%. The samples were fully characterized by scanning electron microscopy, Fourier transform infrared spectrophotometer, refractive index, UV-VIS and Van der Paul sheet resistance measurements.**

Keywords – **Flexible displays, Indium-tin oxide films, RF sputtering.**

I. INTRODUCTION

During the last decade the organic materials are intensively investigated because of their potential application in displays, solar cells, sensors and other types of optoelectronic and microelectronic devices [1]. The highest attention attract the electroluminescent light emitting displays, which tent to replace the current developed liquid crystal technology (LCD) as portative flat displays, because of the lower consumption, wider viewing angle and higher own brightness. According to the contemporary initiatives for environment protection by saving paper (and preserve trees), the idea for electronic papers appears – cheep digital displays, based on organic materials, deposited on flexible substrates, bended and put in the pocket. The first step toward the physical realization is the deposition of electrode, which injects the necessary charge carriers in the organic semiconductor to cause light emission and in the same time it is enough transparent for the visible light. The most suitable and studied material is indium-tin oxide (ITO), which is usually prepared by techniques like vacuum electron beam evaporation, vacuum reactive sputtering, spray pyrolysis [2,3]. Unfortunately the temperatures developed during those processes are higher then the melting point of the substrate, so alternative technological mode must be created.

There are several reports in the literature connected with different approaches for ITO deposition on different types of flexible substrates like crosslinked hydroxypropylcellulose [4], polypropyleneadipate [5] and polyethersulfone [6]. More of the cited substrate materials are heat resistant in the temperature range Ts = 110–170°C. In some of the papers, where PET substrate is used [7] there is no data about the adhesion of the obtained ITO films and there is no information about the possible microcracks in the film during folding.

The new moment in our work is the modification of the sputtering mode in such way that we could achieve comparable parameters with those, obtained on glass substrate, without damaging of the flexible one. The efforts are directed to ITO deposition by varying the oxygen concentration in the vacuum chamber at decreased RF power. In contrast to the known values for the parameters of ITO films grown up on PET foil by vacuum deposition types of processes, here we decreased the thickness of the electrode more than twice [8], but we still kept the parameters in the same ranges as at higher thickness, which is a big advantage. The decrease of the thickness is expected to lead to higher resistance and lower mechanical stability of the film. However, we proved that at suitable deposition conditions and post deposition treatment we can keep the parameters at the desired values and thus we guarantee high quality preparation of electrodes for thinner and lighter displays.

II. EXPERIMENTAL DETAILS

Sheets from amorphous polyethyleneterephtalate foil (APETF), having melting point of 80°C, were cut with sizes 2.5 cm x 2.5 cm and were cleaned by chemical and mechanical method. Detergent solution consists of hydrogen peroxide:ammonia:distilled water in ratio 1:1:3 was prepared as addiction of the ultrasonic treatment in ultrasonic cleaner. The substrates were cleaned in it for 90 seconds. ITO films were prepared by a RF reactive sputter system using a target with contents In_2O_3 and SnO_2 at a weight proportion of 95:5 mol%. The sputtering power input to the target was decreased in comparison with our previous developed technology for electrode on glass, where the target voltage was kept at 210 W. Here, the RF power was set to 60 W (target voltage 500 V and plasma current 120 mA) at deposition time of 20 minutes. The base pressure was 8.10^{-6} Torr, the oxygen pressure varied between 1.10^{-3} and 2.10^{-4} Torr and the total pressure of

[1]Mariya Aleksandrova, Georgy Dobrikov, Ivelina Cholakova and Georgy Kolev are with the Faculty of Electronic Engineering (Department of Microelectronics) at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, Corresponding author's e-mail: m_aleksandrova@tu-sofia.bg.

[2]Kostadinka Gesheva and Georgy Bodurov are with the Central Laboratory of Solar Energy and New Energy Sources, Bulgarian Academy of Science, blvd. "Tzarigradsko chaussee", 72, Sofia, Bulgaria.

reactive gas and sputtering inert (argon) gas had constant value of $2.5.10^{-2}$ Torr. In this way the substrate temperature during film growth was lower then the temperature of PET's mechanical deformation. To decrease the specific and the sheet resistance, the samples were treated with ultraviolet light form exposure source with power 250W for 15 min. Segment structuring of the ITO was performed by standard photolithographic process and etching in oxalic acid.

The thickness of ITO film was measured 58 ± 1 nm. The refractive index, extinction coefficient and physical thickness of the films were determined simultaneously from transmittance and reflectance spectra of the sample deposited on transparent substrates (optical glass) and the reflectance spectra of the corresponding films deposited on opaque silicon substrates [9,10]. The spectra were recorded by a high precision Cary 5E spectrophotometer at normal light incidence in the wavelength region $\lambda = 400 – 800$ nm, with an accuracy of 0.1 and 0.5 %, respectively. The previously developed three-step algorithm [11] was used for reliable isolation of physically correct solutions and for high accuracy determination, for instance, $\Delta d=\pm 1$nm. Transmittance spectra were recorded by UV-VIS Specord equipment. More details about the spectrophotometrical measurements could be found in [12]. The specific and sheet resistances of the films were measured before and after UV treatment by using four-point probe FPP 500. For composition identification FTIR spectroscopy analysis was performed by Shimadzu FTIR Spectrophotometer IRPrestige-21in reflection mode. Layers morphology was observed on Scanning Electron Microscopy JEOL - JXA733 with accelerating voltage 25 kV. Picture of the structured ITO layers on the flexible substrate is also shown.

III. Results and Discussion

After series of experiments at different deposition conditions we established that the optimal values are as they are described above. Our starting point was the lowest tension in the system ITO film/flexible substrate and to save the substrate from the plasma heating. We set the supplied RF power and argon sputtering gas concentration as constants and the oxygen reactive gas concentration as parameter. We had to set the proper concentration, so a compromise to be achieved between transparency and conductivity for the ITO film, which will perform double function in the future display – injecting electrode and viewing electrode. On the following Fig. 1 we have shown the transmission spectrum for the two boundary states: the best and the worst case of oxygen concentration. The reference curve of the transmittance for uncovered APETF is also shown on the graph as a comparison. At the low value the transparency doesn't exceed 68%, but for the higher value it achieves 88%. However for this transmittance the initial conductivity of the film was too poor. The minimal sheet resistance before exposure with UV radiation was 359 Ω/square. That's why we had decided to choose intermediate oxygen pressure value of 4.10^{-4} Torr, with average transparency of 76% and sheet resistance of 510 Ω/square, and further to decrease the resistance's value. The reason for the poor conductivity is the decreasing of the RF power, which leads to producing of amorphous film, with undefined structure. The micrograph on Fig. 2 shows the result of the scanning electronic microscope at low power and the working oxygen pressure before and after UV exposure at one and the same magnification of 4300. We had to introduce additional energy to reorganize the particles in the film and made the structure three orders of magnitude higher conducting. This may happen with ultraviolet exposure source, where the heating infrared component is rejected and there is no danger for the flexible substrate's to be damaged. It can be seen the amorphous structure without content of any microcrystals before treatment and beginning of micrograins formations after treatment. The corresponding minimal sheet resistance before exposure was 359 Ω/square, but after 15 minutes of illumination it decreased to 68,6 Ω/square. For our chosen working pressure it was 74,5 Ω/square (see Fig. 3). For comparison in the literature the typical achieved values are 160 Ω/square on (RF power = 30W, $P_{o2} = 3.10^{-3}$ Torr) [7] and even 4.104 Ω/square (RF power =100 W, Po_2 = 1 Torr) [4] on BDI Hydroxypropylcellulose cross linked with 1,4-Diisocyanatobutane substrate. The resistance is average value from measurements conducted in three different points all over the area of the substrates. The corresponding transmittances in the visible range for the cited papers are 80% for both investigations. The order of the film's thickness is also comparable – 30 and 60 nm. From the SEM observations it could be seen that there are no micro-cracks in the amorphous layer while for the UV treated sample the image represents grain boundaries.

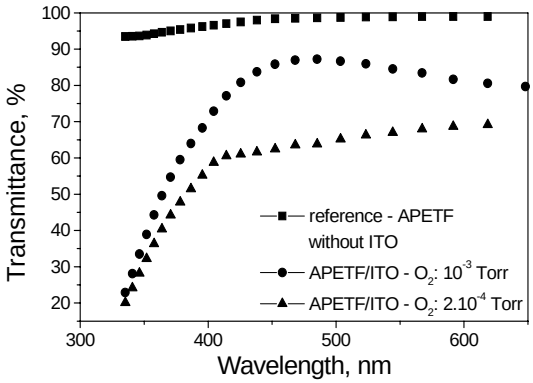


Fig. 1. UV-VIS spectra of systems APETF, APETF/ITO at partial pressure 10^{-3} Torr and 2.10^{-4} Torr.

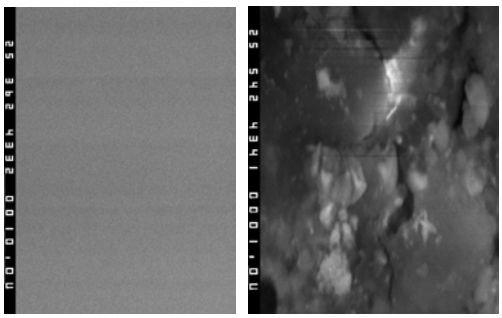


Fig. 2. SEM of untreated with UV (left) and treated with UV (right) ITO surface.

The refractive index is stable and constant in wide range of the wavelengths in the visible spectrum (see Fig. 4). It is lower than the typical measured value [13]. The possible reason is that the light propagation is easier, because the film structure (as well as the flexible substrate's structure) is not defined in certain crystal cells, but it is in form of starting stage of arrangement of non-uniformly distributed nuclei or start of micro-grains formations.

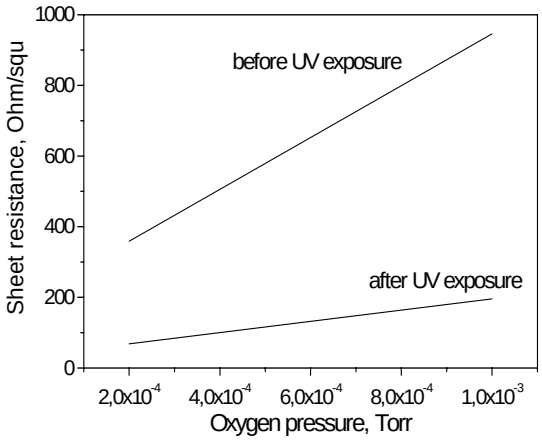


Fig. 3. Sheet resistance of the thin ITO film for different oxygen pressures, including the highest and the lowest possible values.

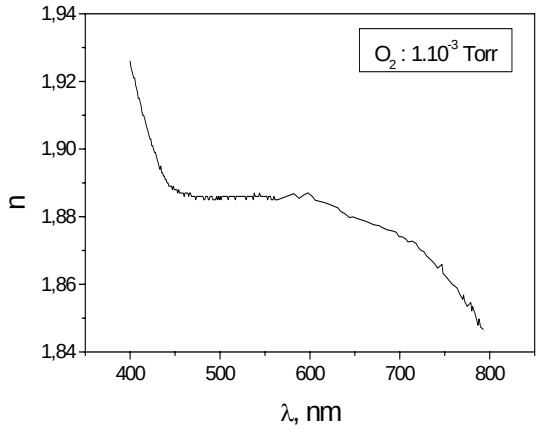


Fig. 4. Refractive index of sample prepared at partial pressure.

Typical FTIR spectra in reflectance mode, acquired in the range 300-4000 cm^{-1} for films deposited on PET at different post-deposition conditions are shown in Fig. 5. The reason for the choice of reflectance mode is that the polymer substrate and the ITO film are not transparent for the infrared (IR) light. After 500 cm^{-1} there are no specific changes in contrast to the values described in the literature, where the characterized peaks are situated between 1000 and 4000 cm^{-1} [14]. It could be seen that the character of the bonds is almost not affected by the UV exposure. The absorption is slightly higher, but the peaks are situated at the same positions, which is prove for preservation of the typical features of ITO film. The peaks situated around 430 cm^{-1} and 560 cm^{-1} confirm the existence of In-O-In bonding, which is in good agreement with the results, reported in [15]. The Sn-O and Sn-O-Sn bonds, which are usually evidenced by peaks situated around cm^{-1} and 616 cm^{-1}, are not observed here.

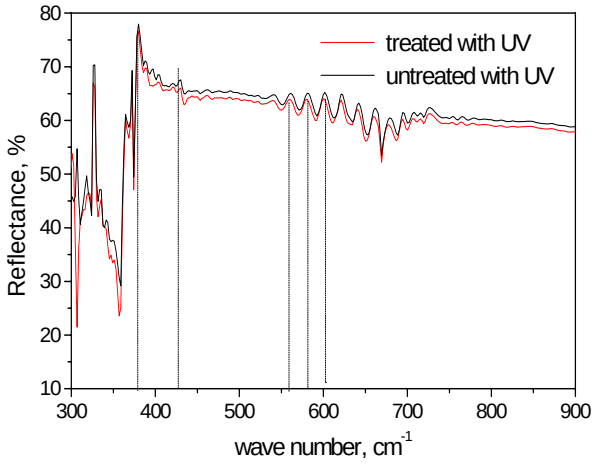


Fig. 5. FTIR spectra of the sputtered ITO layers on flexible APETF substrate before and after UV treatment.

The results show that the behavior of the film can be ascribed to "bending mode", where dipole moment or charge imbalance in the molecule occurs.

According to the measured results the IR reflectance in the range 500 - 2000 cm^{-1} is high (average 43.68 %), which is advantage for another application, different than display – this is so called heat mirror, used as element in the "smart windows".

IV. Conclusion

ITO films were deposited on flexible polymer substrate by using modified method of RF reactive sputtering at different oxygen concentrations. The obtained films have higher than 80 % transmittance, but also resistance higher than 300 Ohm/square. This was the reason to apply UV exposure for self assembly of the ITO particles after deposition. The result was 5 times decreasing of the resistance and almost no changes in the transmittance, as can be seen from UV-VIS spectrum. The UV light doesn't change the molecular structure of the film, which is proved by FTIR measurements. Its influence consists of structure's reorganization in form of microcrystals, shown by SEM observations. After deposition the PET substrates are stable and don't reveal mechanical tensions, which is evident for the ITO film's uniformity with lack of cracks and from the substrate flatness. This is the main advantage of the used modified low power sputtering technique, because of the low substrate temperature at the deposition process. During the conducted measurement another suitable application of the produced ITO films was found, because of the IR reflective properties – the heat mirrors.

ACKNOWLEDGEMENT

The work is financial supported by grant ДМУ 03/5 – 2011 of Fund "Scientific Research", Bulgarian Ministry of Education, Youth and Science. The authors are thankful to the Seriozha Valkanov from the "Institute of Metal Science, equipment, and technologies "Acad. A. Balevski" BAS for the scanning electronic microscopic pictures.

REFERENCES

[1] S.M. Kelly, *Flat Panel Displays: Advance Organic Materials*, Cambridge, Royal Society of Chemistry; 1 edition, 2000.
[2] M. A. Aouaj, R. Diaz, A. Belayachi, F. Rueda, M. Lefdil, "Comparative study of ITO and FTO thin films grown by spray pyrolysis", Materials Research Bulletin, vol. 44, no 7, pp. 1458–1461, 2009.
[3] J. George, C.S Menon, "Electrical and optical properties of electron beam evaporated ITO thin films", Surface and Coatings Technology, vol. 132, no. 1, pp. 45–48, 2000.
[4] A. Luis, C. N. Carvalho, G. Lavareda, A. Amaral, P. Brogueira, M.H. Godinho, "ITO coated flexible transparent substrates for liquid crystal based devices", Vacuum, vol. 64, pp. 475–479, 2002.
[5] Z.W. Yang, S.H. Han, T.L. Yang, Lina Ye, D.H. Zhang, H.L. Ma, C.F. Cheng, "Bias voltage dependence of properties for depositing transparent conducting ITO films on flexible substrate", Thin Solid Films, vol. 366, pp. 4-7, 2000.
[6] L.M. Wang, Y. J. Chen, J. W. Liao, "Characteristics of indium–tin oxide thin films grown on flexible plastic substrates at room temperature", Journal of Physics and Chemistry of Solids, vol. 69, pp. 527–530, 2008.
[7] Y. S. Kim, Y. C. Park, S.G. Ansari, J. Y. Lee, B. S. Lee, H. S. Shin, "Influence of O_2 admixture and sputtering pressure on the properties of ITO thin films deposited on PET substrate using RF reactive magnetron sputtering", Surface and Coatings Technology, vol. 173, pp. 299–308, 2003.
[8] F.L. Wong, M.K. Fung, S.W. Tong, C.S. Lee, S.T. Lee, "Flexible organic light-emitting device based on magnetron sputtered indium-tin-oxide on plastic substrate", Thin Solid Films, vol. 466, pp. 225–230, 2004.
[9] V. Panayotov, I. Konstantinov, "Algebraic determination of thin-film optical constants from photometric (T, Rf, Rm) and (T, Rb, Rm) measurements", Proc. SPIE, vol. 2253, pp. 1070-1079, 1994.
[10] I. Konstantinov, Tz. Babeva, S. Kitova, "Analysis of errors in thin-film optical parameters derived from spectrophotometric measurements at normal light incidence", Appl. Opt., vol. 37, pp. 4260-4267, 1998.
[11] Tz. Babeva, S. Kitova, I. Konstantinov, „Photometric methods of determination of the optical constants and the thickness of thin absorbing films: Criteria for precise and unambiguous determination of n, k and d in a wide spectral range", Appl. Opt. vol. 40, pp. 2682-2686, 2001.
[12] S. Tabakov, V. Yakimov, V. Videkov, "Modernization of Spectrophotometric Measurements", 19-th ISS on Semiconductor and Hybrid Technologies, vol. 19, no. 1, pp. 24-27, Sofia, Bulgaria, 1998.
[13] http://refractiveindex.info/?group=CRYSTALS&material=ITO
[14] S. H. Brewer, S. Franzen, "Optical properties of indium tin oxide and fluorine-doped tin oxide surfaces: correlation of reflectivity, skin depth, and plasmon frequency with conductivity", Journal of Alloys and Compounds, vol. 338, pp. 73–79, 2002.
[15] T.F. Stoica, T.A. Stoica, M. Zaharescua, M. Popescu, F. Sava, N. Popescu-Pogrion, L.Frunza, "Characterization of ITO thin films prepared by spinning deposition starting from a sol-gel process", Journal of Optoelectronics and Advanced Materials, vol. 2, no. 5, pp. 684-688, 2000.

Investigation of Over Voltage Protection Circuit for Low Power Applications

Tihomir Brusev[1], Nikola Serafimov[2] and Boyanka Nikolova[3]

Abstract – **The sophisticated electronic control systems and Integrated Circuits (IC) are widely used nowadays in the modern automotive and industrial applications. The maximum voltages in the modern CMOS technologies go down. The transient and dc voltages which exceed these values can seriously damage the electronic devices. Over voltage protection circuits are needed to prevent from destroying control system and IC for automotive and industrial applications. Investigations results of over voltage protection circuit appropriate for low power applications are presented in this paper.**

Keywords – **Over voltage protection, Surge protection devices, Low power applications.**

I. Introduction

Today the modern semiconductor technologies are widely used often in industrial and automotive applications. Many discrete electronic circuits and Integrated Circuits (IC) are used for measurements, control system, sensors, actuators etc. The maximum voltages in the new CMOS processes became smaller and smaller. The electronic devices for low power applications have to work at various over voltage conditions. The voltages higher than the maximum level for these circuits can lead to serious damaging of the equipment.

The over voltage protection circuits are needed in order to preserve the electronic devices. A potential source of higher input voltage could be for example alternator in the vehicles. These circuits have to prevent the electronics in industrial and automotive applications from damaging. Voltages higher then maximum allowable of discrete circuits and IC have to be stopped.

In some industrial applications surge protection devices (SPDs), such as metal-oxide varistors (MOVs), gas discharge tubes (GDT) and silicon avalanche diodes (SAD) are used. These circuits shunt the input voltage to the ground. They can protect the electronics in the automotive systems when large energy is necessary to be absorbed.

On the other hand the over voltage protection circuits (OVPC) should not degrade the operation of the electronic systems. They should not induce noise in the protected devices. Also voltage drop of the over voltage protection circuits have to be as small as possible. This voltage drop can degrade the performance of the automotive and industrial equipments and especially of system which are used for measurements. Such kind of voltage drops appears in the offset of the measurement system [1].

This paper presents the investigations results of an over voltage protection circuits appropriate for low power applications. In Section II are shown basic types of surge protection devices (SPDs) - as metal-oxide varistors (MOVs), gas discharge tubes (GDT) and silicon avalanche diodes (SAD). In Section III are presented simulation results of over voltage protection circuit. The temperature characteristic of trip voltage is evaluated. The results are achieved by simulations made with Cadence OrCAD PSpice.

II. Metal-Oxide Varistors, Gas Discharge Tubes and Silicon Avalanche Diodes

The transient suppression devices should limit the voltage; limit the current; divert the current; operate fast; be capable of handling the energy; survive the transient; have a negligible affect on the system operation; fail safe; have a minimal cost and size [2]. All of those requirements are desirable and some of them are difficult to be achieved.

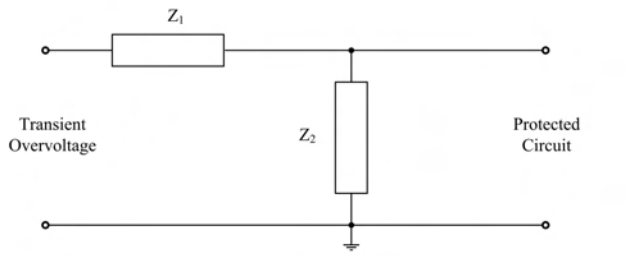


Fig. 1. Transient voltage protection circuit.

The transient voltage protection circuit, shown in Fig. 1, consists both of series and shunt component. For shunt element usually is used breakdown devices that have large impedance during the normal operation of protected circuit. When high transient voltage appears at the input Z_2 has small impedance and transient current is shunt to the ground.

The series element is used to limit the current which is flowing through the shunt. Z_1 will reduce the voltage applied to the protected circuit. This element should exist in the circuits, because the current through the Z_1 will be infinite when transient voltage exceeds the breakdown of the shunt component.

[1]Tihomir Brusev is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: brusev@ecad.tu-sofia.bg.

[2]Nikola Serafimov is with the Faculty of Electrical Engineering, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: nssi@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.

The requirements for shunt components can be fulfilled by electronic devices such as gas discharge tube, silicon avalanche diodes and voltage variable resistor.

A. Gas Discharge Tubes

The gas discharge tube (GDT) is a system which consists of two electrodes fitted inside the tube. The tube is filled with a gas under pressure. The GDT do not limit the bandwidth of high frequency circuits, because they have small shunt capacitance.

At low input voltages gas discharge tubes have high value of impedance. When voltage became bigger than threshold voltage, GDT switches to very low impedance state. The voltage across the GDT is clamping to the threshold level [3].

One of the main disadvantages of GTD is that they conduct slow and conduction threshold voltage depends on the rate of change of transient voltage. They can not be used for low voltage sensitive electronic system, because those voltages are very high – several hundreds of volts. GDT's response time is in microseconds range. They can withstand the current of tens thousand of amperes.

The big problem of the electronic devices which are closed to the GDT is spark developed between the electrodes. These phenomena can seriously damage the system nearby GDT and it's very dangerous in terms of fire. Another disadvantage is that there is current flowing throughout the GDT when the transient over voltage is ended. Therefore the electronic circuits are disconnected from power supply.

B. Silicon Avalanche Diodes

Silicon avalanche diodes are similar to the zener diodes, but the have larger p-n junction area. They are used as signal line suppression and power line transient suppression devices. Silicon avalanche diodes (SAD) clamp the transient overvoltage at a low residual value [4]. The maximum clamping voltage is the voltage that protected circuit should be able to withstand without damage. These devices have to divert the transient current away from the protected circuit. They are fast electronic devices which can respond rapidly to the transient voltage surge.

The disadvantage of SAD is that they can not adsorb large input energy. That's why in most of the applications several SAD are combined together. They can not be used when transient overvoltage occurs frequently. In such cases some of the SAD devices fail.

C. Metal-Oxide Varistors

Metal-Oxide Varistors (MOVs) are voltage clamping devices. They are nonlinear voltage variable resistors which are produced from the mixtures of zinc oxides. The resistance of MOVs decreased when the voltage across the devices exceeds their threshold voltages. MOVs are symmetrical electronic components which can clamp positive and negative voltages.

They are used as surge protection electronic devices which maintain sufficiently low clamping voltage. This is very important for the protected circuits. MOVs can withstand high transient surges. The response time of these components is bigger than the silicon avalanche diodes, but it's smaller than the gas discharge tubes. MOVs can withstand currents in the range of hundreds or thousands of amperes. The multilayer MOVs have sub-nanoseconds response time and they can clamp voltages in the range between 10 V and 50 V.

MOVs are a good choice when circuits have to be protected from an ac power surge. The circuit shown in Fig. 2 is used when the protection requires high level energy range and fast response time [5].

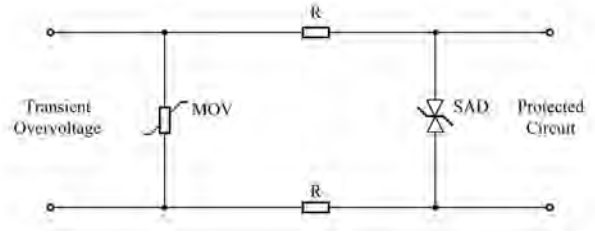


Fig. 2. Transient protection network.

The big problem, when shunt devices are used for over voltage protection of low power applications, is that large amount of energy has to be absorbed [6]. This high power which is dissipated in the surge protection devices, like those described above, can lead to the failure or destruction of low power electronic circuits used in industrial and automotive applications. Therefore they are not suitable for this purpose.

In the section below are presented investigation results of over voltage protection circuit for low power applications.

III. Investigations and Analysis

Over voltage protection circuits for low power applications have been investigated. The achieved simulations results are received with Cadence OrCAD Captures, which is appropriate tool for analysis of discrete electronic devices. In the modern semiconductor technologies power supply voltages decrease. Protection circuits which switch-off protected equipment from input voltages higher than 3 V and 5 V are analyzed in this section.

Block diagram of the simulated over voltage protection circuit is shown in Fig. 3.

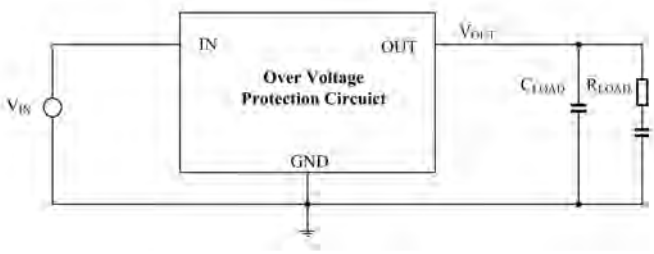


Fig. 3. Over Voltage Protection Circuit for Low Power Applications.

If the over voltage appears at the input, load will be disconnected from the power supply. Simulation results for the circuit which protects load from voltages higher than 3 V are presented in Fig.4.

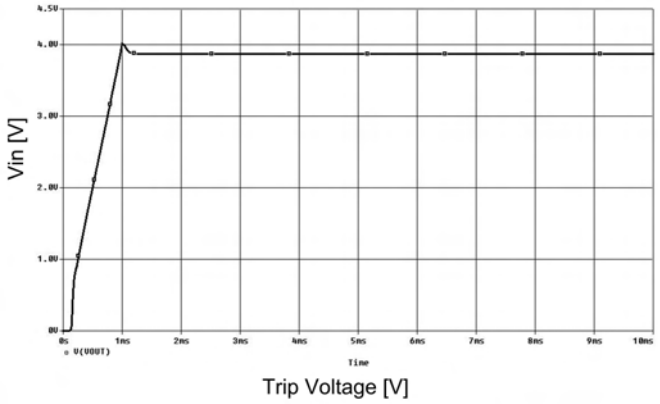


Fig. 4. Simulation results of circuits from Fig.3 when load is protected from input voltages higher than 3 V.

The trip voltage of the investigated electronic device is 3.86 V. When input voltage returns to the normal operation level of the devices under power, it is applied again to the load.

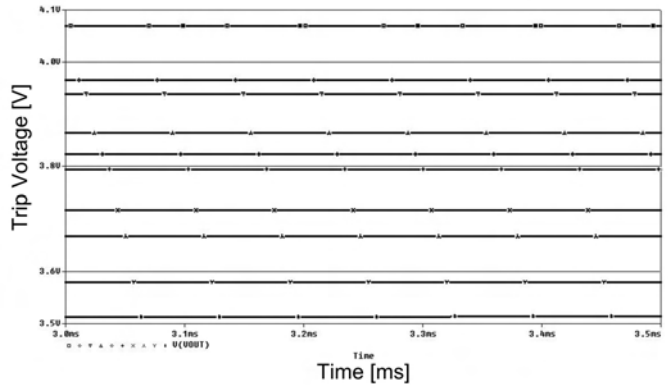


Fig. 5. Trip voltage at different temperatures for circuit from Fig. 3 when load is protected from input voltages higher than 3 V.

TABLE I
TRIP VOLTAGE AS A FUNCTION OF TEMPERATURE WHEN LOAD IS PROTECTED FROM INPUT VOLTAGES HIGHER THAN 3 V.

Temperature [C]	V_{OUT} [V]
-50	4.06
-25	3.96
0	3.93
25	3.86
40	3.82
50	3.79
75	3.71
100	3.66
125	3.57
150	3.51

A temperature analysis of circuit from Fig. 3 is performed. The influence of trip voltage on temperature is simulated. The obtained results are shown in Fig. 5.

The detailed results are presented in Table 1. As can be seen from Fig. 5 and Table 1 when temperature changes from -50 C to 150 C the trip voltage decreases from 4.06 V to 3.51 V.

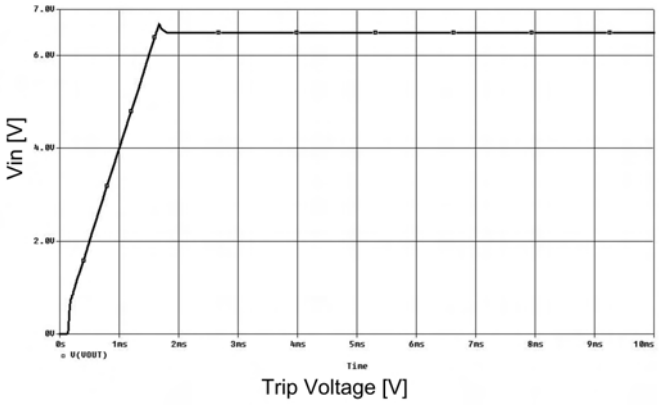


Fig. 6. . Simulation results of circuits from Fig.3 when load is protected from input voltages higher than 5 V.

The temperature coefficient of the trip voltage is -5.5 mV/ C.

The over voltage protection circuit reacts when input voltage is 26 % higher than 3 V. Thus eventual switch-off of the protected circuit from power supply is being eliminated, when small transient voltage fluctuations appear at the input.

Simulation results for the circuit from Fig. 3, which protects load from voltages higher than 5 V are presented in Fig.6. The trip voltage of this over voltage protection circuit is 6.47 V.

Influence of trip voltage on temperature is shown in Fig. 7.

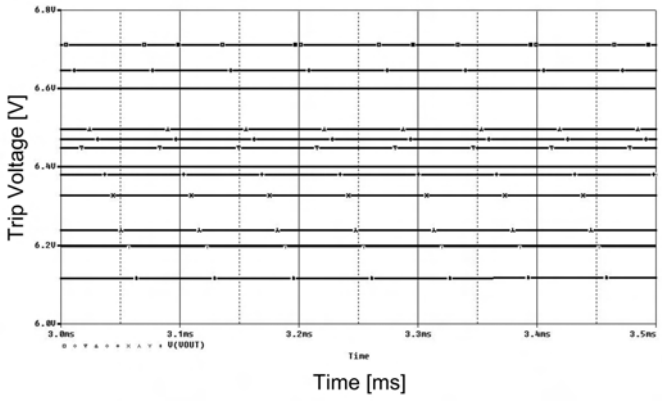


Fig. 7. Trip voltage at different temperatures for circuit from Fig. 3 when load is protected from input voltages higher than 5 V.

Detailed results of characteristics shown in the figure above are presented in Table 2. As can be seen from Fig. 7 and Table 2 when temperature changes from -50 C to 150 C the trip voltage decreases from 6.71 V to 6.19 V.

The simulated temperature coefficient of the trip voltage of the circuit from Fig. 3 when load is protected from input

voltages higher than 5 V is -6 mV/ C. The protection circuit reacts when input voltage is approximately 30 % higher than 5 V.

Temperature characteristics of trip voltage of investigated circuits are presented in Fig. 8.

TABLE II
TRIP VOLTAGE AS A FUNCTION OF TEMPERATURE WHEN LOAD IS PROTECTED FROM INPUT VOLTAGES HIGHER THAN 5 V.

Temperature [C]	V_{OUT} [V]
-50	6.71
-25	6.64
0	6.49
25	6.47
40	6.44
50	6.38
75	6.32
100	6.23
125	6.19
150	6.11

The temperature coefficient of trip voltage is negative due to the semiconductor voltage reference being used.

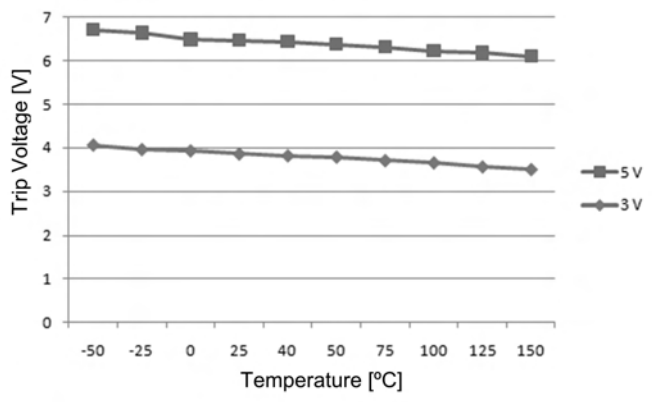


Fig. 8. Trip voltage versus temperature for circuit from Fig. 3 when load is protected from input voltages higher than 5 V.

IV. CONCLUSION

Investigation results of over voltage protection circuits appropriate for low power industrial and automotive applications are presented in this paper. The basic types and applications of surge protected devices, such as metal-oxide varistors (MOVs), gas discharge tubes (GDT) and silicon avalanche diodes (SAD) are shown. The received simulation results are achieved with Cadence OrCAD Captures. Temperature characteristics of trip voltage are analyzed when the electronic equipments are prevented from dc and transient level higher than 3 V and 5 V. The temperature coefficients of the trip voltage are -5.5 mV/ C and -6 mV/ C respectively when load is protected from input voltages higher than 3 V and 5V. They have negative values because used references are semiconductor components.

ACKNOWLEDGEMENT

The research described in this paper was carried out within the framework of Contract No ДУНК - 01/03 – 12.2009 (УНИК).

REFERENCES

[1] Maxim, http://www.maxim-ic.com/an760, “Overvoltage Protection in Automotive Systems”, Apr 02, 2002.
[2] Drabkin, M.M., “Surge Protection of Low-Voltage AC Power by MOV-Based SPDs” Harmonics and Quality of Power, 2002. 10th International Conference on, Vol.1, pp. 13 – 16, 6-9 Oct. 2002.
[3] Zola, J.G. “Gas Discharge Tube Modeling With Pspice”, *IEEE Transaction On Electromagnetic Compatibility*, Vol. 50, No.4, pp. 1022-1025, November 2008.
[4] Samaras K., C. Sandberg, C. J. Salmas, and A. Koulaxouzidis, “Electrical Surge-Protection Devices for Industrial Facilities – A Tutorial Review”, *IEEE Transaction On Industry Applications*, Vol. 43, No.1, pp. 150-161, January/February 2007.
[5] Henrry W. Ott, "Electromagnetic Compatibility Engineering", A John Wiley & Sons, Inc., Publications, 2009.
[6] National Semiconductor, http://www.national.com, “Over Voltage Protection Circuit for Automotive Load Dump”, National Semiconductor Corporation, December, 2006.

Realization of Low-frequency Amplitude Modulator and Demodulator with FPAAs

Ivailo Pandiev[1]

Abstract – **This paper presents a low-frequency amplitude modulator and demodulator based on Field Programmable Analogue Array (FPAA) devices. For creating the FPAA circuits, amplitude modulation and synchronous demodulation methods known from designing analogue RF transmission systems have been adapted. The proposed system allows redesigning the analogue conditioning stage thanks to the use of a FPAA device, which may be adapted to the requirements of the signal shape and the electrical parameters. For a sinusoidal carrier signal the amplitude modulation (AM) is performed with multiplier, while for a square-wave carrier signal, switch is employed instead of multiplier. In the synchronous demodulation the modulated carrier sinusoidal signal is multiplied by an unmodulated carrier signal with the same frequency and phase. For a square-wave carrier signal in the demodulation process the multiplier is replaced with switch, controlled by square-wave signal. The functional elements of the structures are realized by employing available configurable analogue modules (CAMs) of the FPAAs AN231E04 and AN220E04 from Anadigm. The FPAA circuits can operate with single supply voltage of +3,3V and +5V, respectively. The simulation and experimental results show good agreement with the theoretical predictions.**

Keywords – **Analogue circuits, Amplitude modulation, Synchronous demodulation, FPAA, System prototyping.**

I. INTRODUCTION

Low-frequency useful information can be conveyed by amplitude modulation (AM) of a high-frequency carrier wave. Many naturally occurring sounds are modulated in both amplitude and frequency. The speech and other specific communication signals in mammals, marine species and birds include amplitude and frequency modulation components. The ability to detect fluctuations in the amplitude and frequency of sounds is considered as important for normal auditory processing, including the perception of speech. Recently, several studies have been reported in the literature [1-5]. The majority of the papers contain analyses of human speech or present computer based approaches for detection of sinusoidal amplitude modulation components. Moreover, those schemes cannot be used to realize monolithic electronic systems implementing low-frequency amplitude modulation and demodulation. Without any doubt low-frequency amplitude modulators and demodulators are necessary for analog signal processing of the natural sounds. However, powerful low-frequency modulators and demodulators have not been available yet. This paper introduces new low-frequency amplitude modulators and demodulators based on Field Programmable Analogue Arrays (FPAAs). The FPAA integrated circuits (ICs) are flexible, economical and fast prototyping solutions for the design of complex analogue acquisition systems [6-8]. In particular FPAAs are basically suitable for signal conditioning, gain, filtering, summing and multiplying.

[1]Ivailo Pandiev is with the Faculty of Electronics at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ipandiev@tu-sofia.bg.

II. PRINCIPLE OF AMPLITUDE MODULATOR AND DEMODULATOR OPERATION

A. Modulation

In amplitude modulation (AM), the useful signal $u_S(t)$ modulates the amplitude of the carrier signal $u_C(t)$, while the phase of the carrier signal remains constant:

$$u_{AM}(t) = [U_{Cm} + u_S(t)]\cos\omega_C t\,. \tag{1}$$

For AM with carrier signal, the sinusoidal modulating signal

$$u_S(t) = U_{Sm}\cos\omega_S t \tag{2}$$

renders a modulated signal

$$u_{AM}(t) = [U_{Cm} + U_{Sm}\cos\omega_S t]\cos\omega_C t =$$

$$= \underbrace{U_{Cm}\cos\omega_C t}_{u_C(t)} + \underbrace{\frac{K_{AM}U_{Cm}}{2}\cos(\omega_C - \omega_S)t}_{u_{LSB}(t)} +$$

$$+ \underbrace{\frac{K_{AM}U_{Cm}}{2}\cos(\omega_C + \omega_S)t}_{u_{USB}(t)}\,, \tag{3}$$

where $K_{AM} = U_{Sm}/U_{Cm}$ is the modulation depth (≤ 1).

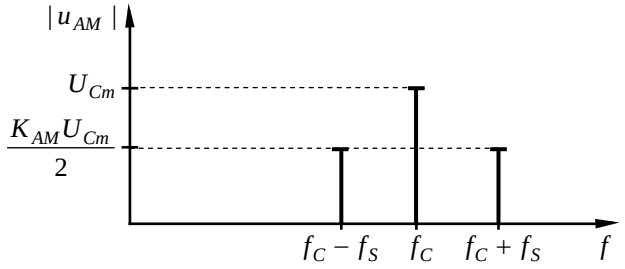


Fig. 1. Spectra of a modulated carrier with sinusoidal signal.

The modulated carrier signal consists of the *unmodulated carrier signal* $u_C(t)$, a useful signal $u_{LSB}(t)$ at the frequency $f_C - f_S$ in the *lower sideband* and a useful signal $u_{USB}(t)$ at

the frequency $f_C + f_S$ in the *upper sideband*. Fig. 1 shows the spectra of the modulated carrier signal.

To produce an amplitude-modulated signal according to (1), it is necessary to use a multiplier and a sinusoidal carrier signal $u_C(t) = U_{Cm} \cos \omega_C t$. Amplitude modulator with multiplier is shown on Fig. 2.

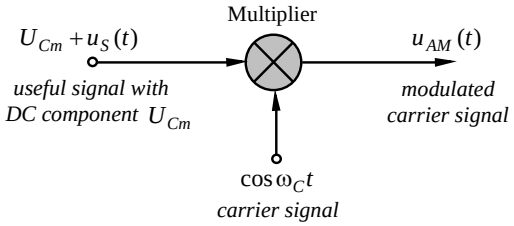


Fig. 2. Amplitude modulator with multiplier.

Instead of the sinusoidal carrier signal $\cos \omega_C t$, a square-wave signal with amplitude levels 0 and 1 and period length $T_C = 1/f_C$ can be used. In this case the multiplier has to be replaced with a switch [9]. From the Fourier series of the symmetrical (pulse duty ratio is equal to 50%) square-wave signal can be seen that besides the desired carrier with frequency f_C , other carrier components also occur with frequency equal to the carrier frequency f_C and multiplied by an odd integer (3, 5, etc.):

$$u_C(t) = \begin{cases} 1 & for \;\; nT_C \le (n+1/2)T_C \\ 0 & for \;\; (n+1/2)T_C < (n+1)T_C \end{cases} \qquad n \text{ integer}$$

$$= \frac{1}{2} + \frac{2}{\pi}\cos \omega_C t - \frac{2}{3\pi}\cos 3\omega_C t + \frac{2}{5\pi}\cos 5\omega_C t ...$$

$$= \frac{1}{2} + \frac{2}{\pi}\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}\cos(2n+1)\omega_C t \,. \qquad (4)$$

In the output resulting signal each components of Eq. (4) is modulated by the useful signal and has the corresponding sidebands. The desired carrier signal with its sidebands has to be extracted from this mixture with band-pass filter. On Fig. 3 amplitude modulator with switch and output band-pass filter is shown.

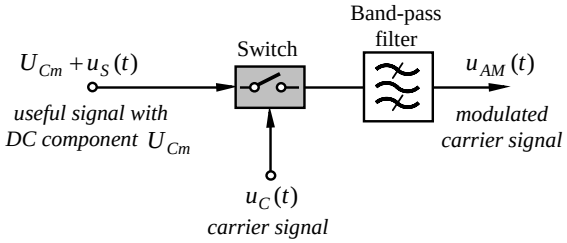


Fig. 3. Amplitude modulator with switch [9].

B. Demodulation

There are two basic types of electronic circuits, which realize amplitude demodulation: (1) *envelope detector* and (2) *synchronous demodulator*. In general the envelope detectors consist of a peak-type rectifier and a high-pass filter for rejecting the DC component [10]. The main advantage of the envelope detector is its simple design. A drawback is the nonlinearity of the transfer characteristic due to the nonlinear I-V characteristic of the diodes within the peak-type rectifier, especially at smaller amplitude of the carrier signal. The envelope detectors are used basically in the AM radio receivers.

Higher quality of the demodulation process can be achieved by synchronous demodulation, although this requires much more complex electronic circuit. In this demodulation method, the modulated carrier signal is multiplied by an unmodulated carrier signal with the same frequency and the same phase. For sinusoidal modulated signal this results in:

$$u_D(t) = u_{AM}(t)\cos \omega_C t =$$

$$= [U_{Cm} + U_{Sm}\cos \omega_S t]\cos \omega_C t \cos \omega_C t =$$

$$= [U_{Cm} + U_{Sm}\cos \omega_S t]\frac{1+\cos 2\omega_C t}{2} =$$

$$= \frac{U_{Cm}}{2} + \frac{U_{Cm}}{2}\cos 2\omega_C t + \frac{U_{Sm}}{2}\cos \omega_S t +$$

$$+ \frac{U_{Sm}}{4}\cos(2\omega_C - \omega_S)t + \frac{U_{Sm}}{4}\cos(2\omega_C + \omega_S)t \,. \qquad (5)$$

Besides the required component $U_{Cm} + U_{Sm}\cos \omega_S t$ the signal product $u_D(t)$ also contains additional components of 1/2 weighting in the double carrier frequency range. Those components can be suppressed by a low-pass filter. The synchronous modulator circuit is shown in Fig. 4. For demodulation of the modulated signal with square-wave signal the multiplier in Fig. 4 has to be replaced with a switch [9]. In this case the modulated signal is multiplied with square-wave signal with a period length $T_C = 1/f_C$. The resulting additional components in the signal product $u_D(t)$ are also suppressed by a low-pass filter.

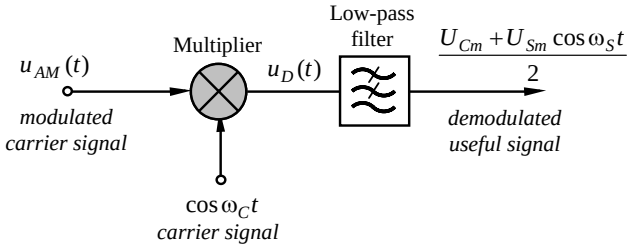


Fig. 4. Synchronous demodulator.

The synchronous demodulator with multiplier or switch largely corresponds to the amplitude modulator with multiplier or switch shown in Fig. 2 and Fig. 3, respectively. They differ only with regard to the necessary filters.

The use of a switch in the modulator requires additional band-pass filter in order to suppress unwanted (high-frequency) signal components. The low-pass filter in the synchronous demodulator is always required, regardless of whether a multiplier or a switch is used.

III. FPAA FOR LOW-FREQUENCY AMPLITUDE MODULATION AND DEMODULATION

This section analyzes the suitability of FPAA for low-frequency amplitude modulation and demodulation. A subsequent presentation of the Anadigm® FPAA architecture and functionalities [11] allows to link them with those required for conditioning of analogue signals with small amplitude and wide bandwidth.

The synthesis of the analogue system for modulation and demodulation is strongly conditioned by the incoming signal features, such as amplitude, frequency and noise level. This means that the implementation of the circuit would have to allow change the multiplication factor when variation in the signal amplitude occur and it should be able to change the pole (centre) frequency of the filters for different noise levels. These requirements are easily achieved using a FPAA to synthesize this analogue system. The basic advantages of the FPAA ICs offered by Anadigm are:

– Dynamically programmable parameters "*on-the-fly*";

– Offer a tradeoff between performance and system design time;

– Possibility for realization of complex analogue or mixed function without external passive components.

The main drawback presented by the use of FPAAs could be their high power consumption, compared with that of some ASICs. This drawback is affordable for implementations that are intended to be a portable instrument.

IV. FPAA CONFIGURATIONS FOR LOW-FREQUENCY AMPLITUDE MODULATION AND DEMODULATION

Once the FPAAs features have been presented, this section is focused on the key qualities and elements of those devices that will be used for amplitude modulation and demodulation. Moreover, the functional circuits of voltage-mode low-frequency modulator and demodulator based on FPAA AN231E04 [12] and AN220E04 [13] are presented.

FPAA configuration of amplitude modulator and demodulator with multiplier based on AN231E04 is shown on Fig. 5. The differential input useful signal is connected to the FPAA input cell 1 (pins 01 and 02). In bypass mode, the input signal is routed directly through the cell, bypassing all active circuit elements. The output differential modulated carrier signal is obtained by the output cell 6 (pins 17 and 18). In bypass mode, the cell's output pins are being driven directly by the low-pass filter connected to the output cell. The carrier sinusoidal signal is generated by internal oscillator, which output in connected to node *n2*. The oscillator is realized by using *OscillatorSine* CAM. The oscillating frequency range is from $10kHz$ to $20kHz$ and the output amplitude is equal to $1V$. The chip clock frequency of the oscillator is *Clock 3* with value $250kHz$. For the synchronous demodulator the input modulated carrier signal can be applied to the output cell 3 (pins 11 and 12) and the corresponding output demodulated signal is obtained by the output cell 7 (pins 19 and 20). Furthermore the demodulation process is synchronized with carrier signal, generated by the *OscillatorSine* CAM.

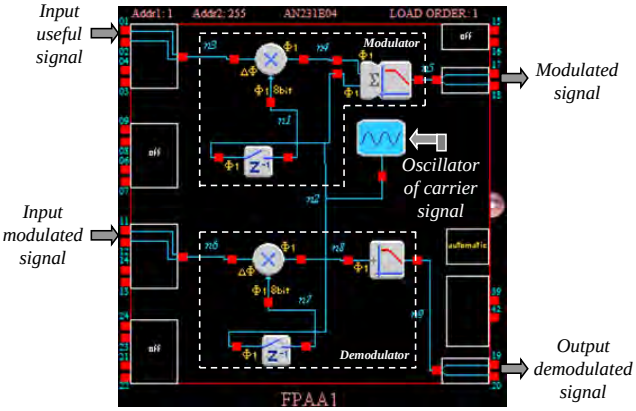


Fig. 5. FPAA configuration of amplitude modulator and synchronous demodulator with multiplier shown on Fig. 2 and Fig. 4, respectively.

FPAA configuration of amplitude modulator and demodulator with switch based on AN220E04 is shown in Fig. 6. The differential input unmodulated signal is connected to the FPAA input cell 1 (pins 09 and 10) and the output modulated signal is obtained by output cell 1 (pins 03 and 04). In the modulator with switch for realization of the switch is used a *GainSwitch* CAM. This CAM creates a gain stage with two switchable input terminals. The useful signal is applied to the upper input with $G1 = -1$ and the signal ground ($VMR = +2V$ (Voltage Mid-Rail − VMR)) is connected to the lower input with $G2 = +1$. Selection of the two inputs is controlled through a comparator, which is part of this CAM. This comparator has options similar to those of the *Comparator* CAM including the ability to select what the Control signal will be compared to. The carrier square-wave signal is generated by internal oscillator, realized by using *OscillatorTriSqr* CAM. The desired modulated carrier signal is extracted with band-pass *FilterBiquad* CAM. The corner frequency is equal to f_S and quality factor is $Q \approx f_C / 2 f_S$. For the synchronous demodulator with switch the input modulated signal is applied to the output cell 3 (pins 37 and 38) and the corresponding output demodulated signal is obtained by the output cell 2 (pins 07 and 08). The switch within the demodulator is realized with second *GainSwitch* CAM. The unwanted signal components are suppressed with low-pass *FilterBilinear* CAM.

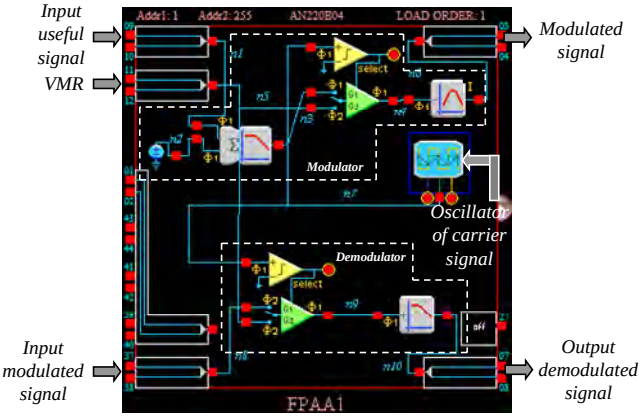


Fig. 6. FPAA configuration of amplitude modulator and demodulator with switch shown on Fig. 3 and Fig. 4 (the multiplier on Fig. 4 has to be replaced with a switch), respectively.

V. SIMULATION AND EXPERIMENTAL RESULTS

The workability of the proposed amplitude modulator and demodulator of Fig. 5 and 6 is presented by the simulation results using the simulator built in *AnadigmDesigner2* and also through the experimental results from the circuit configured on prototype board. The experimental test that has been used for validating the FPAA configuration shown on Fig. 5 is based on the AN231K04-DVLP3 – Development Board, which is built around the AN231E04 device.

The input and output waveforms of the FPAA configuration shown on Fig. 5 at carrier frequency $10kHz$ are illustrated on Fig. 7 and Fig. 8. The amplitude modulation of a useful sinusoidal signal with amplitude $500mV$ and frequency $500Hz$ is presented on Fig. 7. The demodulation process of the modulated carrier signal is shown on Fig. 8. The original unmodulated signal has amplitude $500mV$ and frequency $500Hz$. The goal of a 2% match between the simulation model, built in *AnadigmDesigner2* and the actual device was achieved.

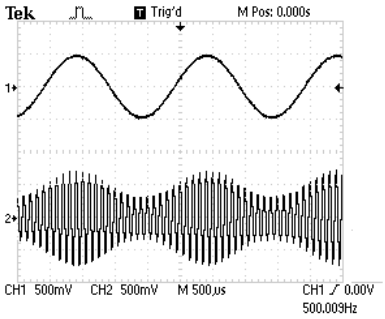


Fig. 7. Modulated signal – CH2 (single-ended – pin 17 on FPAA AN231E04) with carrier at sin-wave – CH1 ($K_{AM} = 0{,}5$).

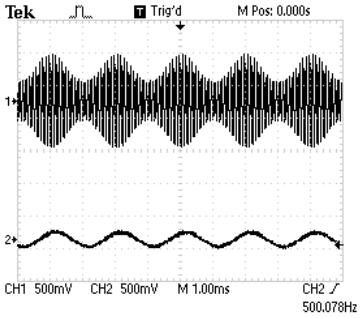


Fig. 8. Demodulated signal – CH2 (single-ended – pin 19 on FPAA) at AM with carrier – CH1 and $K_{AM} = 0{,}5$.

The input and output waveforms of the FPAA configuration on Fig. 6 at square-wave carrier $10kHz$ are shown on Fig. 9.

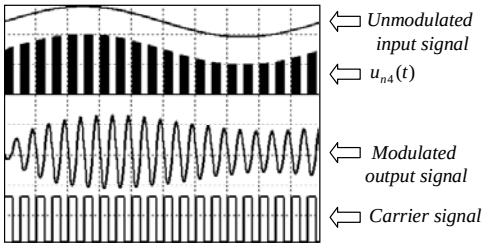


Fig. 9. Simulation results of the FPAA configuration on Fig. 6. The horizontal scale is $200\mu s/div$ and the vertical scale is $2V/div$.

VI. CONCLUSION

In this paper a low-frequency amplitude modulator and demodulator by using of a synchronous modulation and demodulation method has been proposed. The amplitude modulation circuit consists of a multiplier (or a switch and band-pass filter for a square-wave carrier signal). The synchronous demodulator circuit with multiplier or switch largely corresponds to the amplitude modulator with multiplier or switch. They differ only as regard to the necessary filters. The selected FPAAs are an Anadigm AN231E04 and AN220E04, where the analogue signal processing is implemented. The experimental and simulation results, obtained for various input signals confirm the theoretical analysis.

ACKNOWLEDGEMENT

This paper is sponsored by the research program of the Technical University of Sofia, Bulgaria.

REFERENCES

[1] H. Li, T. Li, "Generalized multichannel amplitude-and-phase coded modulation for differential space–time communications", Digital Signal Processing, vol. 17, pp. 261-271, 2007.

[2] I-H. Hsieh, K. Saberi, "Detection of sinusoidal amplitude modulation in logarithmic frequency sweeps across wide regions of the spectrum", Hearing Research, vol. 262, pp. 9-18, 2010.

[3] A. Qayoum, V. Gupta, P. Panigrahi, K. Muralidhar, "Influence of amplitude and frequency modulation on flow created by a synthetic jet actuator", Sensors and Actuators A, vol. 162, pp. 36-50, 2010.

[4] K. Banai, A. Sabin, B. Wright, "Separable developmental trajectories for the abilities to detect auditory amplitude and frequency modulation", Hearing Research, vol. 280, pp. 219-227, 2011.

[5] G. Prendergast, S. Johnson, G. Green, "Extracting amplitude modulations from speech in the time domain", Speech Communication, vol. 53, pp. 903-913, 2011.

[6] E. Manolov, P. Yakimov, M. Hristov, "Controllable square-wave generators by using FPAA", Elektrotechnica & Elektronica E+E, No 11-12, pp. 3-7, 2004 (in Bulgarian).

[7] Zh. Georgiev, E. Manolov, T. Todorov, I. Karagineva, "Synthesis and Experimental Verification of Sinusoidal Oscillator Based on the Modified Van der Pol Equation", Inter. Journal of Electronics, vol. 96, No 5, pp. 467-478, 2009.

[8] D. Morales, A. Garcia, E. Castillo, M. Carvajal, J. Banqueri, A. Palma, "Flexible ECG acquisition system based on analog and digital reconfigurable devices", Sensors and Actuators, vol. 165, pp. 261–270, 2011.

[9] V. Tietze, Ch. Schenk. *Electronic circuits*. 2nd Edition. Berlin, Heidelberg, New York. Springer-Verlag, 2008.

[10] M. Seifart. *Analoge Schaltungen*. 6 Auflage. Verlag Technik Berlin, 2003 (in German).

[11] The Anadigm Product Range, http://www.anadigm.com/products.asp, last accessed January 25, 2012.

[12] AN231E04 Datasheet – Dynamically Reconfigurable dpASP, http://www.anadigm.com/an231e04.asp, last accessed January 25, 2012.

[13] AN220E04 Datasheet – Dynamically Reconfigurable FPAA, http://www.anadigm.com/an220e04.asp, last accessed January 25, 2012.

Modification method to determining the output parameters in the audio power stage with complex load

Plamen Angelov[1]

Abstract –**The aim of the proposed article is to modified a known method to project the power stage in audio power amplifier. This method uses specific parameters of the output load, whose aim is to define the type of the output power transistors. Out of the numerical experiments we can adequately identify the electronics elements (type of the output transistors) in linear power stage. The results obtained in this scientific experiment relate to real complex load - DISCOVERY 15M/4624G00**

Keywords –**modification method, audio amplifier, power transistors**

I. PROBLEM CONDITION

In audio power amplifiers it is necessary harmonization between the high output resistance of the transistors and the low resistance of the load. To deliver the indispensable output current and low output resistance the power stages amplifier work in common collector scheme. The power amplifications is implemented by preamplifier, while the power stage works as current amplifier. At the other hand the load impedance is not completely active and it's value depends on the working frequency. Example of the impedance change of a loudspeaker with change of the frequency is shown on fig.1.1. When designing Hi-Fi audio amplifier this would lead to additional asymmetry in the work of the stage and respectively in the parameters of the output signal. To achieve ultimate effect it has to be considered the variable load impedance while designing an amplifier. When designing low-frequency amplifier class AB in the known so far methods of designing [1-4] account multiple parameters such as: maximal output power, bandwidth, total harmonics distortion, load impedance. In most cases the methodic doesn't account the impedance characteristics of the load, which makes it incomplete. Practically every loudspeaker has different frequency characteristic. Modification of this characteristic leads to change of it's resistance in function of the applied frequency (when this parameter is missing it has to be examined in laboratory). In many cases when designing audio power stage class AB this brings complications with conciliation to the output transistors. This requires additional measurements of the amending of the load impedance with the change of the operating frequency.

[1]Assist.PhD 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

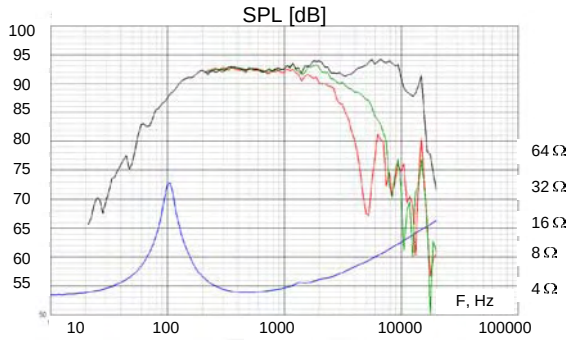


fig.1.1. Impedance characteristics of the loudspeaker Discovery – 15M/4624G00 [7]

The proposed modified method of designing accounts the equivalent electric scheme of the loudspeaker and its impedance characteristic. With this method all this parameters take part in the consistency of design. For this purpose will be examined the equivalent substitute scheme of loudspeaker shown on fig. 1.2. To define all parameters on the shown electric scheme of the loudspeaker are used the know equations [5],[6]

$$R_2 = \frac{(B.L)^2}{R_{ms}},\ L_2 = C_{ms}.(B.L)^2,\ C_2 = \frac{M_{ms}}{(B.L)^2} \quad (1)$$

where: voice coil resistance R_{dc} (dc resistance); voice coil inductance – L_0; Force factor - $B.L$; Suspension compliance – C_{ms}; moving mass of the loudspeaker– M_{ms}; mechanical losses resistance–R_{ms};

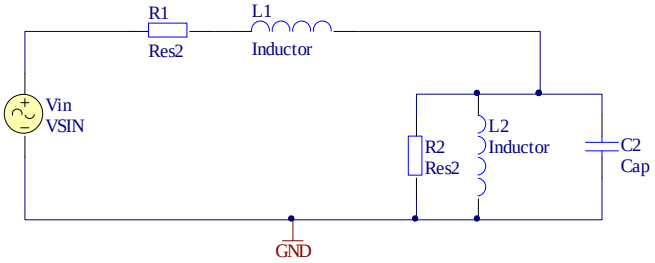


fig.1.2. Equivalent replacing scheme of loudspeaker

First let's convert the parallel double-pole in consecutive with the equations:

$$R_S(f) = \frac{G_2}{G_2^2 + B(f)_2^2},$$
$$X_S = \frac{B_2}{G_2^2 + B(f)_2^2}$$
$$G_2 = \frac{1}{R_2} \quad (2)$$
$$B_2(f) = \frac{1}{2 \cdot \pi \cdot f \cdot L_2} - 2 \cdot \pi \cdot f \cdot C_2$$

After that the shown equivalent scheme is converted to the different type shown on fig 1.3.

When defining the values of the R_s and X_s is more comfortable in their mathematical equations to write down the known limitations for resonance resistance ρ and resonance frequency ω_0. Using these parameters we can standardize the relative resistance of the circuit "k" and the relative frequency „η" with the equations:

$$\rho_2 = \sqrt{\frac{L_2}{C_2}}$$
$$\eta = \frac{R_2}{\rho_2}$$
$$f_0 = \frac{1}{2 \cdot \pi \cdot \sqrt{L_2.C_2}} \quad (3)$$
$$\kappa = \frac{f}{f_0}$$

Then for the serial equivalent transformations we are obtaining the equations:

$$R_S(f) = \frac{G_2}{G_2^2 + B_2^2} = R_2 \frac{1}{1 + \eta^2.(\kappa^{-1} - \kappa)^2} \quad (4)$$

$$X_S(f) = \frac{B_2}{G_2^2 + B_2^2} = R_2 \frac{\eta^{-1}.(\kappa^{-1} - \kappa)}{1 + \eta^2.(\kappa^{-1} - \kappa)^2} \quad (5)$$

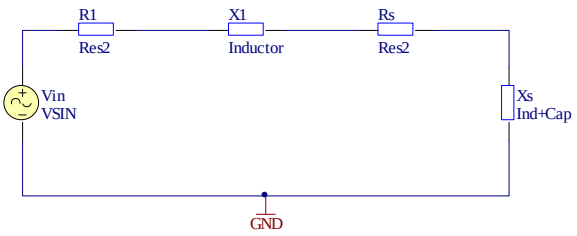


fig.1.3 conversion of parallel to consistent double-pole

$$R(f) = R_1 + R_2 \frac{1}{1 + \eta^2.(\kappa^{-1} - \kappa)^2}$$
$$X(f) = 2 \cdot \pi \cdot f \cdot L_1 + R_2 \frac{\eta^{-1}.(\kappa^{-1} - \kappa)}{1 + \eta^2.(\kappa^{-1} - \kappa)^2} \quad (6)$$
$$Z_{load}(f) = \sqrt{R(f)^2 + X(f)^2}$$

The obtained model of electric substitute scheme of loudspeaker provides simple way to design. The expected result will be refinement of the output parameters of the loudspeaker according to the specific equivalent model. From the obtained parameters we can make numeral experiment which will show us correct result if we are in the right direction. To verify the equivalent electrical model of the selected loudspeaker we derive graphics of the load's impedance characteristic. The simulative examination is performed on MathCAD with the following parameters : Voice coil resistance – R_{dc}=3,2Ω - Voice coil inductance – L_0=230μH; - Force factor - BL=5,3Tm; suspension compliance – C_{ms}=0,41mm/N; moving mass of the loudspeaker – M_{ms}=6,2g; mechanical losses of the loudspeaker –R_{ms}=0,69kg/s. After substituting the obtained data in eqation.1 for the values for the elements L2, R2, C2 occurs: L2=12mH, C2=220,7μF; R2=40,71Ω. When we use the obtained results in the electric substitute scheme for the amendment of the impedance of the loudspeaker we will get as result the result from fig.1.4. Comparing the result from the numerical simulation with the manufacturer's recommendations (fig.1.1) shows that the amendment of the impedance is identical. This means that the obtained mathematical dependencies describe quite accurate the amendment of the load and can be used for determining the active output power with equation 7.

$$P_l(f) = I^2.R(f) = \frac{U_{lm}^2}{2} \frac{R(f)}{R(f)^2 + X(f)^2} \quad (7)$$

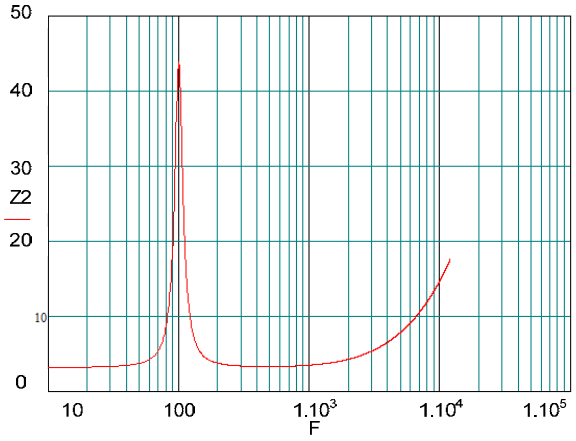


fig.1.4. Numerical experiment for determining the impedance characteristic of the loudspeaker Discovery – 15M/4624G00 [6]

II. MODIFICATION METHODOLOGY FOR DETERMINING THE OUTPUT TRANSISTORS IN LINEAR POWER AUDIO AMPLIFIER CLASS AB

A. *Choice of principle scheme*

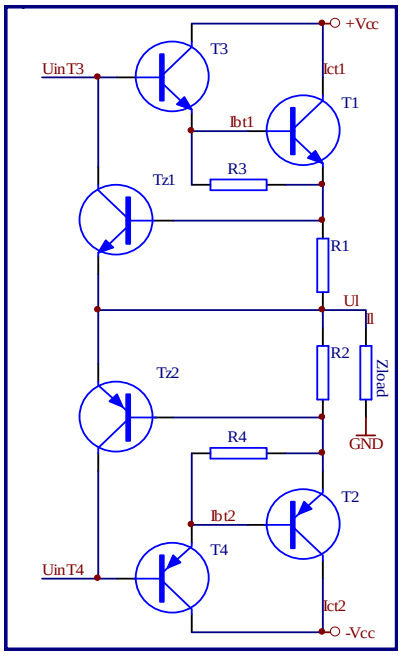


fig.2.1. Audio power stage

There is a big variety of schematic solutions for low frequency audio amplifiers. In the present article is observed circuit of linear power amplifier fog.2.1 class AB. The amplifier is realized with complimentary couple transistors T1 and T2. It's know that the terminal transistors used for relatively high power are with low coefficient h_{21E}. For this reason are added transistors T2 and T4. With their help it's formed darlington transistor T3-T1 which amplification is determinable with the equation:

$$h_{\sum 21e} = \beta_{T3}.\beta_{T1} \qquad (8)$$

where :

β_T - static coefficient of amplification of current (for common emitter circuit). The market offers constituent transistors, which unite as one unit T1 and T3.

In that case to account their power s used the familiar equation:

$$P_{\sum T1T3} = P_{T1} + P_{T3} \qquad (9)$$

where:

P_{T1}, P_{T3} - power dissipation for each transistor.

За проектирането са необходими следните изходни данни:

- The effective output power of the stage - P_l;
- bandwidth $f_l \div f_h$;
- Input Voltage U_i
- all of the above electrical parameters of the selected speaker.

B. *Determining the output voltage*

The first step is to determine the amplitude of the output voltage by the equation:

$$U_{lm}(f) = \sqrt{\frac{2.P_l.(R(f)^2 + X(f)^2)}{R(f)}} \qquad (10)$$

Substituting the result of equations (4) and (5) in (10) for the amplitude of the output voltage will obtain:

$$U_{lm}(f) = \sqrt{2.P_l.\frac{(R_1 + R_2\frac{1}{1+\eta^2.(\kappa^{-1}-\kappa)^2})^2}{R_1 + R_2\frac{1}{1+\eta^2.(\kappa^{-1}-\kappa)^2}}}... \sqrt{\frac{+(2\cdot\pi\cdot f\cdot L_1 + R_2\frac{\eta^{-1}.(\kappa^{-1}-\kappa)}{1+\eta^2.(\kappa^{-1}-\kappa)^2})^2}{R_1 + R_2\frac{1}{1+\eta^2.(\kappa^{-1}-\kappa)^2}}} \qquad (11)$$

where:

U_{lm} - maximum load voltage;

P_l - effective output power of the stage.

The displayed equation shows the frequency dependence of the output voltage according to the frequency change of the load Z_{load}. Using this parameter we according the output resistance of the stage in the specified frequency range.

After calculating the output voltage occurs graphic of the type:

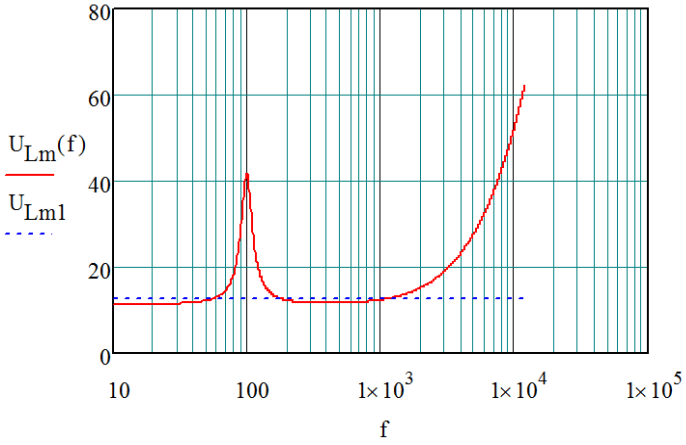


fig.2.2. Numerical experiment to determine the the output voltage

Let's conduct a numerical experiment with the chosen audio speaker with the following output parameters:

P_l=20W; f_l=40Hz; f_h=12kHz; R_1=3,2Ω; L_1=230μH; L_2=12mH; C_2=220,7μF; R_2=40,71Ω

In fig.2.2. with a dashed line is mapped second voltage value (U_{lm1}), obtained by numerical simulation with active load. Comparing the two results shows précising of the methodology to 400%, especially for two operating frequencies $f_1 = 100Hz$ $f_2 = 10{,}5kHz$

C. Determining the load current

Using the known equation for the output power we specifies the load current:

$$P_l = \frac{U_{lm}.I_{lm}}{2} \Rightarrow I_{lm} = \frac{2.P_l}{U_{lm}} \qquad (12)$$

where:

I_{lm} - maximum value of the load current;

U_{lm} - maximum value of of the voltage on the load;

P_l - effective output power of the stage.

Substituting equation (10) in (12) for the size of the load current is obtained:

$$I_{lm} = \frac{2.P_l}{U_{lm}} = \frac{2.P_l}{\sqrt{\frac{2.P_l.(R(f)^2 + X(f)^2)}{R(f)}}} \qquad (13)$$

This equation displays the amendment of the current through the load impedance at the specified limits. This way in the design is determined the maximum output current and design the audio amplifier without overloading the transistors.

Let's conduct a numerical experiment and determine the amendment of load current. After conducting the program examination are obtained graphics of the type:

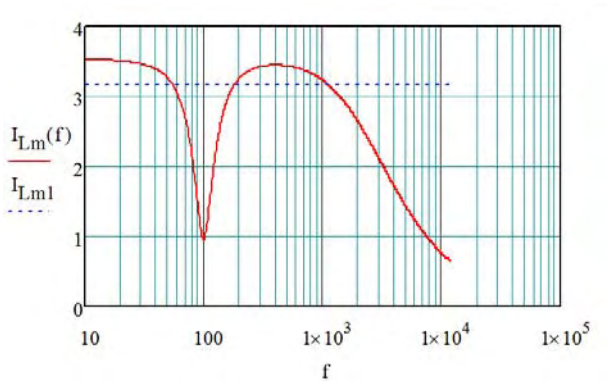


fig.2.3. Numerical experiment to determine the load current

In fig.2.3. with a dashed line is mapped second current value (I_{lm1}), obtained by numerical simulation with active load. Comparing the two results shows a significantly lower consumption at two frequencies $f_1 = 100Hz$ and $f_2 = 10{,}5kHz$ reason for which they are the resonant characteristics of the load.

III. Conclusion

When designing the transistor stage class in AB must be considered the frequency amendment of the loudspeaker. In the process of design the frequency amendment of output impedance determines parameters such as: amplitude of the output voltage, the load current amplitude.This contributes to the linear work of the transistor in the selected area of the class AB audio amplifiers.

References

[1] Г. Ненов „Усилватели изчисляване, измерване, регулиране” С. Техника 2000;

[2] В. Златаров „Електронни аналогови схеми и устройства” С. Техника 1994;

[3] И. Немигенчев „Аналогова схемотехника” Габрово изд. В.Априлов 2006;

[4] И. Атанасов „Ръководство за курсово проектиране на нискочестотни усилватели” ТУ Варна 2000;

[5] Макро модел на нискочестотно озвучително тяло тип band-pass Екатеринослав Събев Сираков - sirakov_akustika_08_3.doc;

[6] Band-Pass Loudspeaker Systems with Single Vent Ekaterinoslav S. Sirakov - ICEST_2009_Sirakov.doc

[7] Discovery 15M/4624G00. Scan Speak 2012, N.C. Madsensvej 1 6920 Videbaek, Denmark, www.scan-speak.dk

Modified method for design of the low-frequency audio driver

[1]Anton Petrov [2]Plamen Angelov - student mentor

Abstract–**The purpose of this study is to illustrate how different kinds of speakers impact on audio amplifiers and their output parameters. For this study using specific parameters of the output load. Results obtained in this study relate to complex real load.**

Keywords –**amplification, speaker, frequency response**

I. STATE OF THE PROBLEM

The condition of the problem raises the question – is it necessary to design the driver stage to be taken into account characteristic impedance of the load? The expected response will show how accurate is the design of the amplifier when the impedance characteristic is not given. A detailed account of this feature shows the influence of parallel resonance in low frequency region. This resonance frequency is caused by the mechanical properties of the speakers. I will do some software simulations and numerical experiments to show you how the resonance of the load affects the linear response of the amplifier.

fig.1.1. Chosen load speaker model 12M/4631G00

For the purpose of numerical experiments I choose a "midrange" speaker, model 12M/4631G00. Its impedance characteristic is shown on fig. 1.1., which shows the resonance of the system.

To confirm the recommended impedance characteristic is necessary to carry out numerical simulation. For the purposes of numerical experiments I will use the following data:

- Force factor [Bl] 5.3 Tm
- Mechanical resistance [Rms] 0.55 kg/s
- Movingmass [Mms] 6.5 g
- Suspensioncompliance [Cms] 0.69 mm/N
- DC resistance [Re] 3.2 Ω
- Voicecoilinductance [Le] 0.22 mH

fig.1.2. From references to speaker Midrange – 12M/4631G00[7]

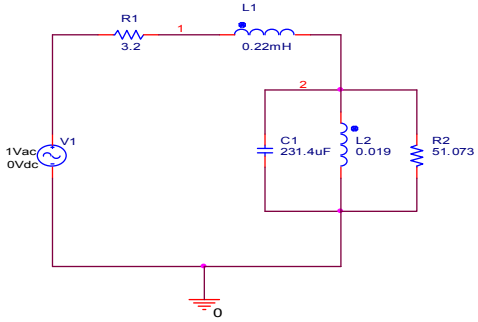
These parameters can be substituted for an equivalent scheme, which presents the speaker with an additional parallel circuit (C1, R2, L2), which characterize certain mechanical resonance of the system. The values for each parameter is determined by the familiar expressions [6]:

$$R_2 = \frac{(B.L)^2}{R_{ms}},\ L_2 = C_{ms}.(B.L)^2,\ C_2 = \frac{M_{ms}}{(B.L)^2} \quad (1)$$

For the purposes of the study (research) it is necessary to convert the resulting parallel circuit. For this purpose the well-known expressions will be used[2]:

$$R_2 = \frac{G^2}{G_2^{\ 2} + B_2^{\ 2}},\ X_s = \frac{B^2}{G_2^{\ 2} + B_2^{\ 2}},\ G_2 = \frac{1}{R_2} \quad (2)$$

After this transformation we get the following equivalent replacing scheme:



[1]Student Anton V.Petrov, Faculty of Computer Science and Engineering,Burgras Free University, 62San Stefano Str., Burrgas-8001, Bulgaria, e-mail: anton.petrov.007@gmail.com
[2]Mentor: assist. PhD. Prof. Plamen A. Angelov,Faculty of Computer Science andEngineering,Burgras Free University, 62San Stefano Str., Burrgas-8001, Bulgaria, e-mail: pangelov@bfu.bg

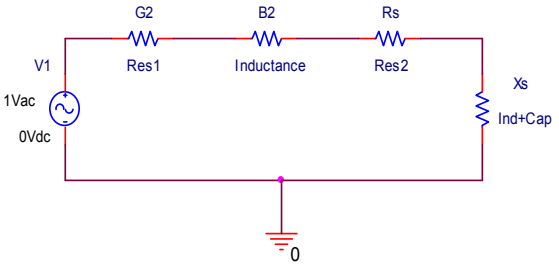


fig. 1.3. Convert parallel to serial circuit

After examine the mentioned parameters with the software, Mathcad I get a result that shows the change of electrical resistance of the speaker when changing its working frequency.

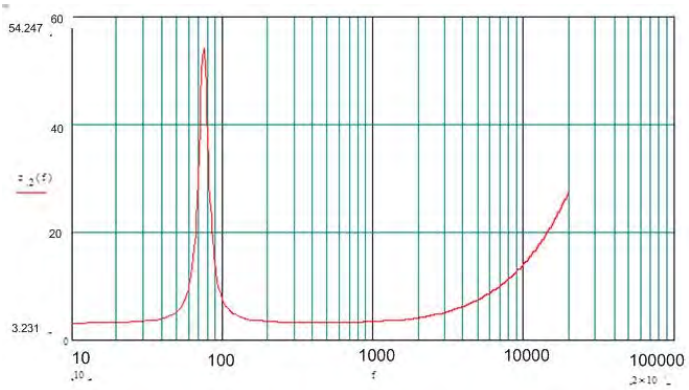


fig.1.4. For determining the impedance characteristics of the load Midrange – 12M/4631G00 I did the numerical experiment with MathCAD

Comparison of results from numerical simulation with those from the manufacturer indicates identical to the load. In consequence we can use the resulting parameters to modify the methodology of Scientifically LF driver stage. For that purpose the values obtained for R2, L2, C2 are replaced by the equivalent model of Pic.1.4.

II. DESIGNING THE DRIVER STAGE

When design the driver stage, the following initial parameters need to be set:

- Input impedance R_{in}=30kΩ;
- Bandwidth - f_l=40Hz f_h=17KHz ;
- The distortion coefficient – M_l = ?
- Input volgate – U_{in} = 0.7V;
- Impedance characteristics of the load – $Z_{load}(f)$

A. *Choice of principle scheme*

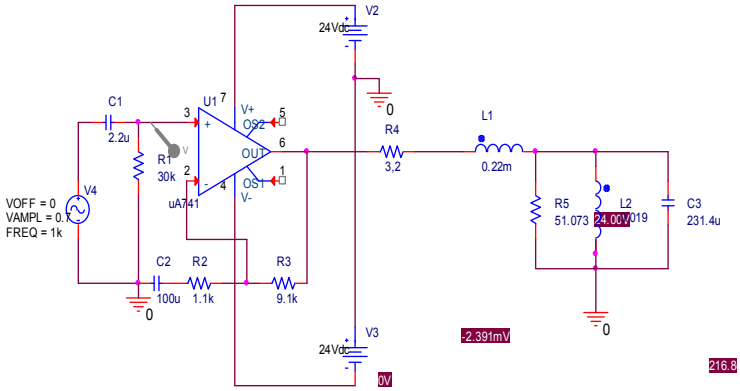


fig.1.5. Principal schematic of the driver stage

Selected scheme is a non-inverting audio amplifier filled with uA741. The difference in this article is the loading of the amplifier with a complex load, which changes the output parameters. To read this change in the design process we have to read all the parameters of the load in the design methodology.

B. *Define the amplification of the audio stage*

To determine the amplification coefficient we are going to use this familiar equation:

$$A_U = \frac{U_l}{U_{in}} \qquad (3)$$

where:
Ui - efficiency load voltage;
Uin - efficiency input voltage.

C. *Determination of the input impedance of the audio amplifier:*

$$Z_{in}(f) = \frac{R_1.R_{inOU}}{R_1 + R_{inOU}} - j.\frac{1}{2.\pi.f.C_1}$$
$$\approx R_1 - j.\frac{1}{2.\pi.f.C_1} \qquad (4)$$

where:

R_{inOU} - input resistance for the selected operational amplifier. When we choose the correct amplifier this resistance must execute the expression: $R_{inOU} >> R_{in}$;

D. *Determining the value of the resistor R1.*

$$A_U = 1 + \frac{R_3}{Z_{in}} = 1 + \frac{R_3}{R_1 - j.\frac{1}{2.\pi.f.C_2}} \Rightarrow$$

$$R_1 - j.\frac{1}{2.\pi.f.C_2} = \frac{R_3}{A_U - 1} \Rightarrow$$

$$R_1(f) = \frac{R_3}{A_U - 1} + j.\frac{1}{2.\pi.f.C_2} \quad (5)$$

Resulting expression shows that the resistor is dependent on frequency.

E. Determination of the dissipation power of resistance R1:

$$P_{R1}(f) = \frac{U_{in}^{\ 2}}{R_1^{\ 2} + (\frac{1}{2.\pi.f.C_2})^2} \cdot R_1 \quad (6)$$

After calculating equation.5 and equation.6 we record the nearest larger value. When calculation we get the result - R1 =30kΩ

F. Determination of the required output voltage for operational amplifier

To determine this parameter the equality must be given:

$$U_{\max OU} = U_{l\max} + U_{linou} \quad (7)$$

where:

$$U_{linou} = 2 \div 3V$$

It is well-known that all operational amplifiers are produced with relatively low supply voltage $U_{\max OU} \le \pm 18 \div \pm 25V$. This limits the stage. For this output signal of the operational amplifier further includes raising the output voltage. For that reason we choose the supply voltage to be 44V, this requires an additional scheme to reduce ten siongripped (the supply) to the operational amplifier.

G. Determine the value of each capacitors C1 and C2.

Capacity C1 is determined by the expression (7):

$$C_1 \ge \frac{1}{2.\pi.f_l.R_1.\sqrt{M_l^{\ 2} - 1}} \quad (7)$$

After calculating (7) the nearest larger standard value is recorded. Objectives of the program study the size of the capacitor C1=2.2 uF

Capacity of capacitor C2 is determined by the expression:

$$(8)\ C_2 \ge \frac{1}{2.\pi.f_l.R_9.\sqrt{M_l^{\ 2} - 1}} \quad (8)$$

After calculating (4) the nearest larger standard value is recorded. For the purpose of the numerical experiment value of the capacitor will be C2=100 uF

H. Determination of operating voltage for each of the capacitors C1 and C2.

Considering the frequency characteristics of the load for voltage U_{C1} we can summarize the following equation [1] [2] [3] and [4]:

$$U_{c1} = U_{c2} \ge (3 \div 5) \cdot U_{in\max} \quad (9)$$

After calculating (9) the nearest larger standard value is recorded. Operating voltage of C1 and C2 comply with the maximum input level - 1.096V when we choose an operating voltage of 25V and a working temperature of 105° C

I. Produce a model for simulation program

The equivalent replacement scheme of the speaker load up the audio step, which is the purpose of our study or this will show how to change the operation of the amplifier with the following parameters of the load. The size of each R2, L2, C1 is determined by expressions. (1).

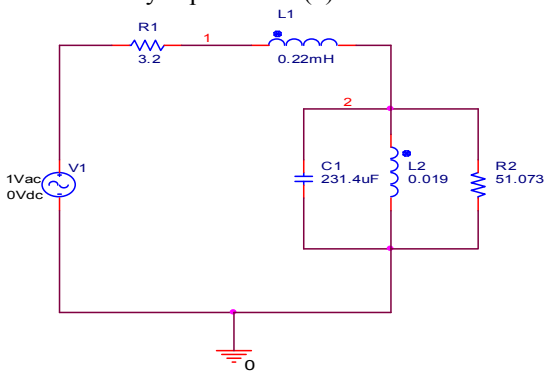


fig. 1.6. Equivalent scheme of replacement speaker

After we determine the load, we should calculate all parameters of the amplifier (fig.1.5.). We get the values for each of the elements after calculate expressions 4. to 9. The model for simulation is made for software "Proteus" fig.1.5. With this model we examine: the resulting gain, frequency response and level of output noise. In that way we study how the output level is changed in the presence of complex load audio amplifier.

III. RESULTS OF THE PROGRAM SIMULATION

Result of the analog analysis

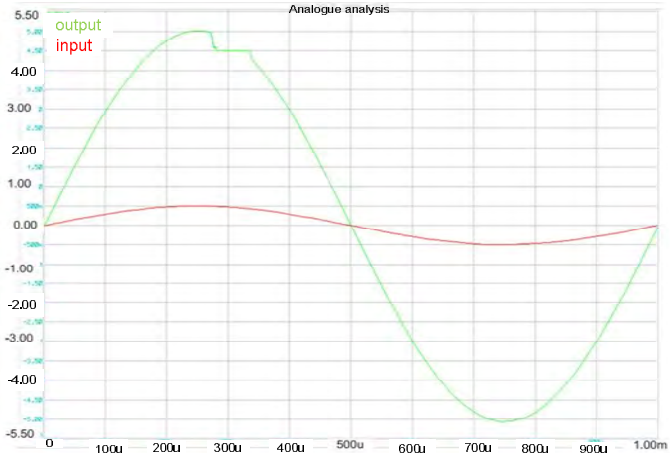


fig.2.1. Define the maximum amplification of the audio stage

This is analog analysis which shows the input and the output signal level. There is a 10-fold amplification of the input signal.

The frequency response

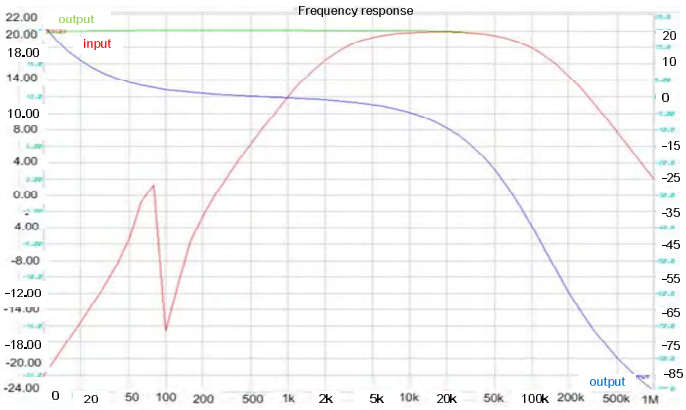


fig.2.2. Program simulation to determine the frequency response

Frequency response, exhibited here shows that the linear region of the step is covered in a range of 10Hz to about 100kHz, but when loaded with complex impedance of the speaker there is drop at a frequency less below 1kHz.

Distortion analysis

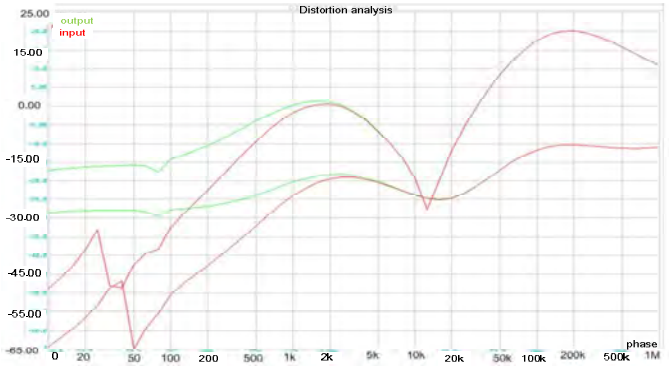
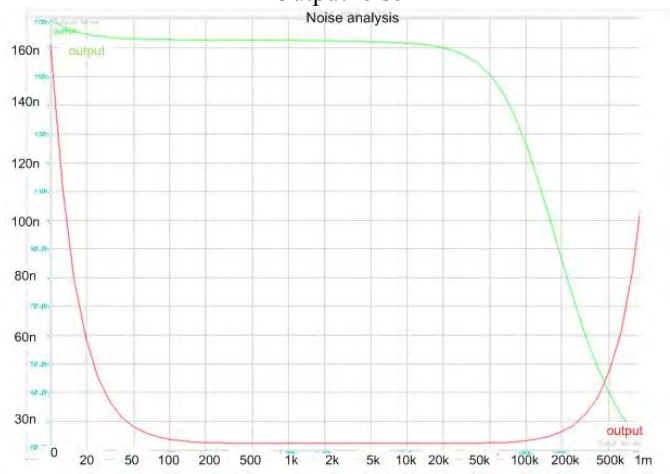


Fig.2.3. Numerical simulation of amplification distortion

Amplification system has a nonlinear form for compensation during the resonance.

Outputnoise



Pic.3.4. Numericalsimulation of theoutputnoise

The change of the output noise amplifier 45nV compared with the input, which is observed 29nV fold increase in noise.

IV. CONCLUSION

When design the driver stage high frequency speaker response must be considered. In the process of designing, this change determines parameters such as choice of supply voltage and amplitude of the output voltage.

REFERENCES

[1] Г. Ненов „Усилватели изчисляване, измерване, регулиране” С. Техника 2000;
[2] В. Златаров „Електронни аналогови схеми и устройства” С. Техника 1994;
[3] И. Немигенчев „Аналогова схемотехника” Габрово изд. В.Априлов 2006;
[4] И. Атанасов „Ръководство за курсово проектиране на нискочестотни усилватели” ТУ Варна 2000;
[5] Макро модел на нискочестотно озвучително тяло тип band-passЕкатеринослав Събев Сираков - sirakov_akustika_08_3.doc;
[6] Band-PassLoudspeaker Systems withSingleVentEkaterinoslav S. Sirakov - ICEST_2009_Sirakov.doc
[7] Midrange 12M/4631G00. ScanSpeak 2012, N.C. Madsensvej 1 6920 Videbaek, Denmark, www.scan-speak.dk

SPICE Modelling of Magnetoresistive Sensors

Boyanka Nikolova[1], Georgi Nikolov[2] and Milen Todorov[3]

Abstract – **In this paper are presented non-linear characteristics of magnetoresistive sensors and methods for their curve fitting. For this purpose is used Curve Fitting Toolbox by MathWorks MATLAB, which provide powerful processing capabilities. The results of curve fitting are used as input data in SPICE model. To creating a SPICE model are necessary ABM (Analog Behavioral Model) sources. They use as input data mathematical expressions from preceding fitting of sensors' characteristics. For this purpose is used National Instruments Multisim.**

Keywords – **Magnetoresistive sensors, Curve fitting, SPICE model, MATLAB, Multisim.**

I. Introduction

Sensors that monitor properties such as temperature, pressure, strain or flow provide an output signal that is directly related to the desired parameter. Magnetic sensors, on the other hand, differ from most of these detectors as they very often do not directly measure the physical property of interest. They detect changes, or disturbances in magnetic fields that have been created or modified by objects or events.

One of many advantages of using magnetic field for sensing position and distance is that any nonmagnetic material can be penetrated by the filed with no loss of position accuracy. Another advantage is that the magnetic sensors can work in severe environments and corrosive situations because the probes and targets can be coated with inert materials that will not adversely affect the magnetic fields.

Magnetoresistive sensors are well suited to measuring both linear and angular position and displacement in the Earth's magnetic field. For functioning, they require an external magnetic field. Hence, whenever the magnetoresistive sensor is used as proximity, position, or rotation detector, it must be combined with a source of a magnetic field. These sensors offer non-contact operation, high reliability, low and stable offset, low sensitivity to mechanical stress, much more insensitive to vibrations than inductive sensors, high operating temperature, wide operating frequency range, short response time and small size. They therefore provide an excellent means of measuring both linear and angular displacement under extreme environmental conditions, because their very high sensitivity means that a fairly small movement of actuating components in, for example cars or machinery can create measurable changes in magnetic field. Other applications for magnetoresistive sensors include rotational speed measurement and current measurement. The output signal requires some signal processing for translation into the desired parameter.

[1]Boyanka Nikolova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: bnikol@tu-sofia.bg.

[2]Georgi Nikolov is with the Faculty of Electronics and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

[3]Milen Todorov is with the Faculty of Electronics and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria.

II. Base Characteristics of Magnetoresistive sensors

For functioning, magnetoresistive sensors require an external magnetic field. Hence, whenever the magnetoresistive sensor is used as a proximity, position, or rotation detector, it must be combined with a source of a magnetic field. Usually, the field is originated in a permanent magnet, which is attached to the sensor. When the sensor is placed in the magnetic field, it is exposed to the fields in both the *x* and *y* directions. The sensor operates like a magnetic Wheatstone bridge measuring nonsymmetrical magnetic conditions [1, 2, 3].

In a Giant Magnetoresistance (GMR) sensors, the resistance of two thin ferromagnetic layers separated by a thin non-magnetic conducting layer is changed if the magnetic moments of the ferromagnetic layers are changed from antiparallel to parallel. Layers with parallel magnetic moments display less scattering at the interfaces, longer mean free paths, and lower resistance than the layers with antiparallel magnetic moments. The layers must be thinner than the mean free path of electrons (typically lower than 10 nm); if not, the spin-dependent scattering cannot be a significant part of the total resistance. The structures used in GMR sensors are unpinned sandwiches, antiferromagnetic multilayers, and spin valves.

Colossal Magnetoresistance (CMR) sensors under certain conditions undergo a semiconductor-to-metallic transition with the application of a magnetic field of a few teslas (tens of kilogauss). The size of the resistance ratios is 103% – 108%.

The Anisotropic Magnetoresistive (AMR) sensors are suitable for use in the measurement of magnetic fields in the range up to 200 μT. These sensors are made of a Permalloy (NiFe) thin film deposited on a silicon wafer and are patterned as a resistive strip. During deposition of the Permalloy strip, a strong external magnetic field is applied parallel to the strip axis. By doing this, a preferred magnetization direction is defined within the strip.

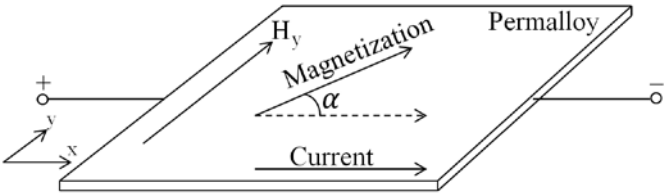


Fig. 1. The magnetoresistive effect in Permalloy.

The properties of AMR films cause the resistance to change by 2-3% in the presence of an external magnetic field.

In absence of any external magnetic field, the magnetization always points into this direction. In Fig.1 this is assumed to be the *x*-direction, which is also the direction of current flow. An AMR sensor now relies on two basic effects:

1. The strip resistance R depends on the angle α between the direction of the current and the direction of the magnetization;

2. The direction of magnetization and therefore α can be influenced by an external magnetic field H_y, which is parallel to the strip plane and perpendicular to the preferred direction.

When no external magnetic field is present, the Permalloy has an internal magnetization vector parallel to the preferred direction and $\alpha = 0^o$. In this case, the strip resistance R has its maximum value R_{max}. When an external field H_y is applied, the internal magnetization vector of the Permalloy will rotate around an angle α. Resistance reaches its minimum value R_{min} at high field strengths because angle $\alpha = 90^o$. Eq. 1 gives the functional dependence between the resistance and the angle:

$$R = R_0 + \Delta R \cos^2 \alpha , \qquad (1)$$

where $R_0 = R_{min}$ and $\Delta R = R_{max} - R_{min}$. Function R versus H_y is:

$$R = R_0 + \Delta R\left(1 - \left(\frac{H_y}{H_0}\right)^2\right). \qquad (2)$$

H_0 is parameter, which depends on material and geometry of the strip. Eq. 1 is defined for field strength magnitudes of $H_y \leq H_y$. R_0 and ΔR are also parameters of material. For Permalloy, ΔR is in the range of 2 to 3% of R_0.

Because quadratic Eq. 2 it is obvious that this characteristic is non-linear and each value of R is not associated with unique value of H.

The magnetoresistive effect can be linearized by depositing aluminium stripes (Barber poles), on top of the Permalloy strip, at an angle of 45º to the strip axis (Fig. 2). Because a higher conductivity of aluminum than Permalloy, the effect is rotation of current direction by 45º. Thus the angle between the magnetization and the electrical current is changed from α to $(\alpha - 45^o)$.

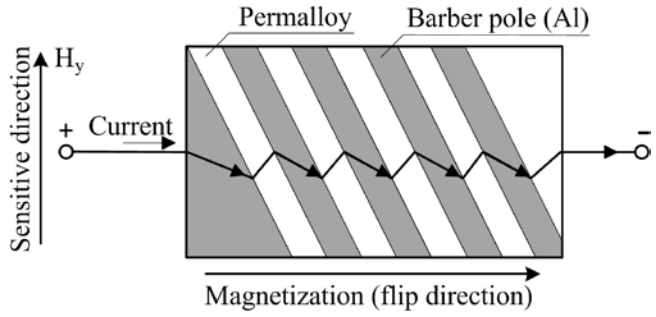


Fig 2. Linearization of the magnetoresistive effect.

AMR sensors are usually used in a Wheatstone bridge configuration with diagonally opposite sensors having barber pole orientations of ±45º (Fig. 3). This arrangement helps to:

– reduce the temperature drift of the sensor;

– increase the sensitivity of the sensor.

The best results are obtained from a Wheatstone bridge configuration when a current source rather than a voltage source drives the bridge. Use of a current drive doubles the bridge linearity and, in the ideal case, the temperature dependence is reduced [1, 3, 6].

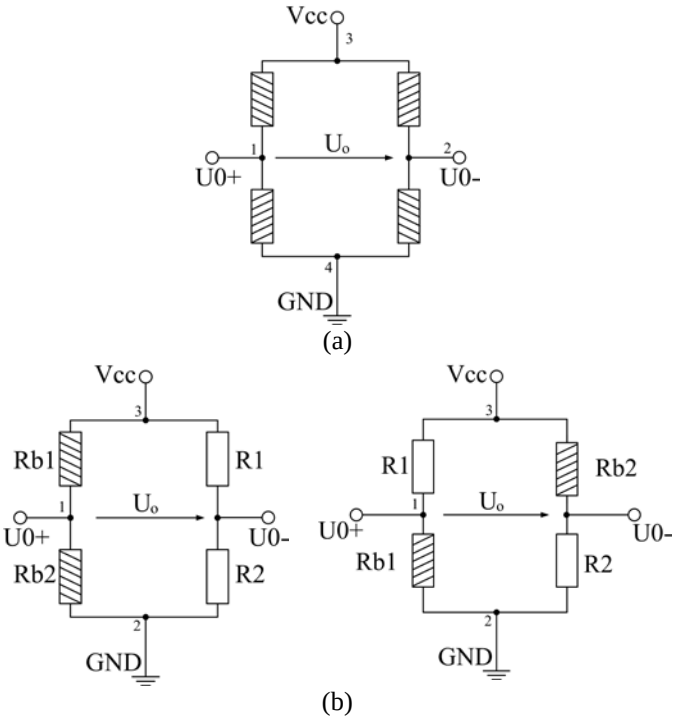


Fig. 3. Types of Wheatstone bridge:
(a) for full bridge sensors, (b) for half bridge sensors

Conventional Wheatstone bridge signal conditioning circuits can be used to process the AMR bridge. The bridge sensitivity and zero offset are proportional to the bridge voltage, so it is important to use a well-regulated supply with low noise and good temperature stability.

III. PARAMETRIC FITTING WITH MATLAB LIBRARY MODELS

Parametric fitting (Table I) involves finding coefficients (parameters) for one or more models that fit to data. The data is assumed to be statistical in nature and is divided into deterministic and random components. The first component is given by a parametric model and the second one is an error, which represents random variations in the data that follow a specific probability distribution. The model is a function of the independent variable and one or more coefficients [7, 8].

After fitting data with one or more models is important to evaluate the goodness of fit. The first step for that is a visual examination of the fitted curve displayed in Curve Fitting Tool.

Beyond that, the toolbox provides these methods to assess goodness of fit:

– Residual analysis;

– Goodness of fit statistics;

– Confidence and prediction bounds.

TABLE I
LIST OF LIBRARY MODELS FOR CURVE FITTING

Library Model	Description
Distribution	Distribution models such as Weibull.
Exponential	Exponential function and sum of two exponential functions.
Fourier	Up to eight terms of Fourier series.
Gaussian	Sum of up to eight Gaussian models.
Interpolant	Interpolating models, including linear, nearest neighbour, cubic spline, and shape-preserving cubic spline.
Polynomial	Polynomial models, up to degree nine.
Power	Power function and sum of two power functions.
Rational	Rational equation models, up to 5^{th} degree/5^{th} degree (i.e., up to degree 5 in both numerator and denominator).
Sin	Sum of up to eight sin functions.
Spline	Cubic spline and smoothing spline models.

IV. MODELLING OF MAGNETORESISTIVE SENSORS

A. Structure of the SPICE model

The ABM feature of SPICE can be used to make flexible descriptions of electronic components in terms of a transfer function or look-up table. The basic structure of the magnetoresistive sensors ABM model using Wheatstone half bridge is shown in fig. 4.

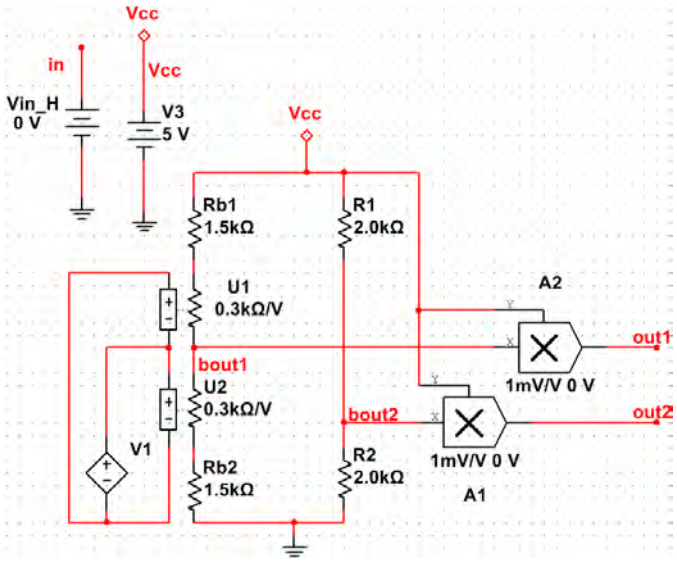


Fig. 4. SPICE Model of Magnetoresisitive Sensors

The main parts of the proposed model are:

1. V1 – ABM voltage source. The polynomial expression is specified as a mathematical equation which presented transfer function of the sensors. The obtained output voltage is equivalent to the relative change of resistance over the simulated magnetic field.

2. U1 and U2 – The resistance of this voltage controlled resistor is controlled by the voltage that is applied across the "+" and "-" terminals.

3. Rb1 and Rb2 – resistors using for modelling the nominal resistance of magnetoresistive elements of the sensor.

4. R1 and R2 – the value of the resistors is equal to the bridge resistance given in the datasheets.

5. A1 and A2 – Multipliers using for calculating the output voltage of the bridge according to the supply voltage.

B. Curve fitting of transfer characteristic of AMR Sensor KMY 21 M

After fitting data with one or more models is very important to evaluate the goodness of fit. The two stages are: visual examination of the fitted curve and estimation of numerical results. Plotting residuals and prediction bounds are graphical methods. They are more beneficial and aid visual interpretation, while numerical results are more narrowly focused on a particular aspect of the data and often try to compress that information into a single number [4, 5, 9].

For fitting models of transfer characteristic of AMR Sensor KMY 21 are chosen 9^{th} and 8^{th} degree polynomial fits. The data, fit, prediction bounds and residuals for fitting are shown on Fig. 5.

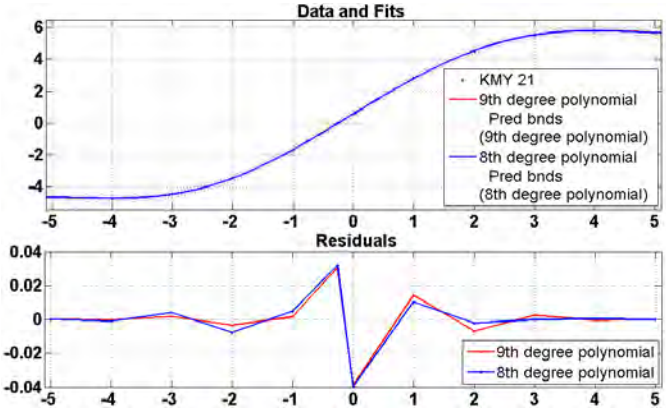


Fig. 5. Graphical results of curve fitting of transfer characteristic of AMR Sensor KMY 21.

TABLE II
NUMERICAL RESULTS OF CURVE FITTING

Fit	Polynomial	
Degree	9^{th}	8^{th}
Coefficients:		
p1	-1.975e-005	-1.172e-005
p2	-1.256e-005	-7.187e-005
p3	0.0009386	0.0005617
p4	0.0006047	0.006016
p5	-0.01049	-0.007533
p6	-0.008237	-0.2229
p7	-0.1276	0.008468
p8	0.01255	5.008
p9	4.866	0.4788
p10	0.4727	–
Goodness of fit:		
SSE	0.007368	0.02864
R-square	1	1
Adjusted R-square	1	0.9999
RMSE	0.0607	0.09771

From Fig. 5 and Table II it is clear that the 9th degree polynomial fit gives better results than 8th degree polynomial fit (lower values of Sum of Squares due to Error (SSE), and Root Mean Squared Error (RMSE), which are closer to ideal case value – 0).

C. Simulation Results

The design approach for magnetoresistive sensors modelling described above is implemented for AMR Sensor KMY 21 M. The equation of ABM source V1 is 9th degree polynomial with coefficients given in Table II. Typical values for the circuit components according to datasheets [6] are $R_1 = R_2 = 2\ k\Omega$ and $R_{b1} = R_{b2} = 1{,}5\ k\Omega$ with temperature coefficient of resistance of 0,0032 1/°C.

Fig. 6 illustrates the simulation results of the transfer characteristic of AMR Sensor KMY 21 M, using the SPICE simulator. The error of the model is formed only from the accuracy of the curve fitting, because the characteristic is described by mathematical polynomial equation. Therefore using the proposed model, the sensor can be simulated and used not only in the linear region of the transfer characteristic but in relatively large scale of input magnetic fields.

The simulation results of temperature simulation of the input offset voltage is shown on fig. 7.

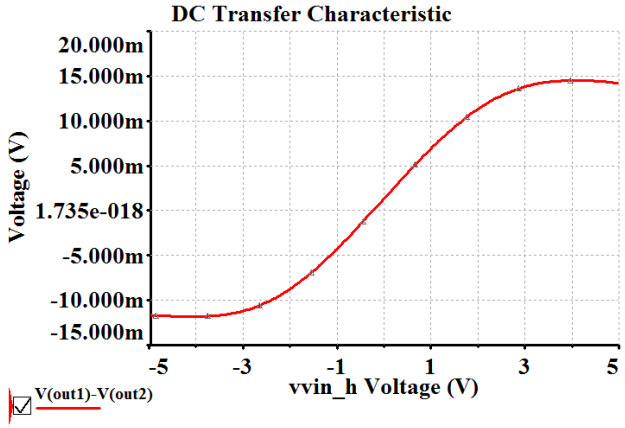


Fig. 6. Transfer characteristic of AMR Sensor KMY 21 M.

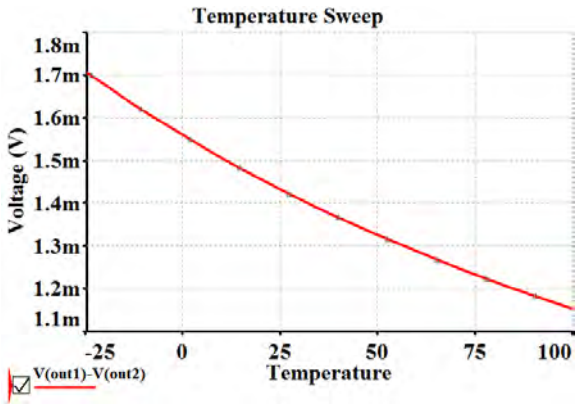


Fig. 7. Offset temperature dependence of AMR Sensor KMY 21 M.

In Table III are summarized main sensors' parameters and appropriate conditions for their simulations.

TABLE III
SENSOR SPECIFICATION

Parameter	Simulation Condition	Value
Supply Current	I_{CC} (V_{cc})	1,67 mA
Output voltage range	$\Delta V_o = (V_{out\ max} - V_{out\ min})$ H_y = -5 kA/m ÷ +5 kA/m	25,86 mV
Offset voltage	$V_{off} = V_{out}$ (H_y) $H_y = 0$	1,42 mV
Sensitivity	$S = \frac{dV_{out}}{dH_y}$ H_y = -1 kA/m ÷ +1 kA/m	5,57mV/kA/m
Temperature coefficient of offset	$TCV_{off} = \frac{V_{off}(T_2) - V_{off}(T_1)}{T_2 - T_1}$ T_1 = -25°C; T_2 = +125°C	-4,16 µV/°C
Temperature coefficient of amplitude	$TCV_o = \frac{\Delta V_o(T_2) - \Delta V_o(T_1)}{(T_2 - T_1)\Delta V_o(T_1)}$ T_1 = -25°C; T_2 = +125°C	-75,58 µV/°C

V. CONCLUSION

In this paper a SPICE behavioral model of magnetoresistive sensors is presented. An systematic approach for implementing curve fitting models and methods is applied in order to achieve equation that precisely describe sensor's transfer function. The main benefits of proposed model is that by using only one ABM voltage source many sensors' parameter can be defined like sensitivity, offset, nonlinearity and input range. The proposed modelling approach is suitable to predict behavior of other type of magnetoresistive sensors.

ACKNOWLEDGEMENT

This investigation has been carried out in the framework of the research project № DUNK-01/03 – 12.2009.

REFERENCES

[1] Ripka, P., A. Tipek. *Modern sensors handbook*, ISTE Ltd, ISBN 978-1-905209-66-8, 2007.
[2] Caruso, M., C. Smith, T. Bratland, R. Schneider, A New Perspective on Magnetic Field Sensing, Honeywell, Inc.
[3] *MagnetoResistive Field Sensor*, Sensitec, http://www.sensitec.com/
[4] Nikolova, B., G. Nikolov, M. Hristov. Analogue Behavioural Modelling of Integrated Sensors. The 2006 International Conference on Computer Engineering & Systems (ICCES'06), ISBN 1-4244-0271-9, pp. 107-112, Cairo, Egipt, November 5-7, 2006.
[5] Nikolova, B., M. Todorov, T. Brusev. Curve Fitting for Sensors' Analog Behavioural Modelling. Vol. 3, pp. 941-944, ISBN 978-86-6125-033-0, June 29-July 1, Niš, 2011.
[6] *KMY/KMZ Linear Magnetic Field Sensors*. Measurement Specialties, Inc., May 2011 (www.meas-spec.com).
[7] Riley, F., Hobson, P., Bence, J. *Mathematical Methods for Physics and Engineering*, Cambridge University Press, New York, 2006.
[8] Karris, Steven T., *Mathematics for Business, Science, and Technology with MATLAB® and Excel® Computations*, Orchard Publications, California, U.S.A., 2007.
[9] Curve Fitting ToolboxTM User's Guide R2012a. The MathWorks, Inc., 2001-2012.

Maximizing Power Transfer to the Remote Terminal of PCM4 System

Zoran Zivanovic[1] and Vladimir Smiljakovic[2]

Abstract – **This paper presents a method that allows maximum power transfer from Central Office Terminal (COT) to Remote Terminal (RT) of PCM 4 system for TELECOM use. This is achieved by employing a PWM controller as a Voltage Controlled Oscillator. The basic theory of FLYBACK converter is also presented in order to show the roots of the idea. The modified DC/DC converter was built and successfully tested. Performance measurements in the laboratory show that the goal is met.**

Keywords – **PCM4, Central Office Terminal, Remote Terminal, FLYBACK, PWM, DC/DC converter.**

I. INTRODUCTION

Telecom operators in the whole world use PCMx systems offering x (4, 6, 8 or 11) voice channels at 64kbps due to lack of copper cables (subscriber loops). Usually in urban areas there is unused switching capacity and lack of installed subscriber loops. The most popular is PCM4 which offers 5km reach, requires a lower voltage feed for remote powering and 4x64 kbps channels. This system consists of two main elements: COT (Central Office Terminal) and RT (Remote Terminal). RT is connected to COT via Unshielded Twisted Pair. The COT is powered from Central Office 48V battery and generates high voltage designed to power the Remote Terminal via twisted pair. The high voltage is no more than 200V and the current is limited to less than 60mA. Both power supplies are realized as a flyback converter for simplicity and lower cost.

II. PRINCIPLE OF FLYBACK CONVERTER

The basic flyback converter circuit is shown in Fig. 1. While the primary power switch (MOSFET) is on, the energy is taken from input and stored in the transformer. Actually it is a coupled inductor, because the current does not flow at the same time in the primary and secondary side. At the secondary the diode is reverse biased and the load takes energy from the output capacitor. When the power switch turns off, the output diode is forward biased and the energy is transferred to output capacitor and the load. There are two basic modes of operation. The first mode is discontinuous conduction mode (DCM) in which all of the stored energy is

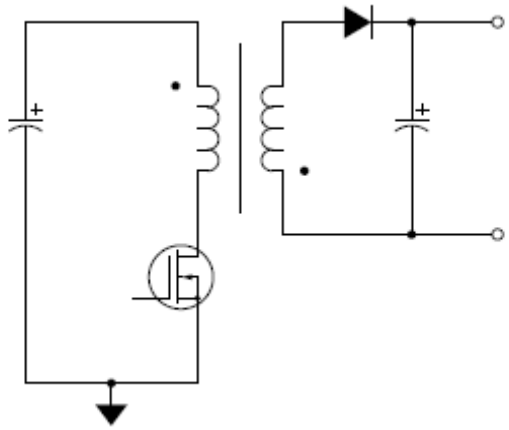


Fig. 1. Flyback converter

transferred to secondary during the time the power switch is OFF and the current has triangular shape. The second mode is continuous conduction mode (CCM) in which the part of the stored energy remains in the transformer when power switch turns on again and the current has trapezoidal shape (Figs.2 and 3). DCM converter requires a transistor and diode with higher current rating and bigger capacitors with low ESR (equivalent series resistance) and vice versa. For a given converter the operating mode depends on the switching frequency, input voltage and output load.

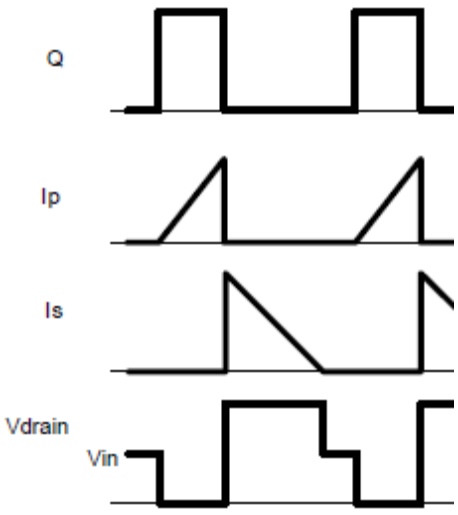


Fig. 2. DCM Waveforms

[1]Zoran Zivanovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina165b, 11070 Belgrade, Serbia, E-mail: zoki@insimtel.com.

[2]Vladimir Smiljakovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina165b, 11070 Belgrade, Serbia, E-mail: smiljac@insimtel.com

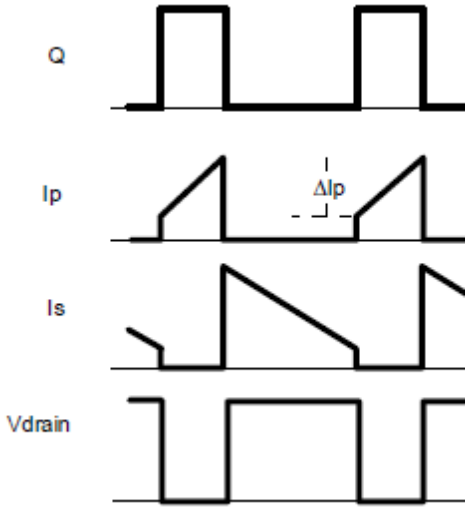


Fig. 3. CCM Waveforms

III. PROBLEM SOLUTION

Remote feeding voltage is 150V and current is limited to 50mA. The maximum loop length is 5km with 0.4mm copper wire and with DC loop resistance around 1350ohms. That means the power available at the end of the twisted pair is 4.125W (83Vx50mA) because the voltage drop over the loop is 67V. If we design the converter for RT with minimum operating voltage of 85V, at the beginning of twisted pair he will draw only 4.125W instead of 7.5W (150Vx50mA) available. The reason for this is that the converter is constant power load. When all four subscribers are off the hook, the power consumption of RT is around 5.5W which means that use of VRLA (Valve Regulated Lead Acid) 12V battery for RT is a necessity. Obviously we need the solution which maximizes power transfer from COT to RT automatic.

Again the magnitude of stored energy in flyback transformer is given in Eq. (1):

$$W = \frac{LI_p^2}{2} \tag{1}$$

where I_p is peak primary current.

The average power is given in Eq. (2)

$$P = \frac{W}{T} = \frac{LI_p^2}{2T} = \frac{1}{2}LI_p^2 f \tag{2}$$

As we see the power is directly proportional to the switching frequency. Varying the frequency we can change the power delivered to the Remote Terminal. We can achieve this by employing a PWM controller (in our case UCC2804 from Texas Instruments) as a voltage controlled oscillator (Fig. 4).

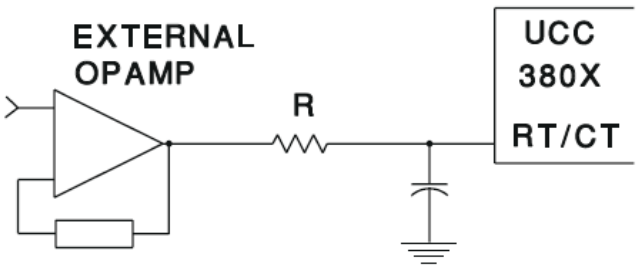


Fig. 4. Using PWM as a VCO

The oscillator section of the UCC2804 has single pin programming. It requires only a resistor to the reference voltage and capacitor to the ground. Using external operational amplifier we can change charging current and thus the frequency.

If we make the loop that regulates the feeding current to slightly lower value from 50mA, we can maximize power transfer from COT to RT using a handful of elements. The circuit is given in Figs. 5 and 6.

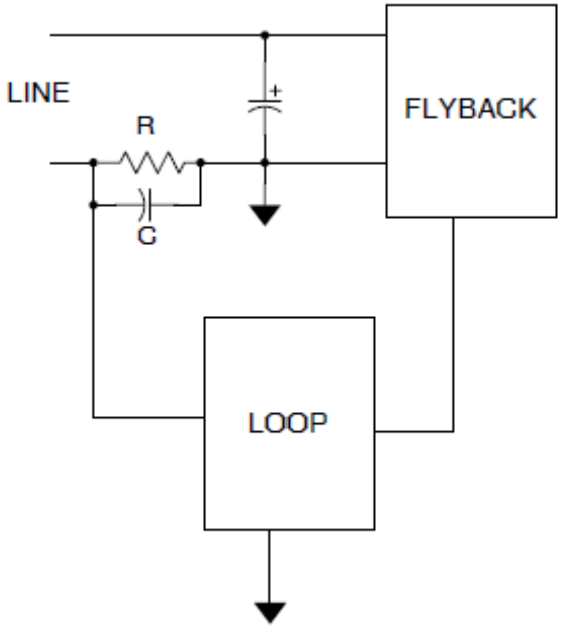


Fig. 5. Basic circuit

Voltage from the current sensing resistor R at the end of the subscriber line is fed via integrator (operational amplifier with R1 and C1) to the oscillator in PWM controller UCC2804 in order to regulate the frequency. Small adjustments of components values are needed in order to make the loop stable.

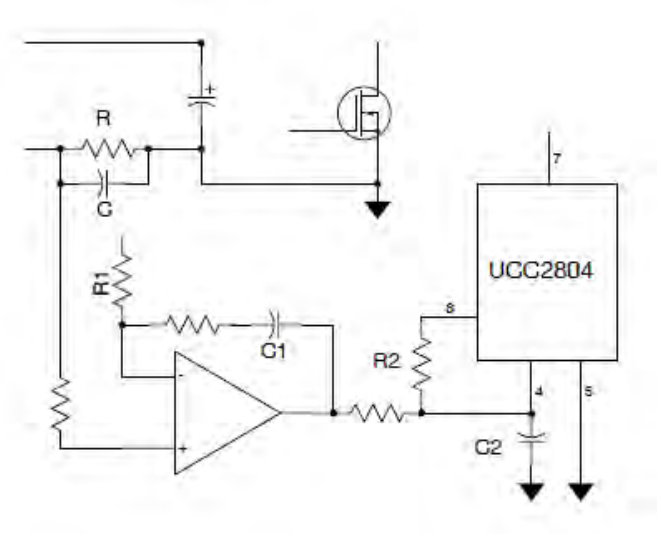


Fig. 6. More detailed circuit

IV. REALIZATION

At the first time flyback converter for RT was built without this automatic feature. Next step was the static test with single resistor and external voltage source. Measurements are taken for different line voltages between 85V and 150V using external voltage to correct the switching frequency in order to maximize the power transfer. The results were very good, so that the complete loop with integrator was built on a separate printed circuit board. Again the measurements are taken, this time only varying the line voltage. Nearly constant feeding current is achieved. The measurements results are given in Table I.

TABLE I
MEASUREMENTS RESULTS

V_L[V]	T[μs]	F[kHz]	I[mA]	P[W]
85	8.7	115	48.8	4.148
90	8.4	119	48.8	4.392
95	8.0	125	48.8	4.636
100	7.6	131	48.8	4.880
105	7.3	137	48.8	5.124
110	7.0	143	48.8	5.368
115	6.8	147	48.9	5.623
120	6.3	159	48.9	5.868
125	6.0	167	48.9	6.112
130	5.7	175	48.9	6.357
135	5.5	182	48.9	6.601
140	5.3	189	48.9	6.846
145	5.2	192	48.9	7.090
150	5.1	196	48.9	7.335

The drain waveforms of the RT flyback converter are given in Figs. 7, 8 and 9.

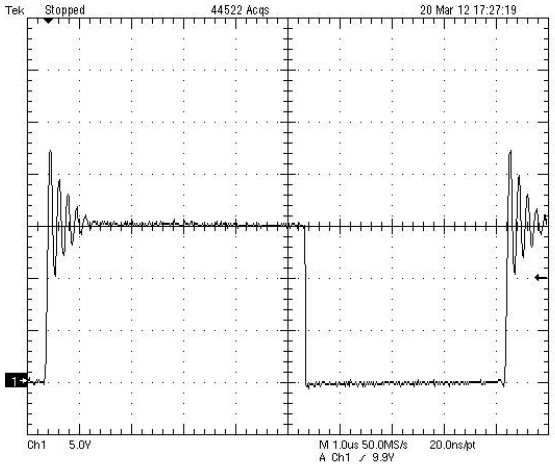


Fig. 7. Drain waveform at 85V

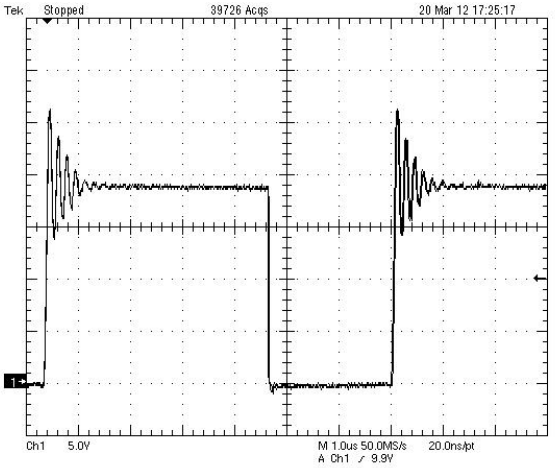


Fig. 8. Drain waveform at 125V

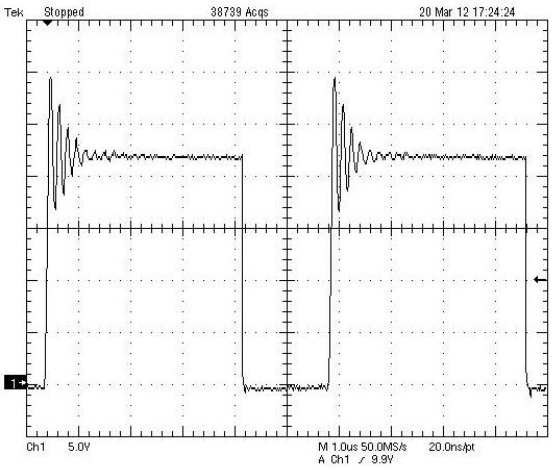


Fig. 9. Drain waveform at 150V

As seen on the waveforms the converter is working in CCM mode. Ringing in the drain waveform of the power switch is not a problem because a maximum drain voltage is well within specifications of used BUZ76A.

All measurements are repeated in a climatic chamber at temperatures between -10º and +55º Celsius.

The picture of the converter prototype is given in the Fig. 10. The size of the board is 100x160mm. Small printed circuit board in the bottom of the picture is a redesigned circuit.

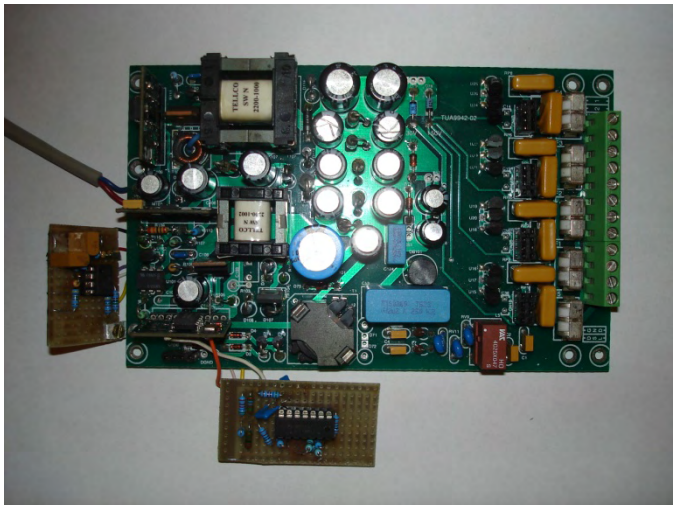


Fig. 10. The converter prototype

V. Conclusion

This paper proposes the simple modification of FLYBACK DC/DC converter in order to maximize the power transfer from Central Office Terminal to Remote Terminal via subscriber line. The topology is simple and easy to adjust. The converter already requires knowledge in loop stability so it is not a problem to compensate new current loop. The prototype was built and successfully tested.

Acknowledgement

The work is partially supported by the Serbian Ministry of Education and Science (Project III-44009).

References

[1] Mindspeed, PCMx Voice Pair Gain (VPG) Framer Bt8954, Data Sheet, www.mindspeed.com
[2] Texas Instruments, "Low Power BiCMOS Current Mode PWM", slus270d Data Sheet, www.ti.com
[3] Texas Instruments, "UCC3800/1/2/3/4/5 BICMOS CURRENT MODE CONTROL ICs", slua149 Application Note, www.ti.com
[4] Texas Instruments, "Under the Hood of Flyback SMPS Design", slup261 Power Supply Design Seminar, www.ti.com
[5] Maxim Integrated Products, "Designing Compact Telecom Power Supplies", AN1062 Application Note, www.maxim-ic.com

Design, Analysis and Modifications of a Telecom Converter

Zoran Zivanovic[1] and Vladimir Smiljakovic[2]

Abstract – **In this paper the straightforward design of a telecom DC/DC converter is presented. The converter was built and tested through lab measurements. Step by step changes are made in order to enhance efficiency. Every step is described and well documented with measurement results.**

Keywords – **DC/DC converter, forward, efficiency**

I. Introduction

The most frequently used DC/DC converter is a single switch forward converter, especially for telecom use with input voltage range from 36 to 72V. It is used in Central Offices, Private Branch Exchanges, Digital Radio Relay Systems, Radio Base Stations etc. Usual power range is from few watts to a couple hundred watts with one or multiple outputs. Very important requirements for these converters are high efficiency and high density. Unlike the flyback converter they have two magnetic components: the transformer and the output inductor.

II. Principle of Forward Converter

The basic forward converter circuit is shown in Fig. 1. During the time when the primary power switch (MOSFET Q) is on, the energy is transferred to secondary. Diode D1 is forward biased and the current flows through inductor L to the capacitor C and the load. When the power switch turns off, diode D1 is reverse biased and forward biased diode D2 provides freewheeling path for inductor current from input and stored in the transformer. Capacitor C acts as a reservoir and holds the output voltage nearly constant.

III. Design and Improvements

The task is to design a 50W telecom forward converter using current mode control IC with careful choice of operating parameters and components. Achieving the efficiency as much as possible near 90% is the primary objective. The footprint size must be smaller than 100x50mm.

First we choose the switching frequency to be around 340 kHz, which is a compromise between the efficiency and size. Knowing that, a good choice of core for the transformer and the inductor is EFD20, N49 material from TDK-EPCOS.

[1]Zoran Zivanovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina165b, 11070 Belgrade, Serbia, E-mail: zoki@insimtel.com.

[2]Vladimir Smiljakovic is with the IMTEL KOMUNIKACIJE AD, Bul. Mihajla Pupina165b, 11070 Belgrade, Serbia, E-mail: smiljac@insimtel.com.

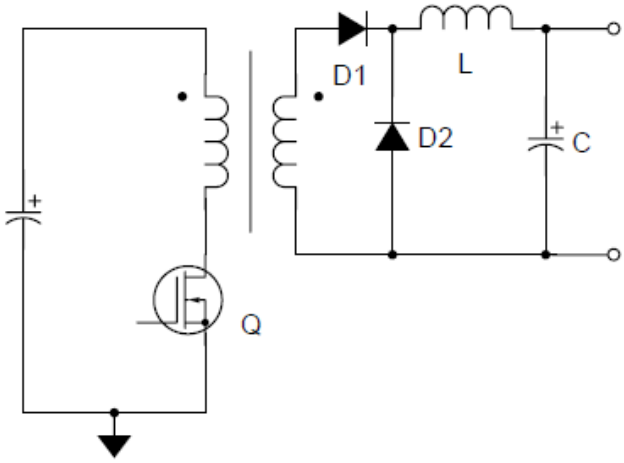


Fig. 1. Forward converter

Design specifications are given in Table I.

TABLE I
DESIGN SPECIFICATIONS

		Min	Typ	Max	
Input voltage	V_{IN}	36	48	72	V
Output voltage	V_O		5		V
Output current	I_O	1.3	10		A
Output current limit	I_{OCL}		12		A
Full load efficiency	η		85		%
Switching frequency	f_{SW}		340		kHz

The main contributors to power losses are transformer, output inductor, power switch, current sensing and secondary rectifiers.

Starting from design specifications we will now calculate basic parameters for the transformer and output inductor (Table II).

TABLE II
BASIC PARAMETERS

		Max	Typ	Min	
Duty cycle	D	0.42	0.31	0.21	
Number of primary turns	N_P		16		
Number of secondary turns	N_S		6		
Primary RMS current	I_{PRMS}	2.64	2.29	1.87	A
Secondary RMS current	I_{SRMS}	7.04	6.10	4.99	A
Output inductance	L		4.75		μH
Number of induct. turns	N		5		

There are several transformer reset methods: Resonant Reset, Active Clamp, RCD Clamp and Third Winding. For simplicity and smaller losses we will use a third winding reset scheme. Now it is time to wind the transformer and output inductor. We will use for the primary 5 parallel strands of 0.22mm copper wire and for the secondary 0.1mm thick copper foil, in order to minimize copper losses. For the output inductor we will use the same foil.

Knowing specific core losses we can now calculate the losses in both magnetic components (Table III). Total transformer power loss at 48V input voltage is 0.94W. This results in approximately 42 °C rise above ambient temperature. For the inductor temperature rise is 23 °C. If we are not satisfied with results, only way to further minimize losses is to change the core geometry.

TABLE III
TRANSFORMER AND INDUCTOR LOSSES

		Max	Typ	Min	
Core effect. volume	V_E		1.46		cm^3
Specific core losses	P_V		0.2		W/cm^3
Primary resistance	R_P		60		mΩ
Secondary resistance	R_S		3.5		mΩ
Core loss	P_{CORE}		0.29		W
Primary loss	P_{PRI}	0.42	0.31	0.21	W
Secondary loss	P_{SEC}	0.17	0.13	0.087	W
Inductor loss	P_{IND}		0.5		W

As a next step we will compare the power losses for two power switches BUZ30A (Table IV) and IRF640 (Table V).

TABLE IV
BUZ30A POWER LOSSES

BUZ30A			Typ		
ON resistance	R_{DS}		0.1		Ω
Reverse transfer capac.	C_{RSS}		130		pF
Conduction loss	P_{CON}	0.69	0.52	0.35	W
Switching loss	P_{SW}	0.31	0.55	1.20	W
Total loss	P_{TOT}	1.00	1.07	1.55	W

TABLE V
IRF640 POWER LOSSES

IRF640			Typ		
ON resistance	R_{DS}		0.1		Ω
Reverse transfer capac.	C_{RSS}		53		pF
Conduction loss	P_{CON}	0.69	0.52	0.35	W
Switching loss	P_{SW}	0.13	0.22	0.49	W
Total loss	P_{TOT}	0.82	0.74	0.84	W

Obviously MOSFET IRF640 is a better choice because of lower switching losses.

For current sensing we can use a current sense resistor or a current transformer. Power dissipated in current sense resistor is given in Table VI.

TABLE VI
CURRENT SENSE RESISTOR POWER LOSSES

		Max	Typ	Min	
Current sense resistance	R_{CS}		0.2		Ω
CS resistance loss	P_{CS}	1.39	1.05	0.7	W

Losses associated with current sense transformer are a little more complex to calculate because dissipation is distributed to the transformer, diode and resistor. The calculation results are given in Table VII.

TABLE VII
CURRENT TRANSFORMER POWER LOSSES

		Max	Typ	Min	
Current sense resistance	R_{CS}		8.2		Ω
Primary resistance	R_{CPRI}		6		mΩ
Secondary resistance	R_{CSEC}		1.14		Ω
Turns ratio	n		40		
CS resistance loss	P_{CST}	36	27	18	mW
Primary loss	P_{CPRI}	42	32	21	mW
Secondary loss	P_{CSEC}	5.8	4.5	2.9	mW
Diode loss	P_D	39	34	28	mW
Total loss	P_{CTOT}	122.8	97.5	69.9	mW

Dissipation of 1.39W on simple current resistor reduce efficiency significantly (near 3%), so that we will use current sensing transformer with dissipation an order of magnitude smaller.

At the end we will compare the power losses for two secondary rectifiers – Schottky diodes MBR2060CT and MBR2535CT (Table VIII).

TABLE VIII
RECTIFIERS POWER LOSSES

		Min	Typ	Max	
MBR2060CT	P_{RECT1}		4		W
MBR2535CT	P_{RECT2}		6.2		W

Comparing to the other losses it is obvious that the choice of secondary rectifier is critical for the converter efficiency. We will use MBR2535CT.

IV. REALISATION

DC/DC converter was built on FR-4 substrate with 70μm copper with footprint 85x40mm. Printed circuit board was mounted on the aluminum cooling plate. The transformer and the output inductor are wounded on thorough hole coil formers according to calculations. Current sensing transformer (of the shelf P8204 from PULSE) is adopted for primary current sensing.

The first step was measuring full load efficiency with BUZ30A and MBR2060CT on board at various input voltages. The results are given in Table IX. The efficiency is around 83%, not so bad for the start.

TABLE IX
EFFICIENCY-FIRST STEP

			Typ		
Input voltage	V_{IN}	36	48	72	V
Input current	I_{IN}	1.657	1.242	0.827	A
Input power	P_{IN}	59.65	59.61	59.54	W
Efficiency	η	83.00	83.00	83.30	%

The next step is to replace BUZ30A with IRF640 and repeat the measurements. The results are given in Table X.

TABLE X
EFFICIENCY-SECOND STEP

			Typ		
Input voltage	V_{IN}	36	48	72	V
Input current	I_{IN}	1.637	1.220	0.813	A
Input power	P_{IN}	58.93	58.56	58.54	W
Efficiency	η	84.00	84.53	84.56	%

Simple change of primary power switch give us the efficiency rise of over 1%. Now we will change the secondary rectifier MBR2060CT with more efficient MBR2535CT. We do the measurements again and the results are in Table XI.

TABLE XI
EFFICIENCY-THIRD STEP

			Typ		
Input voltage	V_{IN}	36	48	72	V
Input current	I_{IN}	1.575	1.170	0.780	A
Input power	P_{IN}	56.70	56.16	56.16	W
Efficiency	η	87.30	88.14	88.14	%

Now we have an efficiency rise of about 3.5% that is expected according to calculations. Full load efficiency is now around 88% - a good result.

Simultaneously with efficiency measurements we have recorded the waveforms at the point of interest.

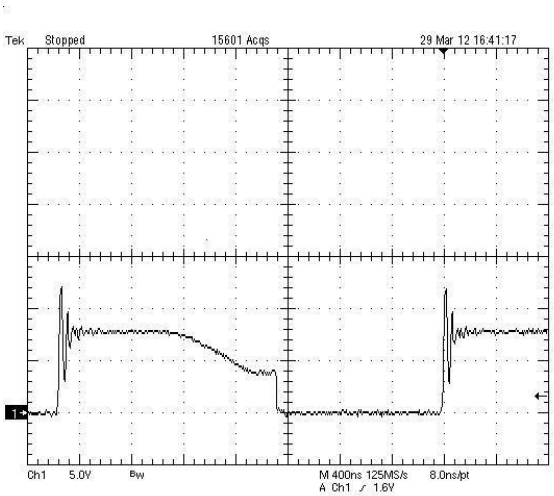


Fig. 2. Drain waveform at 36V

The drain waveforms of primary power switch at full load and input voltages of 36, 48 and 72V are given in Figs. 2, 3 and 4 respectively

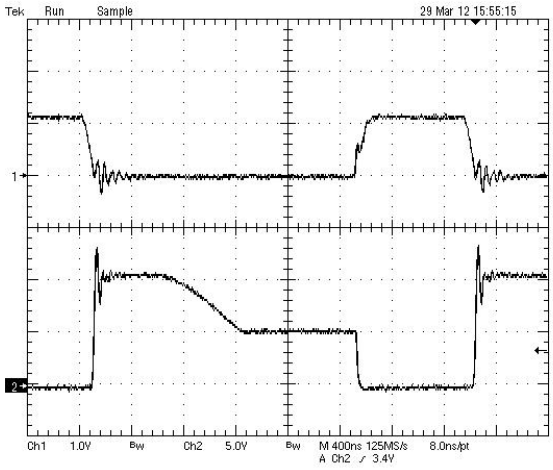


Fig. 3. Drain waveform at 48V with gate waveform above

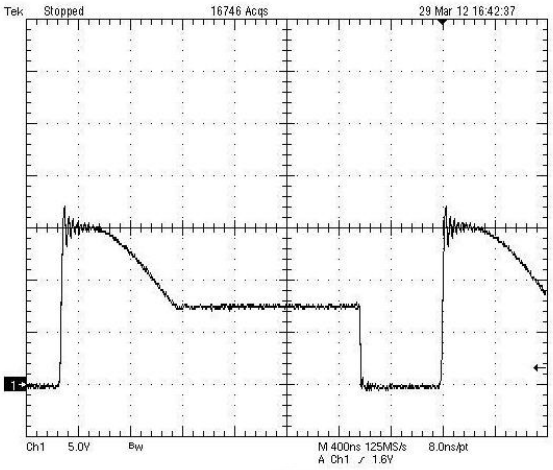


Fig. 4. Drain waveform at 72V

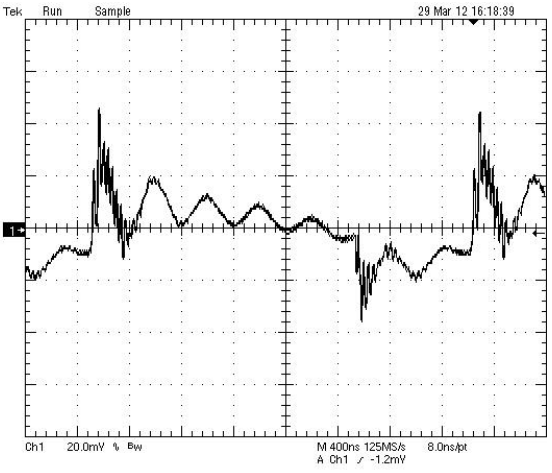


Fig. 5. Output voltage ripple at full load

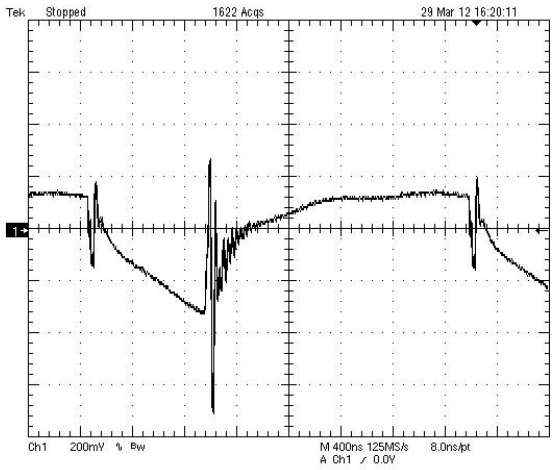


Fig. 6. Input voltage ripple at full load

Output and input ripple voltages are given in Figs. 5 and 6 respectively.

Output voltage rise into full load at 48V input is given in Fig. 7

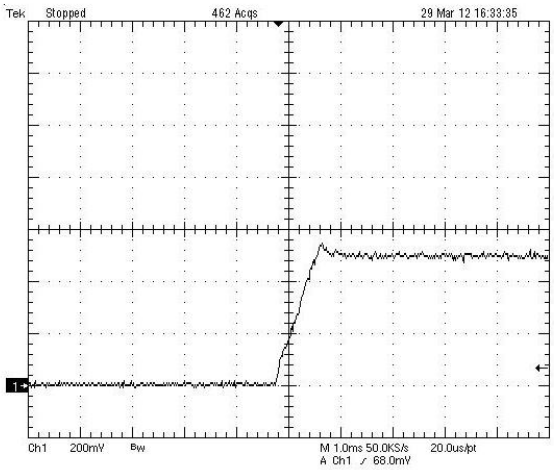


Fig. 7. Output voltage rise into full load at 48V input

The picture of converter prototype is given in Fig.8.

V. Conclusion

In this paper the design and analysis of 50W forward converter are presented.

The prototype was built and tested. The results verified that the full load efficiency is about 88%.

Further improvements are possible thorough Active Clamp Reset with controller change and Synchronous Rectifier in the secondary. The efficiency will go over 90%.

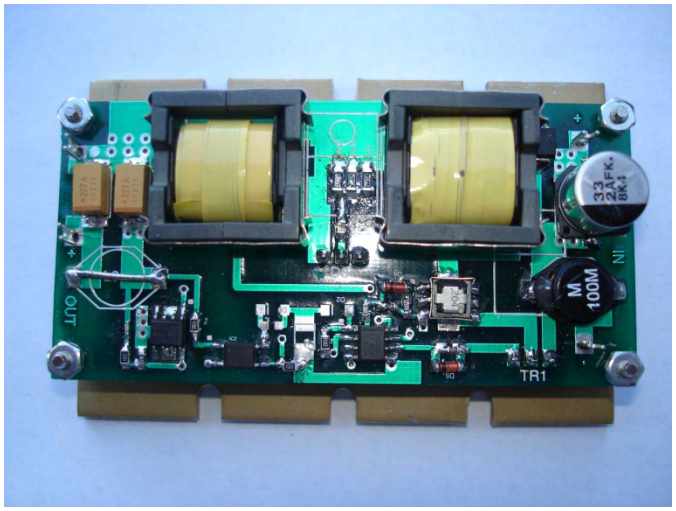


Fig. 8. The converter prototype

Acknowledgement

The work is partially supported by the Serbian Ministry of Education and Science (Project III-45016).

References

[1] Texas Instruments, "Low Power BiCMOS Current Mode PWM", slus270d Data Sheet, www.ti.com

[2] Texas Instruments, "UCC38C42 25W Self-Resonant Reset Forward Converter", SLUA276 Reference design, www.ti.com

[3] TDK-EPCOS, "Ferrites and accessories EFD20/10/7", Data Sheet, www.epcos.com

[4] Vishay Siliconix, "Designing A High-Frequency, Self-Resonant Reset Forward DC/DC For Telecom Using Si9118/9 PWM/PSM Controller", AN70824 Application Note, www.siliconix

[5] Power Integrations, "30W DC-DC Converter with DPA424", EPR21 Engineering Prototype Report, www.powerint.com

Image processing of infrared thermograms for hidden objects

Anna Andonova[1]

Abstract – **In this paper an approach with the flowchart of infrared image processing for hidden objects detection is proposed. The shown results are only referred to the buried mine detection. Because image processing available so far can not detect 100% of hidden subsurface objects it is necessary to combine different technologies to enhance the detection rate.**

Keywords – **Infrared image processing, Thermography, Hidden subsurface object detection**

I. INTRODUCTION

Thermography technique for non-destructive testing is used successfully for the detection of materials discrete defects and the estimation of material properties without causing changes of their performance. Except numerous applications of infrared thermography testing in civilian industry they have found a wide use in rocket, aircraft industry and in military applications. Typical technical objects of non-destructive testing are among other things all types of materials connection (welded, soldered joints etc.) and also constructions and elements made mainly from composite materials.

Hidden object imaging by thermography can be widely used in many application fields, including video surveillance and monitoring, scene understanding, tracking objects and object recognition [1]. However, the major limitation of this approach is that the infrared camera can only get the roughly contour of the object, which is not enough for further image analysis. Infrared detectors are passive sensors that create infrared, or thermal, images without having to expose the subject to any radiation. These images show the heat signature that is given off by objects of interest. Two applications are taken into account buried mine detection or concealed weapon under clothing detection. In the paper are shown only results of mine detection experiments, but analogous approach can be used in the other case, too.

Thermographic inspections produce large amounts of information with infrared images and reports that contain data such as environmental conditions during the inspection, inspected component conditions, data from the operator, etc.

Various methods of data analysis particularly suited in thermography testing have been developed through the World. It is interesting to notice that besides traditional techniques coming from the field of ''computer vision'' imported into specific available camera's software several specific methods have been developed for thermografy testing. These unique techniques are sometimes based on the underlying heat-conduction physics [2]. In the paper, the discussion and implementation of the methods at image processing of infrared thermograms in specific area as hidden objects: buried mines and hidden weapon under clothing detection.

[1]Anna Andonova is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ava@ecad.tu-sofia.bg.

II. EXPERIMENTAL SETUP AND PREPROCESSING

The test minefield was a box of soil (sandy or clay). To measure soil surface temperature, an infrared (IR) camera ThermaCam Sc640 was fixed above the minefield. During the experiment, the measured data were collected for processing. The temperature resolution of the camera is 0.08^0C. Moreover , the data such as air temperature, wind speed and sky irradiance were measured during the experiments. In the experiments are used objects with cylindrical shape equivalent to the shape and dimensions for some type of AP mines with plastic and PVC materials. On the Fig.1 is shown the experimental setup for thermography mine detection.

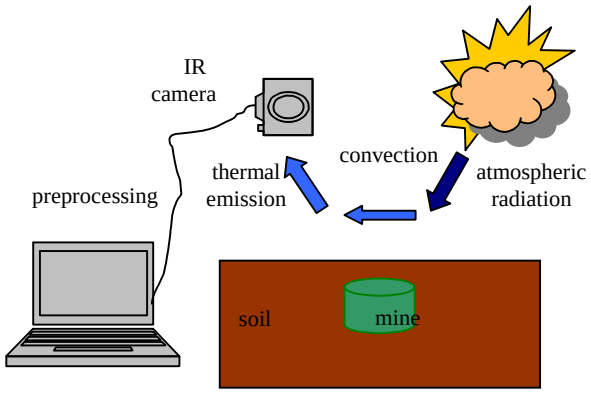


Fig.1.Enperimental setup and different heat transfer processes for thermography mine detection

Before applying the algorithms to the thermograms set, a pre-processing procedure, consisting of radiometric calibration, temporal co-registration, environment correction, apparent temperature conversion and inverse perspective projection was applied to the acquired thermograms sequence.

The previous tests [3] for mine detection have indicated that is not easy to determine accurately the real sizes of different objects buried at different depts. In a single image sequence with usual anomaly detection techniques such as neuron network, mathematical morphology it can even be lose objects whose thermal signature are dominated by others [4].

For example the environment correction is used for estimating the attenuation of the radiated thermal energy through the path from the surface to the camera. After environment (atmospheric) correction, the resulting radiance is an estimation of the radiance originated from the surface. This radiance consists of two components: the radiance due to emission from the objects having a temperature above 0K and the accumulation of all the reflections of the objects in the direction of the camera. Since the examined objects are not blackbodies, then their temperature can still be approximated by so-called apparent temperature. Ground projection maps the IR images onto the surface of the soil producing thermograms (with a resolution of 1 cm^2/pixel) sequence of the soil surface apparent temperature.

The preprocessed thermograms used in the following represent the soil surface temperature.

III. Nondestructive Inspection

The solution of the forward problem F is the solution of the Heat equation. Lets y is measured data and p is the original distribution of parameters that gives rise to y under the application of operator F. In our case p corresponds to the value of the thermal diffusivity at every point within the soil volume α. So the functional equation involving a map F, which represent the connection between the model and data is

$$Y = F[p] \quad (1)$$

where $F[p]$ is the distribution of the temperatures of the surface of the soil.

An inverse problem of this will be the reconstruction of the original distribution of parameters based on measurements of the resulting data. Solving an inverse problem implies approximating the best solution

$$p = F^{-1}[y]. \quad (2)$$

In general y is never known exactly but up to an error of $\delta \neq 0$. It is assumed that we know $\delta > 0$ and

$$\left\| y - y^{\delta} \right\| \leq \delta \quad (3)$$

So y^{δ} is the noisy data and δ is the noise level. To compute numerically a solution of the problem it can be used the regularization techniques to restore stability and existence of the solution and develop efficient algorithms.

IV. Target Detection

Using thermal infrared images to detect subsurface hidden objects is based on the assumption that these objects have different thermal properties than their surrounding area. When an area is heated, the hidden objects will warm up faster or slower, because of the thermal properties of the materials. The influence of the hidden object on the temperature at surface level relates to the distance from the surface. This situation, where hidden objects give relatively warm spots at the surface, is called a situation with positive contrast. For this reason, and under natural heating conditions, sunrise and sunset are the times of the day at which this thermal contrast is largest.

It is assumed that a pre-processing stage is run on a conventional PC in order to align the images and map grayscale colors to temperature values on the surface. Next, the soil inspection procedure itself starts. First, a detection procedure is run to obtain the mask of potential targets. Then, a quasi inverse process operator is used to identify the presence of mines among the potential targets. For those targets that failed to be classified as mines (and are therefore labeled as unknown), a full inverse procedure to extract their thermal diffusivity will be run in order to gain information about their nature. The overall detection process is summarized in Fig. 2, where the processes that require the use of the 3D thermal model are indicated with an ellipse. The detection, quasiinverse and fullinverse procedures are based on the solution of the heat equation for different soil configurations. As explained, this is a very time consuming task that makes the whole algorithm inefficient for real on-field applications.

The use of IR cameras taking images of the soil under inspection gives the exact distribution of temperatures on the surface. On the other hand, the thermal model described previously and extensively validated with experimental data permits to predict the thermal signature of the soil under given conditions. The detection of the presence of potential targets

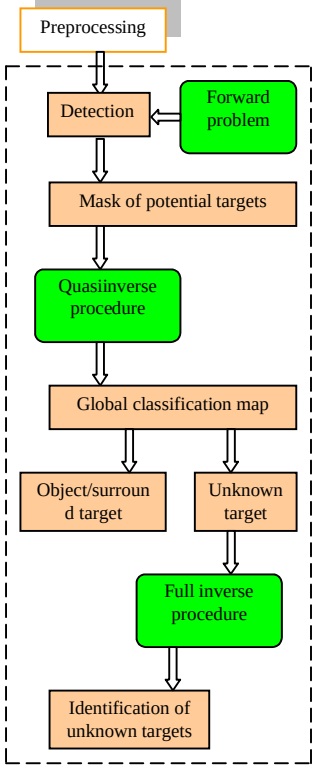
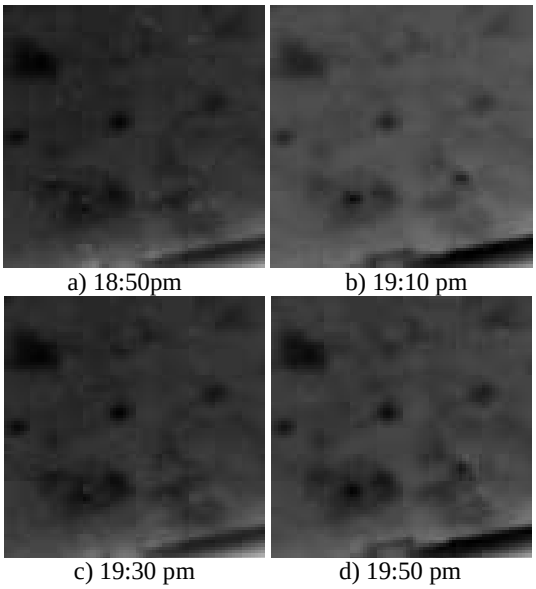


Fig. 2. Structure of the approach to detect hidden objects

on the soil is then made by comparing the measured thermograms with the expected thermal behavior of the soil given by the solution of the forward problem under the assumption of absence of mines on the field.

The surface positions (x, y) are determined where the behavior is different from that expected under the assumption of mine absence, therefore revealing the presence of unexpected objects on the soil. These positions will be classified as potential targets, whereas the rest of the pixels (those that follow the expected pure-soil behavior) will be automatically classified as soil. This process is not trivial.

The most straightforward approach, the threshold detection, has the drawback of setting the threshold, which will vary not only for different image sequences, but it is also likely to depend on the particular frame of the sequence, and on the characteristics of the measured data such as lighting conditions and the nature and duration of the heating. For this reason, the use of a reconfigurable structure, capable of adapting to varying experimental conditions was proposed on [5]. In this work it is demonstrated that it is possible to reduce the time frame of analysis when the maximum thermal contrast at the surface is expected. This phenomenon can be better appreciated in Fig. 3, where a sequence of IR images of a mine field taken between 18:50 am and 19:50 pm is shown. Taking into account the short time interval we can consider that the properties of the soil remain unaltered and that there is no mass transference process during the simulation. The output of the detection stage is a black and white image with the mask of the potential targets.



a) 18:50pm b) 19:10 pm

c) 19:30 pm d) 19:50 pm

Fig. 3. Example of measured thermograms of minefield at sunset

V. Conclusion

Two techniques to reduce the computing time can be used. The first one is an FPGA implementation of the thermal model that speeds up the computations making the system suitable for field work applications. A scheme of the algorithm where the FPGA system can be realized is proposed. The processes makes use of the thermal model, running on the FPGA, are indicated with dashed points. The second approach consists on using non-uniform grids to reduce the number of nodes involved in the computations and therefore, the computing time. The FPGA implementations will be fully realized in recent and presented in some other report.

Acknowledgement

This work was supported by National Ministry of Science and Education of Bulgaria under Contract DDVU 02/4-7: “Thermo Vision Methods and Recourses in Information Systems for Customs Control and Combating Terrorism Aimed at Detecting and Tracking Objects and People”.

References

[1] N. Thanh, “Infrared thermography for the detection and characterization of buried objects”, PhD thesis, Vrije Universiteit Brussel, Belgium, 2007.

[2] P. Lopez, “Detection of landmines from measured infrared images using thermal modeling of the soil”, PhD thesis, University Santiago de Compostela, 2003.

[3] A Andonova, S. Todorov, “Buried object detection by thermography”, Annual Journal of Electronics, vol.4, no 1, pp. 133-136, 2010.

[4] S. SjÄokvist, M. Georgson, S. Ringberg, M. Uppsall, and D. Loyd, “ Thermal efects on solar radiated sand surfaces containing landmines - a heat transfer analysis” Computational Mechanics Publications, Cracow, Poland, 17-19 June, pp. 177-187, 1998. Computational Mechanics Publications.

[5] P. Lopez, L. van Kempen, H. Sahli, D.Cabello, “Improve thermal analysis of buried landmines” IEEE Trans., Geoscience and Remote Sensing, vol. 42, no. 9, pp. 1965-1975, 2004.

Study and analysis of optimization approaches for insulation of an industrial grade furnace with electrical resistance heaters

Borislav Dimitrov[1], Hristofor Tahrilov[2], Georgi Nikolov[3]

Abstract – **The paper discusses several approaches for optimal sizing of a multilayer insulation of an industrial grade furnace with electrical resistance heaters. The analysis is based on an iterative mathematical procedure, that allows simultaneous test of the thermal processes and the design on the device based of an initially set limitations and a dedicated function.**

Keywords – **Energy efficiency, Mathematical methodology, Optimization, electric heating.**

I. INTRODUCTION

Industrial furnaces with electrical resistance heaters are designed for thermal treatment of steal parts – hardening, tempering and others. They are powerful electric loads, and all the efforts about increasing their efficiency are major concern [4, 7].

The mathematical methods for optimization and their application in electro technological apparatuses are discussed in different papers [1, 2, 8]. The possibilities for optimization of multilayer isolation of electric heater furnace (EHF) are considered a special case which makes the description incomplete.

The main goal of the following paper is to summarize the possibilities and the features of the models with concentrated parameters in the research and development of thermal devices with optimal construction, according to predefined constraints or objective function.

The suggested approach for investigation is based on mathematical model [3,4,5,6], realized by a system of differential equation, solved by numeric methods in MATLAB.

II. ANALYSIS

Using mathematical models [3,4,5,6] allows significant expanding of the possibilities for analysis of the processes taking places inside a EHF.

The suggested model [3] is used to optimize the parameters of the isolation. It allows taking into account the specific particularities of the technological process independently of the construction. The process is defined by heating speed, allowed temperature difference according to the thickness of the part and other parameters.

It is also possible to enter the custom operation modes of the EHF and investigate the best ranges in which the optimization criteria are fulfilled with less than the maximum error.

The construction of the EHF sets the optimization procedure as multivariate with different objective function for each case. As a result of this, the scanning method is used, and the constraints are used to choose the construction type.

Major factor, that determines the energy efficiency of EHF are the losses. The following cases for objective function with the set of constraints are examined:

- *Objective function* – minimal heat losses. *Set of constraints* – allowed weight and volume. *Goal* – to find a construction of EHF with given volume and/or weight between two values, for which the losses are minimal.
- *Objective function* – minimization of weight and volume. *Set of constraints* – allowed losses. *Goal* – to find a construction of EHF with smallest external volume for which the losses are in the allowed range.
- *Objective function* – minimal price. *Set of constraints* – allowed volume, mass, losses. *Goal* –to determine the heat isolation materials for the given price range. It is necessary to obtain an up-to date prices for the materials.

With all of the three optimization tasks, the working environment and the load are unchanged. Realization of these tasks is than simplified to find: the minimal losses – for stationary furnaces working in continuous mode (industrial, heavy duty furnaces and so on); the minimum mass and volume for furnaces that are not so often used and are not major equipment – mobile furnaces placed on a platform, laboratory furnaces and others.

Several scanning method approaches are formulated:

- *Calculation of the heat isolation (walls) of the EHF*. A model is created, for which the selected materials are unchanged. Than the thickness of the isolation is recalculated according to the objective function and constraints. This approach is used with all objective functions, but because the type of the material is fixed it is not flexible enough.
- *Selection of isolation materials for EHF*. A model is created with fixed wall thickness (constant volume). Then

[1] eng. Borislav Dimitrov, Ph.D– Technical University of Varna, Bulgaria, assistant. E-mail – bdimitrov@processmodeling.org
[2] eng. Hristofor Tahrilov, Ph.D – Technical University of Varna, Bulgaria, assoc. professor. E-mail – h.tahrilov@gmail.com
[3] eng. Georgi Nikolov, Ph.D – Technical University of Varna, Bulgaria, assistant. E-mail – gtn@gbg.bg

fireproof heat isolation materials are selected from a given database. Possible objective function is minimization of mass and or losses. This requires full information for the database.

- *Calculation of the wall thickness and selection of the material.* A model is created that allows simultaneous work of the former two approaches – variable wall thickness and material. This approach allows great flexibility and gives a multivariate solution to the problem.

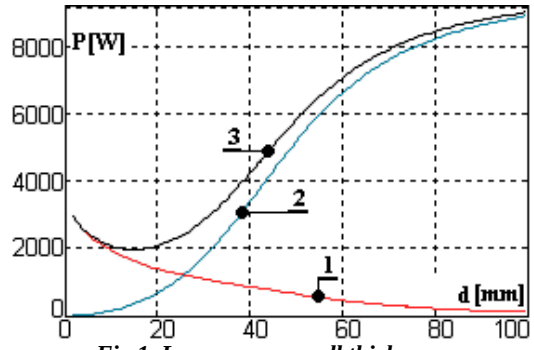


Fig.1. Losses versus wall thickness.
1 – losses to the environment; 2 – energy accumulated in the wall; 3 – total losses

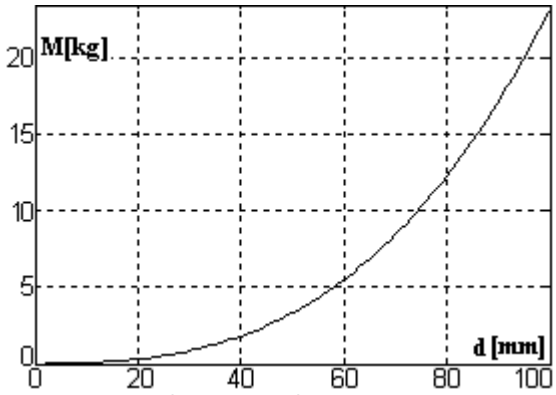


Fig.2. Mass of the wall as function to its thickness.

The calculation process is iterative and realized by changing the wall thickness with predefined step. An EHF is investigated with objective function loss minimization. The set of constraints are minimum wall thickness *gD1* = 10 *mm* and maximum *gD2* = 40*mm*. These are standard dimensions of isolation materials, supplied by various manufacturers. In order to obtained more detailed analysis the range for the wall thickness is expanded from 3mm to 100mm. The obtained results are graphically presented as follows:

- *Fig.1.* Minimum losses are calculated at wall thickness $gD = 18$ *mm*, fulfilling the original constraints: $gD1 < 18 < gD2$. The same approach is used for the mass constraints. Line 1 shows the decrease of the heat flow from the furnace to the environment with increase of the wall thickness. Simultaneously the accumulated energy inside the isolation is increased as shown by line 2. The total losses are increased as a result.
- *Fig.2.* The change of the wall mass is according to the density of the used materials and their volume. In this case the graph shows the mass of a single wall of a furnace that has a total of six walls, but the same approach can be used to determine the total mass of the device.
- *Fig.3.* The distribution of the temperature inside the wall, allows coordinated heat loading of the materials. The presented results are for the middle of the wall, with the transient process completed.
- *Fig.4.* The model gives information about the accumulated energy inside the walls of the furnace and the heat losses to the environment. They are similar to the graphs presented in fig. 1.

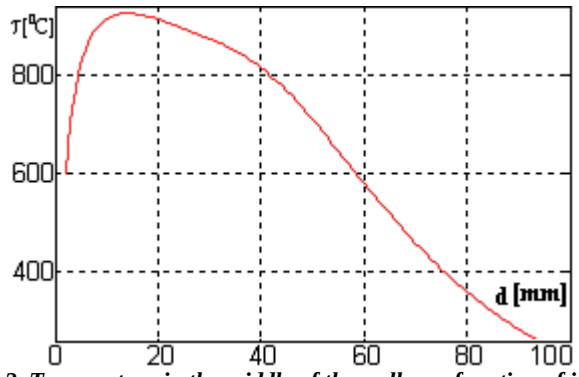


Fig.3. Temperature in the middle of the wall as a function of its thickness.

The optimization task is more complicated when multilayer isolation is used. This case also requires investigation, as the walls of medium and high temperature furnaces are built in exactly this way. The presented example has two layers of isolation, each having thickness from 5 mm to 100 mm. The results are presented below.

- *Fig.5.* presents the losses versus wall thickness. The behavior of graphs 3 and 4 is determined by thermal characteristics of the materials λ and *c*, which are different for heatproof materials (layer 1) and thermal isolation material (layer 2).
- *Fig.6.* presents the mass of the furnace under test as a function of the wall thickness. The maximum weight is 221,6 kg;

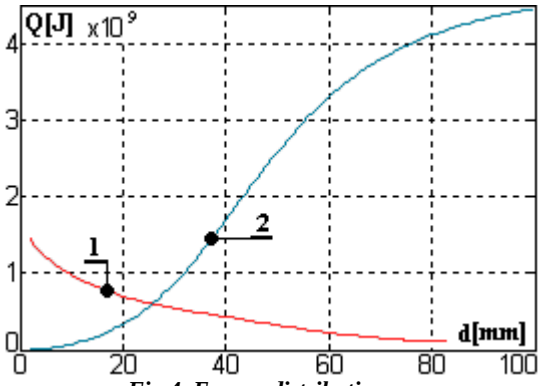


Fig.4. Energy distribution.
1 – energy to the environment;
2 – energy accumulated in the wall;

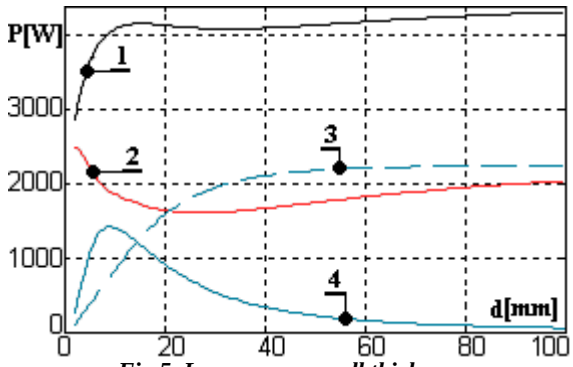


Fig.5. Losses versus wall thickness
1 – total losses; 2 – losses in the environment;
3, 4 – accumulated in layer 1 and 2 of the isolation;

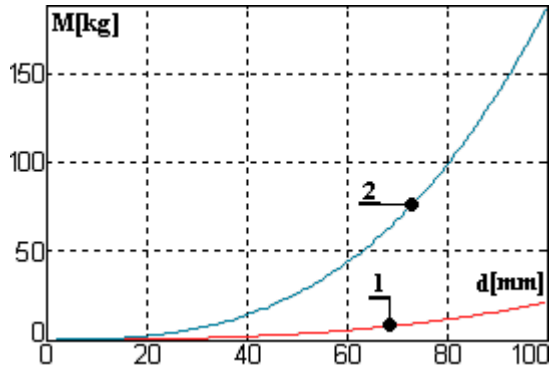


Fig.6. Mass of the layers. 1,2 – for layer 1 and 2;

- *Fig.7.* and *Fig.8.* show the same process but the thickness of the first layer (fireproof) is increased to 10 mm, and 100 mm for the second (heat isolation) layer. The results shows that the total losses (fig.7 graph 1) are considerably increased compared to the previous case (fig. 5 graph 1), while the mass of the furnace is reduced to 32,92 kg. This shows that the obtained results can be used to solve an optimization problem for the mass and the volume of the furnace. This is the problem to be solved for most small volume laboratory furnaces, which are not used in continuous manufacturing process.
- *Fig.9.* and *Fig.10* present the transient process when heating a component, for both the initial, and the final dimensions for both layers. According to the data, when the wall thickness is increased, the temperature inside the furnace is increased, as the supplied power is constant. The presented processes are the final result in the designed apparatus. They give information about the adaptation of the EHF for realization of the required technological process.

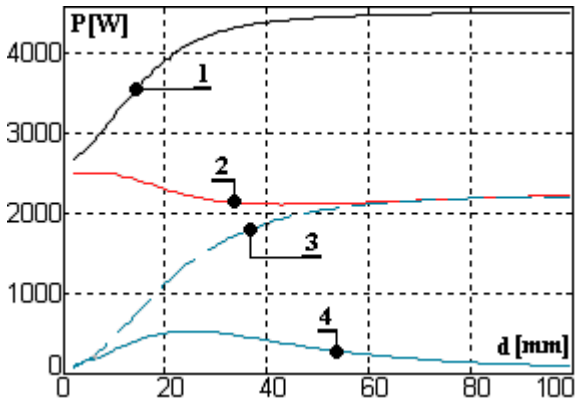


Fig.7. Losses versus wall thickness
1 – total losses; 2 – losses in the environment;
3, 4 – accumulated in layer 1 and 2 of the isolation;

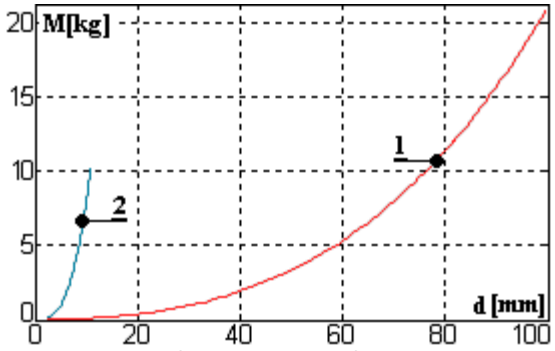


Fig.8. Mass of the layers. 1,2 – for layer 1 and 2;

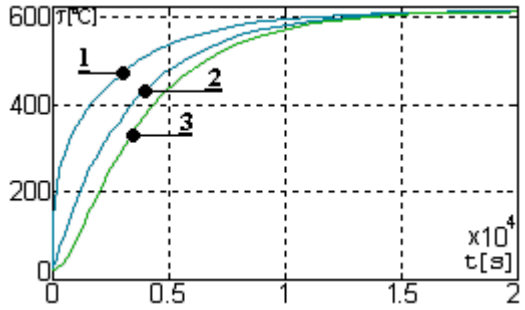


Fig.9. Transient process when heating a component, wall thickness 5 mm. 1 – heater temperature;
2,3 – temperature on the surface and in the middle of the component.

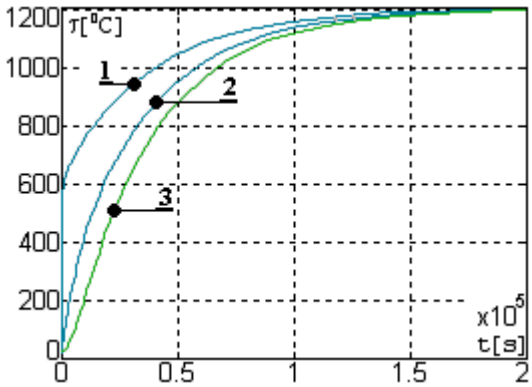


Fig.10. Transient process when heating a component, wall thickness 100 mm. 1 – heater temperature;
2,3 – temperature on the surface and in the middle of the component

III. CONCLUSION

The results show different possibilities and applications of the suggested approach. The following advantages are observed:

- Full information, about the processes in the EHF is obtained. This allows evaluation of the influence of the corresponding factors on the furnace characteristics.
- Obtaining the dimensions of the isolation layers, with predefined constrains is in conjunction with the optimization requirements respectively creating equipment that fulfills some operation conditions.
- The mathematical models [3,4,5,6] should be used in the process of designing new equipment or reconstruction of existing one.

ACKNOWLEDGEMENT

This paper is developed in the frames of project "Improving energy efficiency and optimization of electro technological processes and devices", № МУ03/163 financed by the National Science Fund

REFERENCES

[1] Ashlock D. "Evolutionary computation for modeling and optimization" 2006
[2] Chong E, Zak St. " An introduction to optimization", 2001.
[3] Димитров Б. "Анализ на изчислителния процес за математичен модел на съпротивителни електропещи", Списание Електротехника и електроника Е+Е 11-12 2006, страници 78-83.
[4] Димитров Б. "Анализ и моделиране на електротермични процеси и устройства /електросъпротивителни пещи с периодично действие/. Дисертация ТУ-Варна 2009.
[5] Tahrilov H., Dimitrov B., "The analysis of the processes in the electric resistance furnace with the general model", Tenth international conference on electrical machines, drives and power systems ELMA 2002, page 75-80
[6] Tahrilov H., Dimitrov B. "The general model for researching work of electric resistance furnace", Tenth international conference on electrical machines, drives and power systems ELMA 2002, page 81-86
[7] Йорданова М. Й., "Изследване върху режимите на работа и конструктивните параметри на камерни електросъпротивителни пещи" Дисертация ТУ-Варна 1996 год.
[8] Rao Singiresu, "Engineering optimization. Theory and practice" 2009.

Improving energy efficiency of industrial grade furnaces with electrical resistance heaters and comparative model-experiment analysis

Borislav Dimitrov[1], Hristofor Tahrilov[2], Angel Marinov[3]

Abstract – **The current paper suggests a methodology for optimization of insulation for industrial grade furnaces with electric resistance heaters, by coordinated selection of insulating materials. The methodology is based on an objective function for which – minimal losses, mass and size are selected. The results from the simulation procedure are compared against the data from experimental prototype, in order to evaluate the functionality of the model and suggested methodology.**

Keywords – **Energy efficiency ,Mathematical methodology, Optimization, electric heating**

I. INTRODUCTION

The process analysis for industrial grade furnaces based on electrical resistance heaters (IGF), done with dedicated models and numerical methods, allows for their optimization, using different objective functions [5]. In order to obtain valid results, the analysis has to be conducted using specified models, with parameters set as close as possible to those of the IGF device.

The aim of the current paper is to suggest a numerical optimization method [1,2,3,4] and apply it with a specialized simulation procedure. In order to be validated the results obtained through the simulation - using a model will be compared against experimental data. The experimental data describes a transient process of preheating the camera of the IGF. The data is obtained from a device with parameters corresponding to those set in the model.

For those specific aims, the mathematical model [1] will be used for optimization of the thermal insulation and the refractory layer of the furnace. The results will be applied to a practical construction that will be tested in order to acquire the experimental dataset.

II. ANALYSIS

The analysis of the set problem can be conducted in the following order:

1. A database of refractory layers and thermal insulations is created. The database is then sorted based on the physical parameters of each representative, more specifically: coefficient of heat transfer, specific heat and weight, price, etc.
2. The geometrical size of the working area of the furnace is determined – this is done when the volume and the shape of the load are taken in account.
3. An initial calculation of the construction is carried out. The initial calculation is used in order to determine the number of required layers of insulation, as well as their width.
4. A model of the furnace is created [2]. The model is evaluated on an iterative basis, using the previously prepared database. The iterations are based on the compatibility of the materials: first layer - refractory, second layer - thermal resistance
5. The analysis of the data allows to select a combination of materials that is closest to the objective function.

The object of the study is the chamber of the IGF, which walls are made of two layers – refractory and thermo resistive. Thus the objective function can be for:

- Minimal weight of the materials for the walls:

$$\begin{aligned} G &= G_1 + G_2 = \\ &= \gamma_1.\delta_1^3 + \gamma_2.\delta_2^3 + 3.\gamma_2.\delta_1.\delta_2.(\delta_1+\delta_2) + \\ &+ (a_1+b_1+c_1).(\gamma_1.\delta_1+\gamma_2.\delta_2)^2 + \\ &+ (a_1.b_1+b_1.c_1+a_1.c_1).(\gamma_1.\delta_1+\gamma_2.\delta_2) \end{aligned} \quad (1)$$

- Minimal losses for the working time - t_{work}, based of the thermal conductance λ of the materials – $q_{heat}=f(\lambda)$. This can also optionally include the energy accumulated in the walls of the furnace - Q_{ak}, determined by the specific heat C of each layer: $Q_{heat} = \Sigma q_{heat}.\Delta t + Q_{ak}$, where:

$$q_{heat} = \frac{\tau_1 - \tau_{ок.ср.}}{R_\Sigma} \quad (2)$$

$$R_\Sigma = \frac{\delta_1}{\lambda_1.S_{avg1}} + \frac{\delta_2}{\lambda_2.S_{avg2}} + \frac{1}{\alpha_{kl}.S_3} \quad (3)$$

$$\begin{aligned} Q_{ak} &= Q_{ak1} + Q_{ak2} = C_1.\rho_1.\delta_1^3 + C_2.\rho_2.\delta_2^3 + \\ &+ 3.C_2.\rho_2.\delta_1.\delta_2.(\delta_1+\delta_2) + \\ &+ (a_1+b_1+c_1).(C_1.\rho_1.\delta_1 + C_2.\rho_2.\delta_2)^2 + \\ &+ (a_1.b_1+b_1.c_1+a_1.c_1).(C_1.\rho_1.\delta_1 + C_2.\rho_2.\delta_2) \end{aligned} \quad (4)$$

[1] eng. Borislav Dimitrov, Ph.D– Technical University of Varna, Bulgaria, assoc professor E-mail – bdimitrov@processmodeling.org
[2] eng. Hristofor Tahrilov, Ph.D – Technical University of Varna, Bulgaria, assoc. professor. E-mail: h.tahrilov@gmail.com
[3] eng. Angel Marinov, – Technical University of Varna, Bulgaria, assistant professor E-mail – igdrazil@abv.bg

Geometrical sizes *a*, *b*, *c*, the width δ_1, δ_2 and their distribution are presented at figure 1.

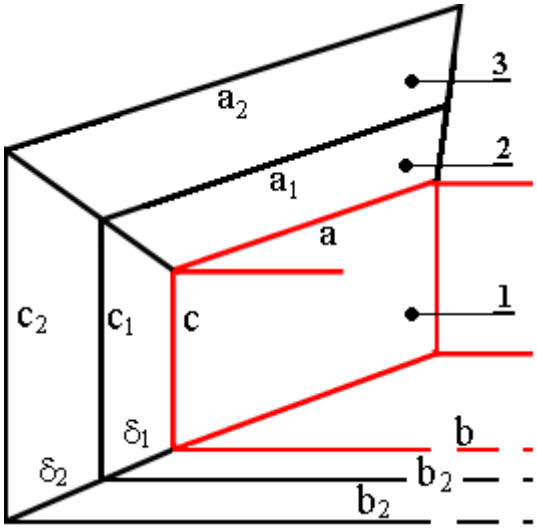


Fig.1. Geometrical sizes a, b, c and widht δ1, δ2 of the walls of IGF:
1 – chamber of the furnace; 2 – refractory layer;
3 – Thermal resistance layer

The optimization procedure is realized using the scan method [5] that allows minimization of the objective function for the given range. [$q_{heat,}$, Q_{ak}]

The limitation conditions are based on the modes of operation that determine different distribution of the two components of the losses.

The database created for the model consists of different materials that are used in the construction of industrial furnaces. The materials are produced and distributed by different companies, thus due to these commercial specifics the materials are cited in table 1 only with their parameters.

The results obtained through the model are presented in graphical and numerical form as follows:

- Table 2 and figure 2 – present the losses dissipated to the surrounding environment when refractory and thermal insulating materials are used. Minimal values are obtained when materials with reduced thermal conduction coefficient are used. The selection of the required combination can be determined based on – weight and price.
- Table 3 and figure 3 – efficiency for one heating cycle: heating and pause during 1 hour.

When the objective function concerns minimal mass with limitation the geometrical size of the furnace, the optimization algorithm remains the same. In this case the database from table 1 is sorted based of the density of the material. The obtained results are as follows:

- Table 4 and figure 4 - present the losses dissipated to the surrounding environment. Based on the data for minimal mass and minimal losses - refractory material number four can be selected from table 5.
- Figure 5 shows the results from a mathematical model (curve 1) and experimental study (curve 2) of the temperature in the transient process of heating. The experimental measurements are made based of a low temperature furnace. The experimental prototype is made of insulation using material number 5 from table 1. Further data for the experimental prototype, its technical parameters as well as additional experimental results are presented in [6].

III. Conclusion

The suggested methodology of optimization of IGF insulation based on mathematical model, allows obtaining notable results. The same method can be used in construction and design of new equipment, as well as in reconstruction for energy efficiency of current equipment.

The difference between results obtained from the model and the experiment, varies between 3% and 7%. Accuracy can be improved by précising the initial parameters set in the model trough series of experiments. This methodology allows simulations to be used in the exploitation of the furnace.

There are practically no limitations to the model as long as the correct objective function is given, and a database for the given elements exists.

The obtained results presented in table and graphical form allow for a continuous correction, limiting the possibility for error.

Acknowledgment

This paper is prepared in the frames of Project MU03/163 “Increasing energy efficiency and optimization for electrotechnological processes and devices”, Ministry of Education Youth and Science, Bulgarian National Science Fund.

References

[1] Димитров Б. “Анализ на изчислителния процес за математичен модел на съпротивителни електропещи”, Списание Електротехника и електроника E+E 11-12 2006 страници 78-83.

[2] Димитров Б. “Анализ и моделиране на електротермични процеси и устройства /електросъпротивителни пещи с периодично действие/. Дисертация ТУ-Варна 2009.

[3] Tahrilov H., Dimitrov B., “The analysis of the processes in the electric resistance furnace with the general model”, Tenth international conference on electrical machines, drives and power systems ELMA 2002 page 75-80

[4] Tahrilov H., Dimitrov B. “The general model for researching work of electric resistance furnace”, Tenth international conference on electrical machines, drives and power systems ELMA 2002 page 81-86

[5] Rao Singiresu, “Engineering optimization. Theory and practice” 2009.

[6] http://processmodeling.org

TABLE I
REFRACTORY (1) AND THERMAL INSULATING (2) MATERIALS

Material Number	1		2		3		4		5		6		7		8		9		10	
Type	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Coefficient of thermal conductance λ[W/m.K]	0,6	0,1	0,65	0,15	0,7	0,2	0,75	0,25	0,8	0,3	0,85	0,35	0,9	0,4	0,95	0,45	1	0,5	1,05	0,55
Specific heat C [J/kg.K]	400	300	850	450	500	240	350	190	770	270	650	550	230	660	715	333	908	130	381	381
Density γ[kg/m^3]	1000	300	800	100	1200	200	1350	350	1700	700	950	250	1230	310	1015	115	800	180	2000	420

TABLE II
POWER LOSSES DISSIPATED TO THE ENVIRONMENT (W)WITH MATERIAL COMBINATIONS

		1	2	3	4	5	6	7	8	9	10
		Refractory material									
1	Thermal insulation materials	4796,3	4807,97	4897,66	5009,62	4883,63	5036,72	5295,88	5065,67	5084,91	5091,77
2		6094,05	6139,31	6283,28	6471,96	6352,71	6603,88	6947,43	6702,28	6755,16	6792,92
3		7023,32	7116,97	7317,88	7577,72	7465,37	7795,22	8227,6	7970,64	8063,23	8129,09
4		7733,89	7844,07	8123,58	8437,19	8340,9	8742,2	9252,06	8933,95	9119,14	9216,43
5		8296,32	8435,65	8763,38	9136,23	9050,87	9513,59	10075,5	9764,35	9991,88	10120,2
6		8694,06	8915,72	9252,11	9642,67	9588,05	10080,3	10709,4	10453,1	10638,5	10792,1
7		9064,59	9311,51	9685,07	10111,6	10072,8	10607,6	11284,1	11036,5	11251,6	11430,3
8		9374,83	9648,69	10085,2	10586,9	10536,7	11067	11797,2	11547,9	11788	11988,9
9		9634,94	9933,32	10404,4	10936	10895,5	11456	12217,2	11984,1	12249,3	12471,4
10		9859,26	10178,1	10628,4	11151,6	11150,5	11794,3	12585,1	12364,1	12646	12889,2

TABLE III
EFFICIENCY FOR COMBINATION OF DIFFERENT MATERIALS

		1	2	3	4	5	6	7	8	9	10
		Refractory material									
1	Thermal insulation materials	0,8566	0,8356	0,8296	0,828	0,796	0,8049	0,8309	0,7852	0,7781	0,7697
2		0,8608	0,8384	0,8329	0,832	0,7984	0,8087	0,8379	0,7888	0,7818	0,7733
3		0,8628	0,8399	0,8347	0,8348	0,799	0,8107	0,8411	0,7907	0,7838	0,7753
4		0,8639	0,8407	0,835	0,8360	0,8001	0,8119	0,8428	0,7919	0,7850	0,7765
5		0,8663	0,8425	0,8378	0,8387	0,8013	0,8143	0,8472	0,7943	0,7875	0,7790
6		0,8641	0,841	0,8361	0,8365	0,8005	0,8125	0,8429	0,7926	0,7858	0,7773
7		0,8632	0,8407	0,8355	0,8356	0,8002	0,8119	0,8410	0,7921	0,7852	0,7768
8		0,8703	0,8455	0,8412	0,8430	0,8034	0,8186	0,8543	0,7985	0,7919	0,7834
9		0,8717	0,8466	0,8427	0,8449	0,8042	0,8201	0,8570	0,8000	0,7936	0,7852
10		0,8674	0,8437	0,8390	0,8404	0,8024	0,8163	0,8489	0,7964	0,7897	0,7813

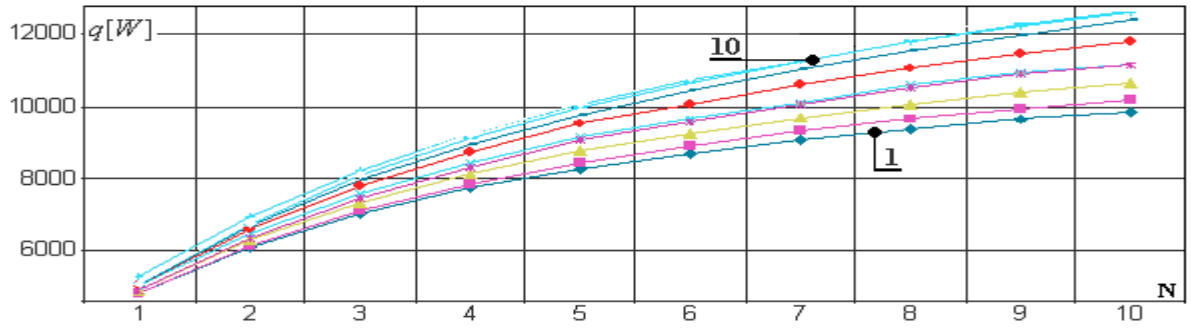


Fig.2. Power losses dissipated to the environment using data from table 2

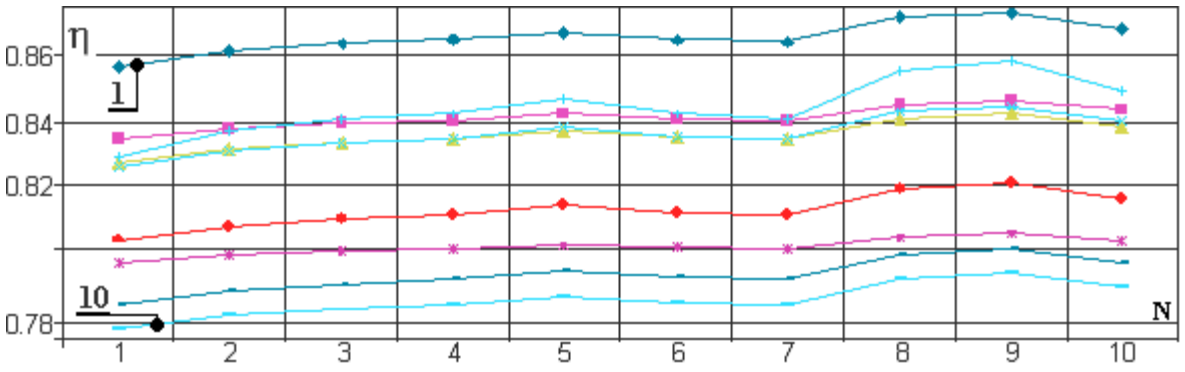


Fig.3. Efficiency using data from table 2

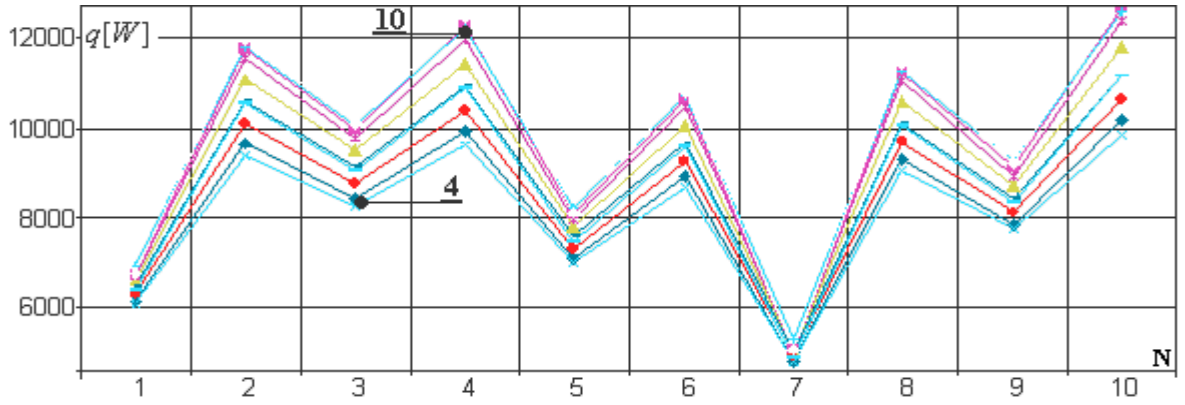


Fig.4. Power losses dissipated to the environment using data from table 4

TABLE IV
POWER LOSSES DISSIPATED TO THE ENVIRONMENT (W)WITH DIFFERENT MATERIAL COMBINATIONS

		1	2	3	4	5	6	7	8	9	10
		Refractory material									
1	Thermal insulation materials	6139,31	6755.16	6603,88	6094,05	6702,28	6283,28	6947,43	6471,96	6352,71	6792,92
2		9648,69	11788	11067	9374,83	11547,9	10085,2	11797,2	10586,9	10536,7	11988,9
3		8435,65	9991,88	9513,59	8296,32	9764,35	8763,38	10075,5	9136,23	9050,87	10120,2
4		9933,32	12249,3	11456	9634,94	11984,1	10404,4	12217,2	10936	10895,5	12471,4
5		7116,97	8063,23	7795,22	7023,32	7970,64	7317,88	8227,6	7577,72	7465,37	8129,09
6		8915,72	10638,5	10080,3	8694,06	10453,1	9252,11	10709,4	9642,67	9588,05	10792,1
7		4807,97	5084,91	5036,72	4796,3	5065,67	4897,66	5295,88	5009,62	4883,63	5091,77
8		9311,51	11251,6	10607,6	9064,59	11036,5	9685,07	11284,1	10111,6	10072,8	11430,3
9		7844,07	9119,14	8742,2	7733,89	8933,95	8123,58	9252,06	8437,19	8340,9	9216,43
10		10178,1	12646	11794,3	9859,26	12364,1	10628,4	12585,1	11151,6	11150,5	12889,2

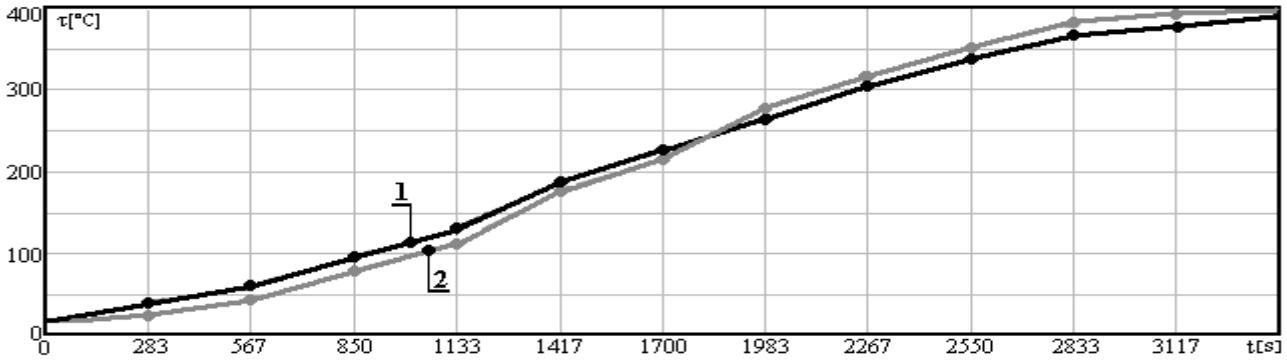


Fig.5. Comparison between experimental (2) and model data(1)

Dynamic Braking in Induction Motor Adjustable Speed Drives

Nebojsa Mitrovic[1], Milutin Petronijevic[2], Vojkan Kostic[3] and Bojan Bankovic[4]

Abstract – **A voltage source PWM inverter with diode front-end rectifier is one of the most common power configurations used in modern variable speed AC drives. However, it allows only unidirectional power flow. Various alternative circuits can be used to recover the load energy and return it to power supply. This paper presents the most popular topology used in adjustable speed drive (ASD). The diode rectifier is replaced with PWM voltage source rectifier. This is already an industrially implemented technology and known as most successful active front end (AFE) solution in ASD if regenerative operation is needed.**

Keywords – **Adjustable speed drive, Active front end.**

I. Introduction

Today's adjustable speed drives (ASD) in the low and mid power range are normally based on the concept of variable voltage, variable frequency (VVVF). Fig.1 shows the basic concept of a single variable speed drive. The three-phase AC supply network is rectified. The DC capacitor, which links the supply rectifier to the inverter, assures that the inverter sees a constant DC voltage from which it generates the required supply voltage and frequency to the motor.

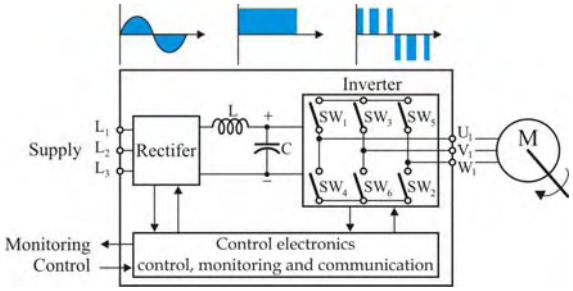


Fig. 1. Basic concept of a variable speed drive.

General classification divides induction motor control schemes into scalar and vector-based methods [1]. Opposite to scalar control, which allows control of only output voltage magnitude and frequency, the vector-based control methods enable control of instantaneous voltage, current and flux vectors. Type of the front end converter, regardless of the control schemes, depends on the power and torque requirements of the drive [2].

In order for an AC drive to operate in quadrant II or IV in speed-torque plane, rectifier must be able to deal with the electrical energy returned to the drive by the motor [3]. The typical pulse width modulated AC drive is not designed for regenerating power back into the three phase supply lines. Electrical energy returned by the motor can cause voltage in the DC link to become excessively high when added to existing supply voltage. Various drive components can be damaged by this excessive voltage. These regenerative conditions can occur when:

- quickly decelerating a high inertia load,
- controlling the speed of a load moving vertically downward (hoist, declining conveyor),
- a sudden drop in load torque occurs,
- the process requires repetitive acceleration and deceleration to a stop,
- controlling the speed of an unwind application.

In standard drives the rectifier is typically a 6-pulse diode rectifier only able to deliver power from the AC network to the DC bus but not vice versa. If the power flow changes as in two or four quadrant applications, the power fed by the process charges the DC capacitors and the DC bus voltage starts to rise. The capacitance is a relatively low value in an AC drive resulting in fast voltage rise, and the components of a frequency converter may only withstand voltage up to a certain specified level.

In order to prevent the DC bus voltage rising excessively, the inverter itself prevents the power flow from process to frequency converter. This is done by limiting the braking torque to keep a constant DC bus voltage level. This operation is called overvoltage control and it is a standard feature of most modern drives. However, this means that the braking profile of the machinery is not done according to the speed ramp specified by the user.

There are two technologies available to prevent the AC drive from reaching the trip level: Dynamic braking and active front end regeneration control [4,5]. Each technology has its own advantages and disadvantages.

II. Dynamic Braking

A dynamic brake consists of a chopper and a dynamic brake resistor. Fig.2 shows a simplified dynamic braking schematic. The chopper is the dynamic braking circuitry that senses rising DC bus voltage and shunts the excess energy to

[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]Milutin Petronijevic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: milutin.petronijevic@elfak.ni.ac.rs.
[3]Vojkan Kostic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vojkan.kostic@elfak.ni.ac.rs.
[4]Bojan Bankovic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: bojan.bankovic@elfak.ni.ac.rs.

the dynamic brake resistor. A chopper contains three significant power components: The chopper transistor is an IGBT. The chopper transistor is either ON or OFF, connecting the dynamic braking resistor to the DC bus and dissipating power, or isolating the resistor from the DC bus. The current rating of the chopper transistor determines the minimum resistance value used for the dynamic braking resistor. The chopper transistor voltage control regulates the voltage of the DC bus during regeneration. The dynamic braking resistor dissipates the regenerated energy in the form of heat.

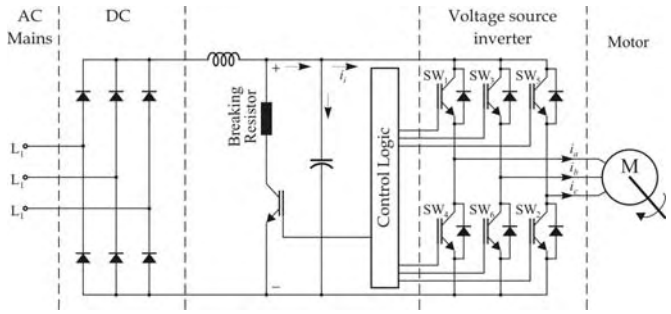


Fig. 2. Voltage source inverter with diode front end rectifier and dynamic brake module.

As a general rule, dynamic braking can be used when the need to dissipate regenerative energy is on an occasional or periodic basis. In general, the motor power rating, speed, torque, and details regarding the regenerative mode of operation will be needed in order to estimate what dynamic braking resistor value is needed. The peak regenerative power of the drive must be calculated in order to determine the maximum resistance value of the dynamic braking resistor [3].

The peak breaking power required to decelerate the load, according to equation (1) is:

$$P_b = \frac{J\,\omega_b(\omega_b - \omega_0)}{t_b} \tag{1}$$

where t_b represents total time of deceleration, ω_b and ω_0 initial and final speed in the process of braking.

The value of P_b can now be compared to the drive rating to determine if external braking module is needed. If peak braking power is 10% greater than rated drive power external braking module is recommended. Compare the peak braking power to that of the rated motor power, if the peak braking power is greater than 1.5 time that of the motor, then the deceleration time, needs to be increased so that the drive does not go into current limit.

Maximum dynamic brake resistance value can be calculated as:

$$R_{db} = \frac{V_{dc}}{P_b} \tag{2}$$

The choice of the dynamic brake resistance value should be less than the value calculated by equation (2). Once the maximum resistance value of the dynamic braking resistor current rating is known, the required rating of dynamic braking resistors can be determined. If a dynamic braking resistance value greater than the minimum imposed by the choice of the peak regenerative power is made and applied, the drive can trip off due to transient DC bus overvoltage problems. Once the approximate resistance value of the dynamic braking resistor is determined, the necessary power rating of the dynamic braking resistor can be calculated. The power rating of the dynamic braking resistor is estimated by applying what is known about the drive's motoring and regenerating modes of operation.

To calculate the average power dissipation the braking duty cycle must be determined. The percentage of time during an operating cycle (t_c) when braking occurs (t_b) is duty cycle ($\varepsilon = t_b/t_c$). Assuming the deceleration rate is linear, average power is calculated as follows:

$$P_{av} = \frac{t_b}{t_c}\frac{P_b}{2}\frac{\omega_b + \omega_0}{\omega_b} \tag{3}$$

Steady state power dissipation capacity of dynamic brake resistors must be greater than that average.

Fig.3a) shows the experimental results (DC voltage and chopper current) for the variable frequency drive with braking module in DC link and external braking resistor, under a step change of induction motor load in regenerative regime. Danfoss frequency (series VLT 5000) converter is used in experimental set-up. For the supply voltage of 400 V, DC link voltage is about 540 V.

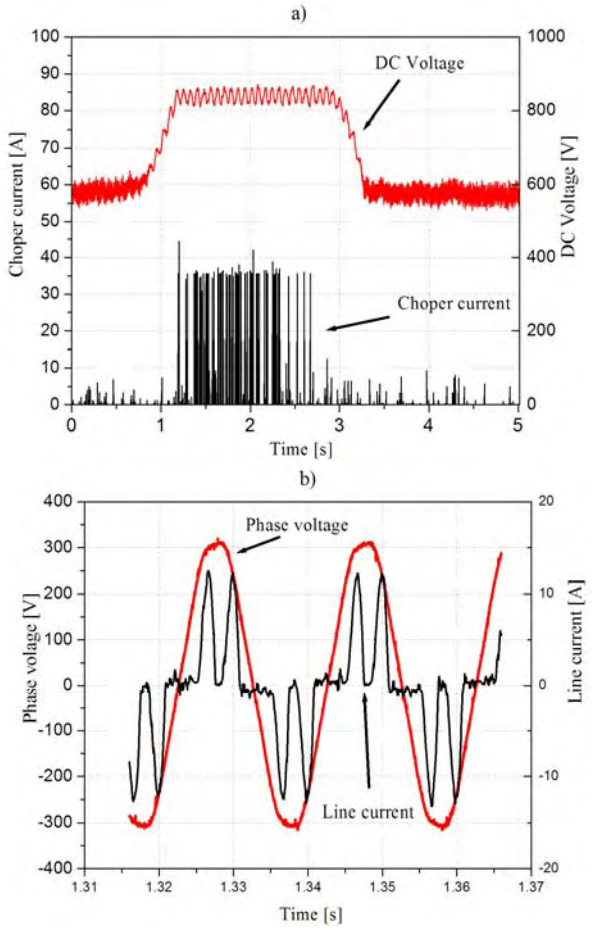


Fig. 3. a) DC voltage and chopper current, b) line voltage and current.

When negative load torque is applied, DC link voltage rises. The chopper transistor voltage control regulates the voltage of the DC bus during regeneration to near 800 V allowing current flow in the resistor. Regenerative energy is then realised into heat. After the end of the regenerative period, DC voltage returns to a value that corresponds to a motor regime. The Fig.3b) shows the line voltage and current at the input of the diode rectifier.

A voltage source PWM inverter with diode front-end rectifier is one of the most common power configurations used in modem variable speed AC drives, (Fig. 2). An uncontrolled diode rectifier has the advantage of being simple, robust, and low cost. However, it allows only unidirectional power flow. Therefore, energy returned from the motor must be dissipated on a power resistor controlled by a chopper connected across the dc link. A further restriction is that the maximum motor output voltage is always less than the supply voltage.

III. Active Front End Rectifier

Various alternative circuits can be used to recover the load energy and return it to power supply. One such scheme is shown in Fig. 4 and presents the most popular topology used in ASD. The diode rectifier is replaced with PWM voltage source rectifier. This is already an industrially implemented technology and known as most successful active front end (AFE) solution in ASD if regenerative operation is needed (e.g. for lowering the load in crane) and therefore was chosen by most global companies: Siemens, ABB, and others.

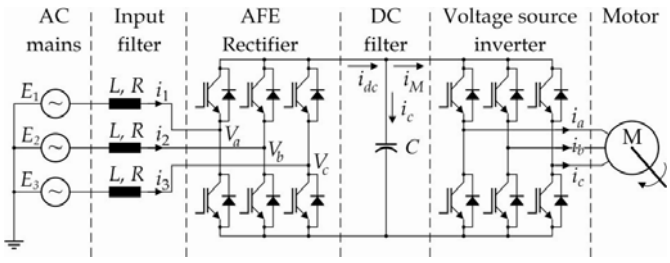


Fig. 4. Active front end inverter topology.

The term Active Front End Inverter refers to the power converter system consisting of the line-side converter with active switches such as IGBTs, the DC link capacitor bank, and the load-side inverter. The line-side converter normally functions as a rectifier. But, during regeneration it can also be operated as an inverter, feeding power back to the line. The line-side converter is popularly referred to as a PWM rectifier in the literature. This is due to the fact that, with active switches, the rectifier can be switched using a suitable pulse width modulation technique.

The PWM rectifier basically operates as a boost chopper with AC voltage at the input, but DC voltage at the output. The intermediate DC-link voltage should be higher than the peak of the supply voltage. The required DC-link voltage needs be maintained constant during rectifier as well as inverter operation of the line side converter. The ripple in DC link voltage can be reduced using an appropriately sized capacitor bank. The AFE inverter topology for a motor drive application, as shown in Fig.4, has two three-phase, two-level PWM converters, one on the line side, and another on the load side. The configuration uses 12 controllable switches. The line-side converter is connected to the utility through inductor. The inductor is needed for boost operation of the line-side converter.

For a constant dc-link voltage, the IGBTs in the line-side converter are switched to produce three-phase PWM voltages at *a*, *b*, and *c* input terminals. The line-side PWM voltages, generated in this way, control the line currents to the desired value. When DC link voltage drops below the reference value, the feed-back diodes carry the capacitor charging currents, and bring the DC-link voltage back to reference value.

The steady state characteristics as well as differential equations describing the dynamics of the front-end rectifier can be obtained independent of an inverter and motor load. This is because the DC-link voltage can be viewed as a voltage source, if V_{dc} is maintained constant for the full operating range. The inverter is thus connected to the voltage source, whose terminal voltage V_{dc}, remains unaffected by any normal inverter and motor operation [6].

Furthermore, as shown in Fig.4, the rectifier can also be viewed as connected to the voltage source V_{dc}. Thus, the rectifier is able to control magnitude and phase of PWM voltages V_{abc} irrespective of line voltages E_{123}.

The simple proportional-integral (PI) controllers are adopted in the current and voltage regulation. The control scheme of the PWM rectifier is based on a standard cascaded two-loop control scheme implemented in a *d-q* rotating frame: a fast control loop to control the current in the boost inductors and a much slower control loop to maintain constant dc-link voltage. The reference angle for the synchronous rotating *d-q* frame θ, is calculated, based on the three input phase voltages.

For the current control loop *d-q* synchronously rotating reference frame with the fundamental supply voltage frequency is used [7]. The line currents (i_1, i_2, i_3) are measured and transformed to the *d-q* reference frame, Fig.5.

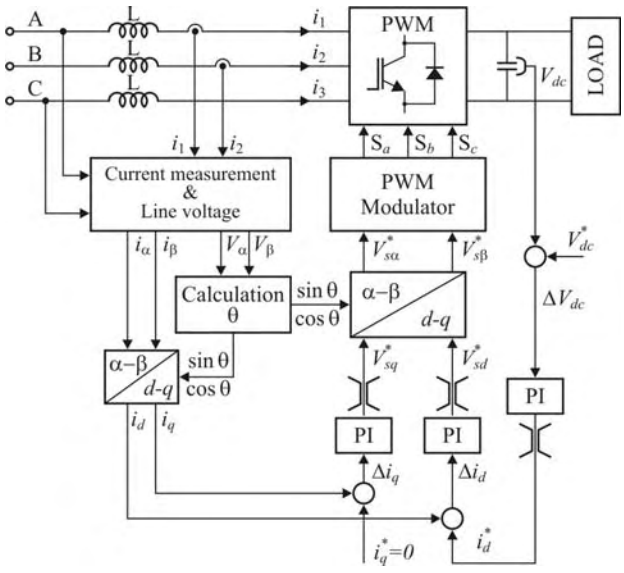


Fig. 5. Simplified block diagram of the AFE.

To get information about the position of the line voltage vector PLL (phase locked loop) is implemented. PI controllers for the *d-q* components of line current are identical and ωL terms are included to eliminate the coupling effect among the *d* and *q* components [7]. Outputs of the line current PI controllers present *d* and *q* components of the voltage across the line inductance. Subtracting this voltage from the supply voltage gives the converter voltage from the AC side that is used to get the modulation signal for proper switching of six switching devices.

The main task of the sinusoidal front end is to operate with the sinusoidal line current; so *d* and *q* components of the line current reference are DC values. Using this approach of control it is possible to control the output voltage of converter as well as the power factor of converter in the same time. To achieve unity power factor the reference of *q* current component need to be set on zero.

Different tests have been performed on the AFE inverter system to show some of the capabilities: simovert masterdrives 3-phase, 380-460V, 260A, inverter nominal power rating 132 kW. The measurements are done at steady-state operation. During experiments, the DC link voltage is boosted to 650 V. The first test is rectifier system operation when the induction machine operates as motor, Fig.6a), and second test is regenerative operation during regenerative operation, Fig.6.b).

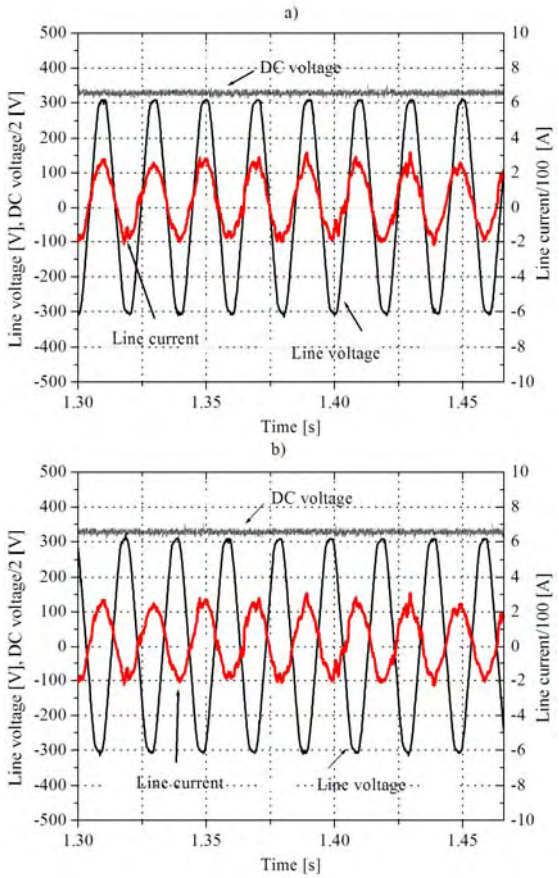


Fig. 6. Waveforms under steady-state operation: Line voltage, line current and dc link voltage a) motor operation, b) generator operation.

Both figures show the measured line currents, line voltages and DC voltage. It can be observed a high stationary performance both in motor and generator operation. The line current is nearly a sine wave with unity power factor while DC voltage is unchanged.

IV. Conclusion

The application of squirrel cage induction motors supplied from the frequency converters (also known as adjustable speed drive - ASD) have become the standard solution for the modern drives applications. However, the standard configuration of the inverter can not be used for some drives primarily due to regenerative operation, which in some cases may be intermittent and continuous.

This paper describes the solutions that are commonly used in modern crane drives. In case that it is a casual recuperating the dynamic braking is used. If continious regeneretation occur active front end rectifier capable to returning energy into the supply network is used. In addition 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.

Acknowledgement

This paper is supported by Project Grant III4400X (2011-2014) financed by Ministry of Education and Science, Republic of Serbia."

References

[1] M. Petronijevic, B. Veselic, N. Mitrovic, V. Kostic & B. Jeftenic, "Comparative Study of Unsymmetrical Voltage Sag Effects on Adjustable Speed Induction Motor Drives", Electric Power Applications, IET , vol.5, no.5, pp.432-442, May 2011

[2] B. Jeftenic, M. Bebic, S. Statkic, "Controlled multi-motor drives", International Symposium on Power Electronics, Electrical Drives, Automation and Motion, SPEEDAM 2006, Taormina (Sicily) - ITALY, 23-26 May 2006 , pp. 1392-1398.

[3] Rockwell Automation, Load Sharing for the 1336 PLUS II AC Drive, *Publication number* 1336E-WP001A-EN-P, June 2000.

[4] N. Mitrovic, M. Petronijevic, V. Kostic, B. Bankovic, "Active Front End Converter in Common DC Bus Multidrive Application", XLVI Inter. Conf. ICEST 2011, Serbia, Nis, June 29 - July 1, 2011, Proc. Vol.3, pp. 989-992, 2011.

[5] Rashid H., *Power Electronics Handbook*, Academic Press, San Diego, 2001.

[6] K. Hartani, Y. Miloud**,** "Control Strategy for Three Phase Voltage Source PWM Rectifier Based on the Space Vector Modulation"**,** Advances in Electrical and Computer Eng., Vol.10, Issue 3, pp. 61-65, 2010.

[7] M. Odavic, Z. Jakopovic, F. Kolonic, "Sinusoidal Active Front End under the Condition of Supply Distortion", AUTOMATIKA 46(2005) 3–4, pp.135–141, 2005.

Application of Active Front End Rectifier in Electrical Drives

Bojan Banković[1], Nebojša Mitrović[2], Vojkan Kostić[3] and Milutin Petronijević[4]

Abstract – **The term Active Front End Inverter (AFE) refers to the power converter system consisting of the line-side converter with active switches such as IGBTs, the dc link capacitor bank, and the load-side inverter. The line-side converter normally functions as a rectifier. During regeneration it can also be operated as an inverter, feeding power back to the line. This paper presents some applications of active front end rectifiers in induction motor drives.**

Keywords – **Induction motor, Active front rectifier.**

I. INTRODUCTION

Modern electrical drives can consist of different electrical machines such are induction motors and different converter topologies with various control algorithms. Some of them had to have adjustable speed and should be capable of energy saving. It could be achieved with the usage of very efficient machines and converter components and with the recuperation capability of electrical power into the grid. AC/DC Voltage Source Converters (VSC) are widely used in industrial AC drives as Active Front End (AFE), DC-power supply, power quality improvement and harmonic compensation (active filter) equipments [1], [2], [3]. The use of the AFE meets the requirements for efficient energy feedback to the network.

There are several possibilities for realization of breaking or recuperation in the electrical drives. One of them is usage of the inverter with diode rectifier on line side, DC-link capacitor with brake resistor and brake chopper as it is shown on Fig. 1.

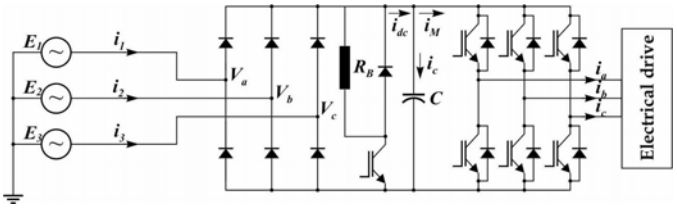


Fig.1. The AC-AC inverter with diode rectifier, DC-link capacitor, brake resistor, brake chopper and DC-AC drive inverter

The breaking of the induction motor leads to an energy flow back into DC link where the braking chopper and resistor limits the voltage rise and converts energy into heat. This method is not appropriate for high power induction motors where conversion of energy into the heat could last very short period of time. A better way for realization of induction motor breaking is application of an inverter on the drive side with line side step-up inverter and DC link between them as shown in Fig. 2.

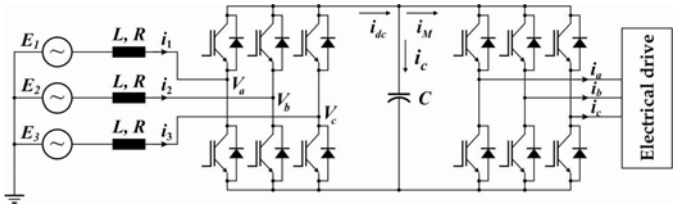


Fig.2. Active front end inverter topology with line-side inverter, DC-link and DC-AC drive inverter

On that way the recuperation energy is feeded back into the mains by a line side inverter in combination with a line side choke. The AC/DC converters become very important part of an AC/DC/AC line interfacing converters in renewable and distorted energy systems. The advantages of such topology (AFE) are[4], [5]:

- bi-directional power flow,
- sinusoidal current waveform,
- high input power factor including unity power factor operation,
- control and stabilization of the DC-link voltage,
- low harmonic distortion of line current
- reduced size of line inductor
- operation under line voltage distortion

Power factor can be controlled by phase angle between voltage and current waveform (lagging and leading power factor). The problem that may occur with AFE is high frequency harmonics which are function of the switching frequency. On that basis harmonics should be filtered. The complexity of the control for the active front end is comparable with the complexity of the field oriented control of the inverter for the motor. The line side and motor side inverter sections are thereby almost stressed by switching losses in the same amount.

II. SYSTEM DESCRIPTION

A. Mathematical model of Voltage-Source PWM Rectifier

To design control system and simulation model, it is necessary to define mathematical model of the Voltage Source Converter (VSC). As we mentioned before the VSC consists of IGBT module, which consists of six transistors and parallel connected freewheeling diodes. However, it is difficult to

[1]Bojan Bankovic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: bojan.bankovic@elfak.ni.ac.rs.

[2]Nebojsa Mitrovic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: nebojsa.mitrovic@elfak.ni.ac.rs.

[3]Vojkan Kostic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: vojkan.kostic@elfak.ni.ac.rs

[4]Milutin Petronijevic, University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: milutin.petronijevic@elfak.ni.ac.rs.

include all nonlinearities of IGBT module, and therefore following assumptions have been taken into considerations:

- transistors are assumed to be ideal switches
- time delay between control signals and IGBT module is neglected,
- transistor switching off time delay is zero,
- no switching loses.

Also, AC-side has been simplified as follows:

- symmetrical line voltages
- internal resistance of voltage supplies and wires resistance is zero,
- negligible small supply inductance
- ideal VSC input choke

The steady state characteristics as well as differential equations describing the dynamics of the front-end rectifier can be obtained independent of an inverter and motor load. This is because the DC-link voltage can be viewed as a voltage source. If V_{dc} is maintained constant for the full operating range, the terminal voltage remains unaffected during any electrical drive operation. The three-phase line voltages are shown in Eq. 1:

$$E_1 = E_m \sin(\omega t)$$
$$E_2 = E_m \sin\left(\omega t - \frac{2\pi}{3}\right) \quad (1)$$
$$E_3 = E_m \sin\left(\omega t - \frac{4\pi}{3}\right)$$

and since there is no neutral connection we obtain Eq. 2:

$$I_1 + I_2 + I_3 = 0 \quad (2)$$

where E_{123} and I_{123} are line voltages and currents.

The rectifier is able to control magnitude and phase of PWM voltages V_{abc} irrespective of line voltages E_{123}. The dynamic equations for each phase can be written as [6]:

$$\begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} = L\frac{d}{dt}\begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} + R\begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} + \begin{bmatrix} V_{a0} \\ V_{b0} \\ V_{c0} \end{bmatrix} \quad (3)$$

In synchronous rotating dq reference frame Equations 4 and 5 represent the dynamic d-q model of an active front end inverter in a reference frame rotating at an angular speed of ω.

$$L\frac{di_{qe}}{dt} = E_{qe} - \omega L i_{de} - R i_{qe} - V_{qe} \quad (4)$$

$$L\frac{di_{de}}{dt} = E_{de} + \omega L i_{qe} - R i_{de} - V_{de} \quad (5)$$

To completely define system dynamics the DC side of the three-phase PWM rectifier can be expressed by Eq. 6:

$$i_{dc} = C\frac{dV_{dc}}{dt} + i_M \quad (6)$$

where, i_{dc} is the total dc-link current supplied by the rectifier, while i_M is the load-side DC current which is the result of induction motor operation.

The voltage components E_{qe} and E_{de} are computed from source voltages, E_1, E_2, and E_3. Since line voltages are known, the angular frequency, ω, can be easily estimated. The PWM voltages V_{qe} and V_{de} are the two inputs to the system which are generated using the sine-triangle PWM controller. L and R represent series impedance.

Eq. 4 and Eq. 5 shows that d-q current is related with both coupling voltages $\omega L i_q$ and $\omega L i_d$, and main voltage E_d and E_q, besides the influence of PWM voltage V_{qe} and V_{de}. Voltage V_{qe} and V_{de} are the inputs, controlled in such a way as to generate desired currents. Now new variables V'_{qe} and V'_{de} are defined,

$$V_{qe} = -V'_{qe} - \omega L i_{qe} + E_{qe} \quad (7)$$

$$V_{de} = -V'_{de} + \omega L i_{de} + E_{de} \quad (8)$$

So that the new system dynamic equations become,

$$L\frac{di_{qe}}{dt} = -i_{qe}R + V'_{qe} \quad (9)$$

$$L\frac{di_{de}}{dt} = -i_{de}R + V'_{de} \quad (10)$$

In Eq. 9 and Eq. 10 two axis current are totally decoupled because V'_{qe} and V'_{de} are only related with i_{qe} and i_{de} respectively. On this way we eliminate the coupling between two current components. To ensure constant DC-link voltage a simple proportional-integral (PI) controller is applied to the DC-link voltage error, resulting in the current reference command $i^*_{d\_e}$. Two simple proportional-integral (PI) controllers are adopted in the current regulation Fig. 3 and one proportional-integral (PI) controller in voltage regulation, Fig. 4.

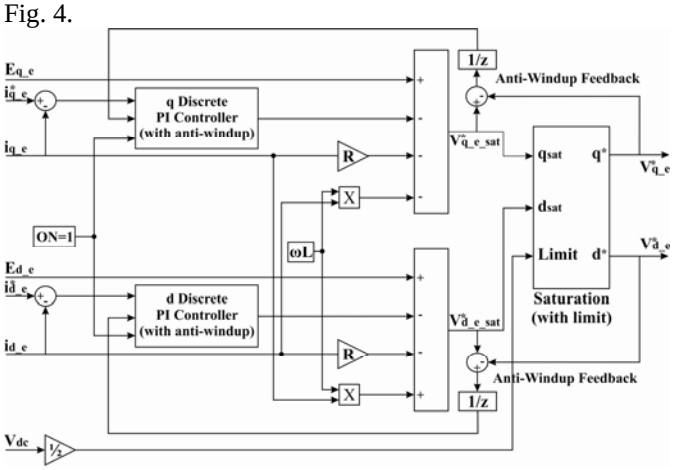


Fig. 3. Decoupled current controller with active damping

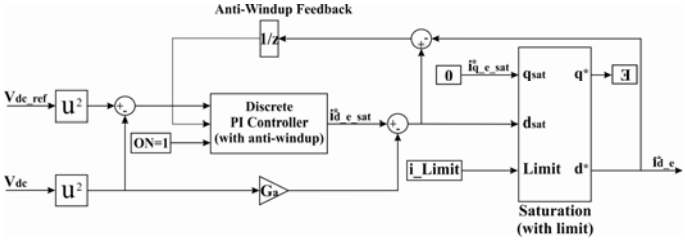


Fig. 4. Voltage controller with active damping

The control scheme of the PWM rectifier is based on a standard cascaded two-loop control scheme implemented in a d-q rotating frame: a fast control loop to control the current in the boost inductors and a much slower control loop to maintain constant DC-link voltage. The reference angle for the synchronous rotating d-q frame $\boldsymbol{\theta}$, is calculated, based on the three input phase voltages. For the effective PI control of

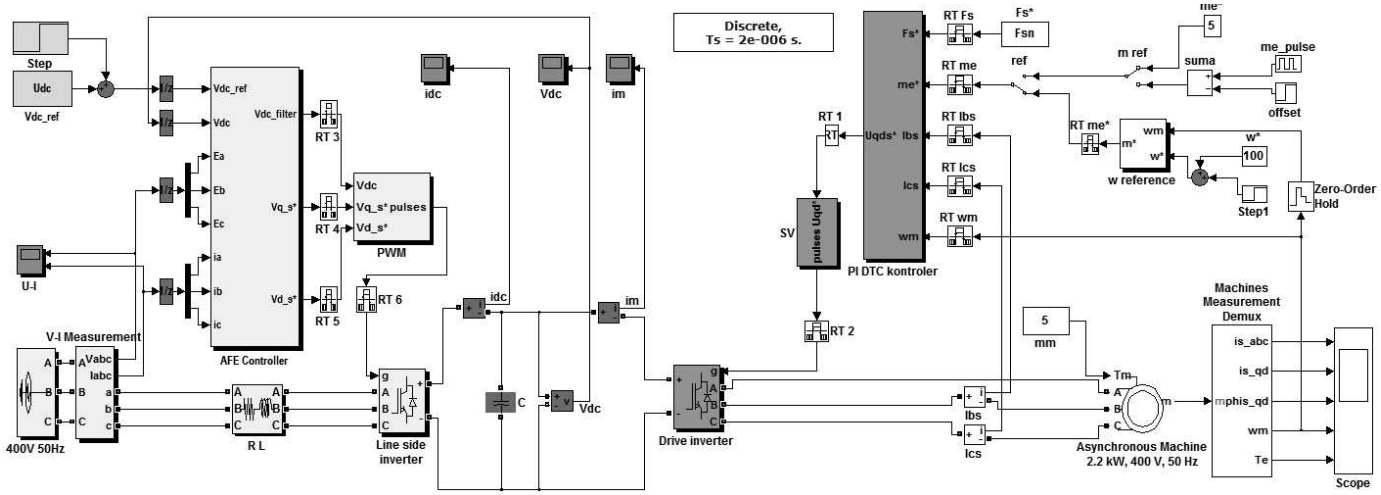


Fig. 5 Three-phase voltage source PWM rectifier with DTC induction drive mode

the DC-link voltage the linearization of the rectifier's output volatge contol loop has been carried out. The control variable (V^2_{dc}) has been taken into account instead the V_{dc}. On this way we eliminated the nonlinearity of PWM rectifier and made the linear PI voltage controller more robust. The line currents (i_1, i_2, i_3) are measured and transformed to the *d-q* reference frame in AFE controller block as it shown at Fig. 3. PI controllers for the *d* and *q* components of line current are identical and ωL terms are included to eliminate the coupling effect among the *d* and *q* components as it is shown at Fig. 5. Outputs of the line current PI controllers present *d* and *q* components of the voltage across the line inductance. Subtracting this voltage from the supply voltage gives the converter voltage from the AC side that is used to get the modulation signal for proper switching of six switching devices. Using this approach of control it is possible to control the output voltage of converter as well as the power factor of converter in the same time. To achieve unity power factor the reference of *q* current component need to be set on zero.

Over the years many control techniques have been adopted for the rectifier device in order to improve the AC power input power factor and reform the input current shape. The pulse-with modulation techniques include sinusoidal pulse-width modulation (SPWM), space vector pulse-width modulation (SVPWM), delta modulation techniques. According to traditional SPWM, the SVPWM has characteristic of high utilization rate of DC supply voltage and fast dynamic response [7], [8]. The current control strategies for voltage source converter on line side with PWM control scheme for three phase rectifier are used in this paper. On the load side the Direct Torque Control (DTC) of the motor is used as a control scheme.

III. Simulation Results

According to the analysis above, the simulation model of whole system is built based on Matlab/Simulink and shown on Fig. 5. The whole system behavior is simulated as a discrete control system. The main parameters of the simulated circuit are given as following: Input phase to phase voltage is 400 V/50 Hz where the AC source is an ideal balanced three-phase voltage source. The inductance of AC side filter reactor L is 2.4 mH the AC side resistance R is 0.05 Ω, the DC side output capacitor C is 360 μF, the DC voltage is set to 513 V and the switching frequency is set to 10 kHz, nominal motor power 2,2 kW, nominal speed 302 rad/s, nominal torque 6.67 Nm and nominal motor current 5,2 A.

The induction machine is initially running at a constant speed reference 100 rad/s and under a 5 Nm load regime. In 0,5 s, we apply a rated speed reference 200 rad/s under 5Nm load, Fig. 6. Changing the speed reference corresponds to a step up speed perturbation. This is motor mode of induction machine when energy flows from line side to load side.

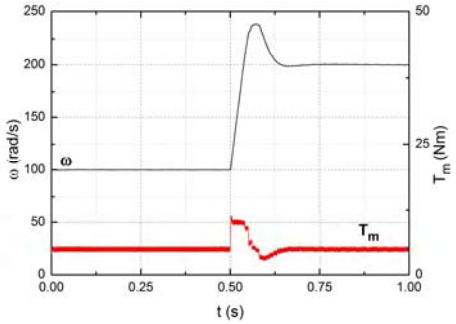


Fig. 6. Motoring mode: Reference speed and motor load

Because of unity power factor the line voltages and currents are in phase as it is shown on Fig. 7.

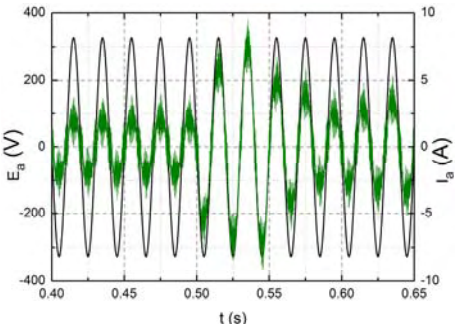


Fig. 7. Motoring mode: The waveform of A phase input voltage and current (unity power factor)

During this simulation voltage in DC-link V_{dc} was constant with small peak in 0,5 s, Fig. 8.

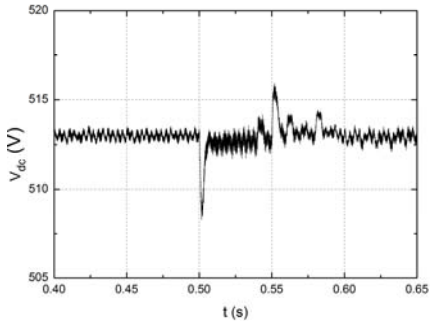


Fig. 8. Motoring mode: The waveform of DC output voltage

In second case the induction machine is initially running at a constant speed reference of 100 rad/s but with -5 Nm load regime. In 0.5 s we apply a rated speed reference 200 rad/s under the same load regime, Fig. 9. This corresponds to the regenerative mode of induction machine. On this way energy goes from induction machine to line side through the DC-link.

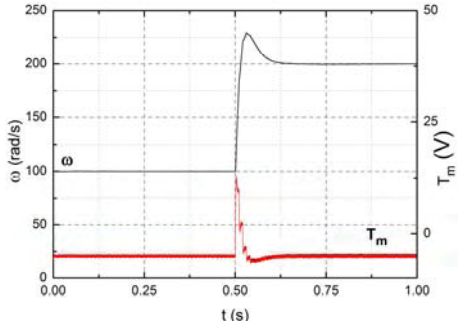


Fig. 9. Generating mode: Reference speed and motor load

The unity power factor in regenerative mode is shown at Fig. 10.

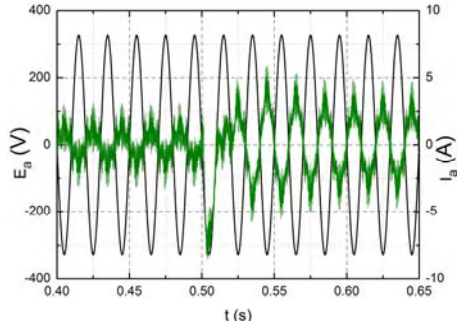


Fig. 10. Generating mode: The waveform of A phase input voltage and current (unity power factor)

Also the V_{dc} voltage was constant with small peak in 0,5 s, Fig. 11.

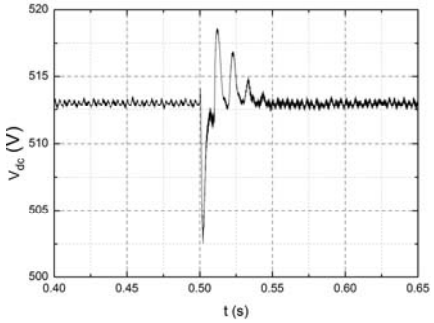


Fig. 11. Generating mode: The waveform of DC output voltage

IV. Conclusion

The efficiency and recuperation possibilities of AFE topologies become very important aspects in electric drives because of the possibility of bi-directional power flow, sinusoidal current waveform, high input power factor, control and stabilization of the DC-link voltage and low harmonic distortion of line current. A simple unity power factor of AFE rectifier and its application in electrical drives is introduced in this paper. Simulation result shows that unity power factor is achieved in both working mode of induction machine. Constant dc voltage reaches the requirement of design.

Acknowledgement

This paper is supported by Project Grant III44004 and III44006 (2011-2014) financed by Ministry of Education and Science, Republic of Serbia."

References

[1] M. P. Kazmierkowski, R. Krishnan, and F. Blaabjerg, Control in Power Electronics. Academic Press - USA, 2002.

[2] M. Lissere, P. Rodriguez, J. Guerrero, and R. Teodorescu, "Power electronics for pv energy systems integration," 39th IEEE PESC, 15–19 June 2008, tutorial.

[3] M. P. Kazmierkowski, M. Jasinski, and H. C. Sorensen, "Ocean wave energy converters-wave dragon mw," Przeglad Elektrotechniczny, vol. 84, no. 2, pp. 8–13, 2008.

[4] Ricardo Luiz Alves and Ivo Barbi, "A New Hybrid High Power Factor Three-Phase Unidirectional Rectifier" Industrial Electronics, 2006 IEEE International Symposium on Volume 2, pp. 1046-1051, July 2006.

[5] Ye, Y., Kazerani, M., Quintana, V.H., "A Novel Modeling and Control Method for three-phase PWM converters", Power Electronics Specialists Conference, 2001. PESC. IEEE 32nd Annual Volume 1, 17-21, pp.102-107, June 2001.

[6] Wang Xu, Huang Kaizheng, Yan Shijie, Xu Bin. " Simulation of Three-phase Voltage Source PWM rectifier Based on Space Vector Modulation". Control and Decision Conference, Chinese, IEEE, Pages: 1881-1884, 2008.

[7] Ian Wallace, Ashish Bendre, Jonathan P. Nord and Giri Venkataramanan,"A unity-power-factor three-phase PWM SCR rectifier for high-power application in the metal industry," IEEE Transactions on Industry Applications, Vol. 38, No. 4, pp. 898-908, July/August 2002.

[8] Poh Chiang Loh and Donald Grahame Holmes, "A variable band universal flux/charge modulator for VSI and CSI modulation," IEEE Transactions on Industry Applications, Vol. 38, No. 3, pp.695-705, May/June 2002.

Cyclic Current Rating of Single-Core XLPE Cables with Respect to Designed Life Time

Miodrag Stojanović[1], Dragan Tasić[2] and Aleksa Ristić[3]

Abstract – **This paper deals with cyclic current rating of single-core XLPE cable buried in the ground. Cyclic current rating obtained by IEC60853-1 and IEC60853-2 standards is conservative because it is based on the assumption that maximum temperature during the cycle is equal to maximum allowed steady-state temperature. Using the Arrhenius-IPM electrothermal life model, cyclic rating factor of single-core XLPE cables is calculated for assigned design life of cables.**

Keywords – **Single-core cables, Cyclic current rating, Electrothermal life model.**

I. Introduction

The rated current of HVAC cables involves operation at the rated temperature, i.e. the maximum allowed continuous temperature of cable insulation, and cables are designed under the conservative constraints that their maximum temperature is steady and equal to rated temperature. The methods for calculation of overload capacity of cables under cyclic loading are given in [1-6]. Overload capacity is quantified by cyclic rating factor. Standard IEC60853-1 gives method for calculating the cyclic rating factor for cables where internal thermal capacitance can be neglected. Simplified method in this standard requires only knowledge of shape of the load variation for not more than six hours immediately preceding the time of maximum temperature and an average value for times before that. The method can be applied to all sizes and types of cable for nominal voltage up to and including 18/30 (36) kV. Standard IEC60853-2 gives manual method for calculating cyclic rating factor for cables whose internal thermal capacitance cannot be neglected. In general, this method is applied to cables for voltages greater then 18/30 (36) kV. Cyclic ratings include temporary overloads under condition that rated temperature is not exceeded.

In this paper overload capability of single-core XLPE cables buried in the ground is analyzed. Firstly the temperature of cable is calculated for typical daily cycle diagram of HVAC cables. Using the Arrhenius-IPM electrothermal life model, expected life time of single-core XLPE cables is calculated for different values of overload factor, as well as failure rate at the end of design life time.

[1]Miodrag Stojanović is with the University of Niš, Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: miodrag.stojanovic@elfak.ni.ac.rs.

[2]Dragan Tasić is with the University of Niš, Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: dragan.tasic@elfak.ni.ac.rs.

[3] Aleksa Ristić is with the University of Niš, Faculty of Electronic Engineering, A. Medvedeva 14, 18000 Niš, Serbia, E-mail: aleksa.ristic@elfak.ni.ac.rs.

II. Thermal Response of Cable and Cyclic Rating Factor

Loading of HVAC power cables varies on the daily cycles. In order to perform computations for variable loading, a daily load curve is divided into a series of steps of constant magnitude and the same duration as that illustrated in Fig. 1. Duration of one step is $\Delta t_i=24/N$, where N is the number of steps. For different successive steps, the computations are done repeatedly, and the final result is obtained using the principle of superposition. The variation of the cable conductor temperature during a stepwise-constant load cycle can be determined by calculating the thermal response of the HVAC cable and the surrounding environment to each step change of load current. The two partial temperature transients are solved separately in sequence and then combined, thereby finding an analytical solution for the whole transient as follows.

Transient temperature rise of the conductor above the ambient, due to the i^{th} step of stepwise-constant load cycle is:

$$\theta_i(t) = \theta_{c,i}(t) + \alpha_i(t)\theta_{e,i}(t) \qquad (1)$$

where $\theta_{c,i}(t)$ is transient temperature rise of the conductor above cable surface, $\theta_{e,i}(t)$ is cable surface temperature rise above ambient and $\alpha_i(t)$ is attainment factor. Transient temperature rise of the conductor above cable surface is calculated as:

$$\theta_{c,i}(t) = W_{J,i}[T_a(1-e^{-at}) + T_b(1-e^{-bt})] \,, \qquad (2)$$

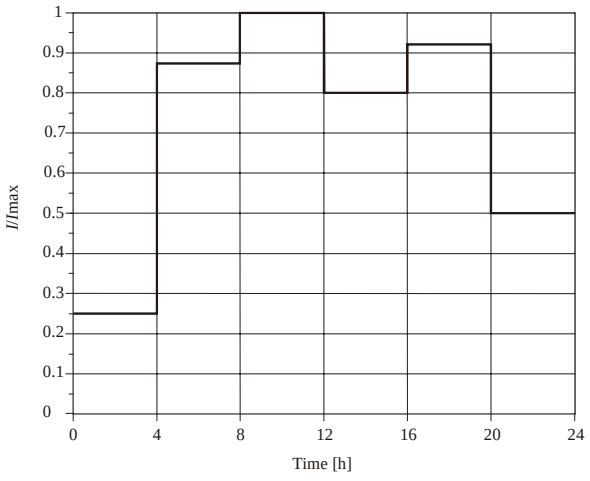


Fig. 1. Daily load cycle

where $W_{J,i}$ is power loss per unit length in a conductor due to the i^{th} current step, and T_a, T_b, a, b are thermal resistances and corresponding coefficients of the first loop of CIGRE transient two-loop network [6].

For the case of three single-core cables buried directly in the ground in a flat formation temperature rise of cable surface above the ambient is:

$$\theta_e(t) = \frac{\rho_{tz} W_{1,i}}{4\pi}\left\{-\mathrm{Ei}\left(-\frac{D_c^2}{16t\delta}\right)+\mathrm{Ei}\left(-\frac{L^2}{t\delta}\right)+ +2\left[-\mathrm{Ei}\left(-\frac{d_1^2}{16t\delta}\right)+\mathrm{Ei}\left(-\frac{d_1'^2}{t\delta}\right)\right]\right\}, \quad (3)$$

where ρ_{tz} is thermal resistivity of soil, δ is diffusivity, D_c is outer diameter of cable, d_1 is center-to-center distance between cables, $d_1' = \sqrt{(2L)^2 + d_1^2}$, $W_{1,i} = W_{J,i}(1+\lambda_1)$, λ_1 is sheath loss factor, and Ei is exponential integral [9].

The attainment factor in equation (1) is calculated as:

$$\alpha_i(t) = \frac{\theta_{c,i}(t)}{W_{J,i}(T_a + T_b)} . \quad (4)$$

Finally, temperature of conductor is obtained adding the temperature rise for each step of the load diagram, calculated by (1), to temperature of the ground and temperature rise due to dielectric losses. Electrical resistance of conductor and corresponding joule losses are calculated with respect to temperature of conductor reached in each step of daily diagram.

The cyclic rating factor is the factor by which rated current may be multiplied to obtain the permissible peak value of current during a daily cycle such that the conductor attains, but does not exceed rated temperature. For calculating cyclic rated factor, according to standard IEC60853-2, only load cycle over a period of six hours before the time of maximum temperature is needed, while earlier values are replaced with constant one, proportional to loss-load factor. Therefore, the cyclic rating factor is given by:

$$M = \frac{1}{\left(\sum_{i=0}^{5} Y_i\left(\frac{\theta_R(i+1)}{\theta_R(\infty)} - \frac{\theta_R(i)}{\theta_R(\infty)}\right) + \mu\left(1 - \frac{\theta_R(6)}{\theta_R(\infty)}\right)\right)^{1/2}} \quad (5)$$

where:

$$\frac{\theta_R(t)}{\theta_R(\infty)} = \alpha(t)(1 - k_1 + \beta(t)k_1) \quad (6)$$

$$k_1 = \frac{W_1(T_4 + \Delta T_4)}{W_J(T_A + T_B) + W_1(T_4 + \Delta T_4)} \quad (7)$$

$$\beta(t) = \frac{\rho_{tz}}{4\pi(T_4 + \Delta T_4)}\left\{-\mathrm{Ei}\left(-\frac{D_c^2}{16t\delta}\right)+\mathrm{Ei}\left(-\frac{L^2}{t\delta}\right)+ +2\left[-\mathrm{Ei}\left(-\frac{d_1^2}{16t\delta}\right)+\mathrm{Ei}\left(-\frac{d_1'^2}{t\delta}\right)\right]\right\} . \quad (7)$$

Thermal resistance T_4 for cable buried directly in the ground is:

$$T_4 = \frac{\rho_{tz}}{2\pi}\ln\left(\frac{2L}{D_c} + \sqrt{\left(\frac{2L}{D_c}\right)^2 - 1}\right) , \quad (8)$$

while thermal resistance ΔT_4 in the case cable line consisted of three single-core cables is:

$$\Delta T_4 = \frac{\rho_{tz}}{\pi}\ln\left(\frac{d_1'}{d_1}\right) . \quad (9)$$

Obviously, the method for cyclic rating factor calculation does not respect variation of conductor temperature and assumes that joule losses are directly proportional to square of current. The results obtained on this way are therefore also conservative.

III. Electrothermal Life Model

Electrical stress and thermal stress are dominant aging factors for HVAC cables. The most popular electrothermal life model is the combination of two popular single-stress life models, the Arrhenius model for thermal life and the Inverse power model for electrical life [7,8]. Expected life time of the cable whose temperature is *T* and electric field in insulation is E can be calculated:

$$L = L_0\, e^{-BcT}\left(\frac{E}{E_0}\right)^{-(n_0 - bcT)} , \quad (10)$$

where E_0 is value of electric field below which electrical aging is negligible, $cT = 1/T_0 - 1/T$ is conventional thermal stress, T_0 is reference temperature, n_0 is voltage endurance coefficient, L_0 is life at $T=T_0$ and $E=E_0$, $B=\Delta W/k$, ΔW is activation energy of the main thermal degradation reaction, k is Boltzmann constant, and *b* is parameter that rules synergism between electrical and thermal stress. The parameters of Arrhenius-IPM model for XLPE insulation are given in Table I. During one step of daily stepwise-constant load cycle temperature of insulation varies. Therefore, different values of thermal life of cable insulation are obtained for each moment during the day. Loss-of-life fraction relevant to the i^{th} step of stepwise-constant load diagram (Fig.1) is defined as:

$$LF_i = \int_0^{\Delta t_i} \frac{dt}{L_i(T)} . \quad (11)$$

TABLE I
PARAMETERS OF ARRHENIUS-IPM MODEL

Parameter	Value
b [K]	4420
B [K]	12430
n_0 [non-dimensional]	15
E_0 [kV/mm]	5
T_0 [K]	293
β_t	2

According to Miner's cumulative damage theory, the sum of all loss-of-life fractions should yield 1 at failure. So, number of cycles to failure (number of days to failure in case of daily cycles) is:

$$K = \left(\sum_{i=1}^{N} LF_i \right)^{-1} . \quad (12)$$

Design life of cable corresponds to certain design failure probability. The cumulative probability distribution function that is commonly used for associating time to failure probability in case of polymeric insulation for power cables is the Weibull's one. Failure probability at mission time t_p is:

$$P(t_p, E, T) = 1 - e^{[-(t_p / L_{63\%})^{\beta_t}]} . \quad (13)$$

where β_t is share parameter of cumulative probability distribution function and $L_{63\%}$ is failure-time with 63.2% probability. The relevant failure rate at time t_p can be estimated through the following hazard function:

$$h(t_p, E, T) = \frac{\beta_t}{L_{63\%}} \left(\frac{t_p}{L_{63\%}} \right)^{\beta_t - 1} . \quad (14)$$

Based on the equation (14) failure rate can be calculated for insulation of cable loaded by defined daily stepwise-constant cycle at the end of design life time.

IV. TEST EXAMPLE

The procedure described in the previous sections is applied to HVAC XLPE insulated single-core cables [10] with aluminum conductors and copper wire screen. Maximum voltage of cables is 123 kV. The data about cables are shown in the Table II. In Table II S_c is conductor cross section, d_c is diameter of conductor, D_i is diameter above the insulation, δ_i is insulation thickness, D_c is outer diameter of cable, and I_r is rated current for considered cable formation and used bonding method of metal screens. It is assumed that three single-core cables laid in a flat formation and metal screen of cables are cross-bounded.

Laying depth of cables is 1 m, ground temperature 20°C and distance between cables is $d_1=D_k+70$ mm. Design life of cables is 30 years, while rated temperature of conductor is 90°C. Thermal resistivity and thermal capacity of cross-linked polyethylene is 3.5 Km/W and $2.4 \cdot 10^6$ J/(m^{3}K), thermal resistivity of soil 1 Km/W, and thermal capacities of aluminum and copper are $2.5 \cdot 10^6$ J/(m^{3}K) and $3.45 \cdot 10^6$ J/(m^{3}K) respectively. An assumption of constant electric field equal to design values is used. Currents that correspond to each step of daily load cycle are given in Table III.

Daily variation of conductor temperature for single-core XLPE cable with 1000 mm^2 cross-section is shown in Fig. 2 for different values of overload factor (maximum current relative to rated current). For overload factor value of 1.1 difference between maximum and minimum temperatures is only 17ºC (varies between 48ºC and 65ºC), while for overload factor value of 1.4 this difference is 41ºC. For overload factor value of 1.2, the maximum temperature of conductor is close to 90ºC. The last row of Table II shows the results of cyclic rating factor calculated by (5). As can be seen, these values vary in very narrow range between 1.173 and 1.185. For 1000mm^2 cross-section cyclic rating factor has value of 1.176.

TABLE II
CABLE DATA

S_c [mm^2]	630	800	1000	1200	1400
d_c [mm]	29.8	33.7	37.9	42.8	46.4
D_i [mm]	58.6	62.5	67.3	73.8	77.4
δ_i [mm]	13	13	13	13	13
D_c [mm]	72.3	76.8	82	89.5	93.3
I_r [A]	740	845	950	1025	1100
M	1.173	1.176	1.176	1.182	1.185

TABLE III
DAILY LOAD CYCLE

Step. No.	Time [h]	I/I_{max}
1	00-04	0.25
2	04-08	0.875
3	08-12	1
4	12-16	0.8
5	16-20	0.917
6	20-24	0.5

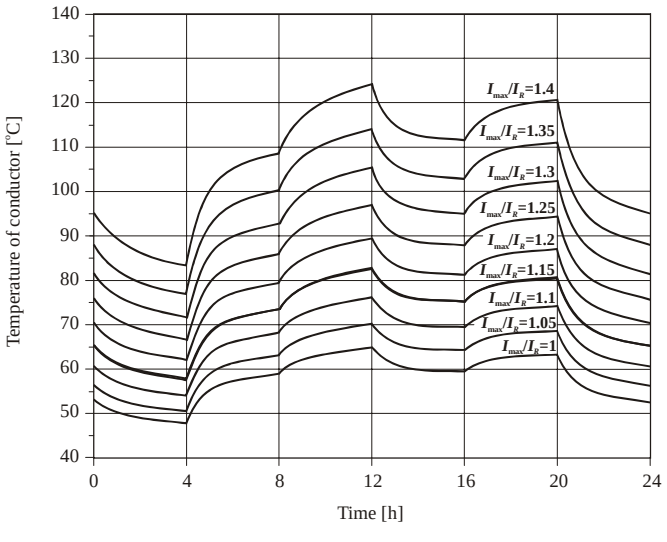


Fig. 2. Temperature variation during daily cyclic loading

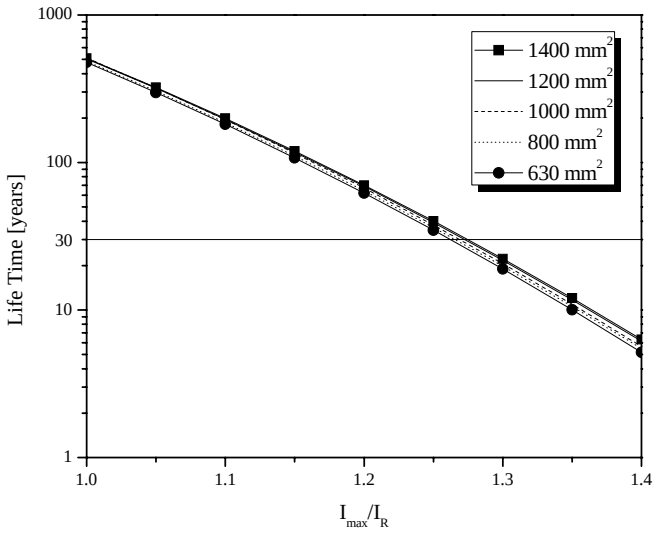


Fig. 3. Life time of cables for cycle loading

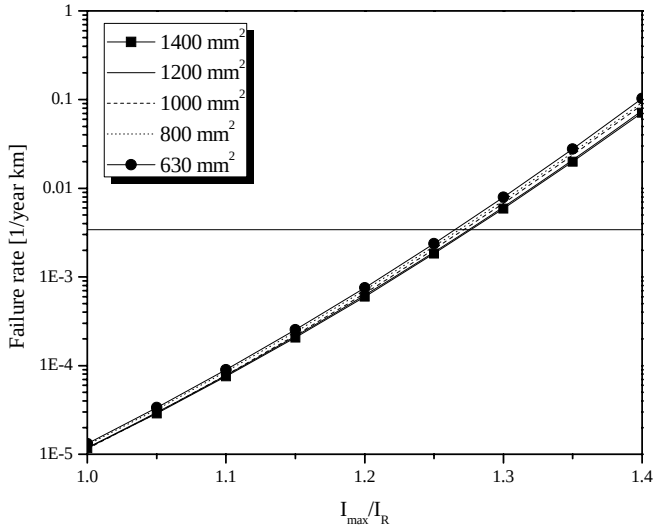


Fig. 4. Failure rate of cables at the end of design life

Fig. 3 shows expected life time of single-core XLPE cables with different cross-section areas in the case of daily cyclic loading as function of overload factor. Having in mind that design life of cable is 30 years, it is obvious that for daily cyclic diagram shown in Table III cables can be overloaded more than 27% without shortening the life time. As previously noticed, according to the Fig. 2, overload factor of the cable for given daily cycle diagram and maximal temperature at 90ºC is 1.2. Fig. 3 shows that cable life time from the aspect of electrothermal aging, for given overload factor of 1.2, is 65 years. For cyclic rating factor of 1.176 calculated according to relevant standard, life time from the aspect of electrothermal aging is approximately 100 years. Fig. 4 shows results of failure rate calculation at the end of design life time. As can be seen from this figure, for overload factor of 1.176 or 1.2, values of failure rates at the end of design life time are very low.

V. Conclusion

In this paper, a heating analysis of single-core XLPE cables buried in the ground, at daily load cycle, is conducted. For the assumed daily load cycle, cyclic rating factors and daily temperature variation at different values of overload factor are determined. On the basis of obtained results, it is shown that rated temperature is attained at value of overload factor slightly higher than cyclic rating factor value. Expected cable life time for different values of overload factor, respecting the daily cable temperature variation, is determined. On the basis of obtained results, it is concluded that overload capability of single-core XLPE cable is greater than one obtained by relevant IEC standard, when calculating with design life time.

Acknowledgement

This work has been funded by the Serbian Ministry for Science under the project TR-33008.

References

[1] J.H. Neher, M.H. McGrath, “The Calculation of the Temperature Rise and Load Capability of Cable Systems”, AIEE Transactions, Part III, Power Apparatus and Systems, vol. 76, no. 3, pp. 752-772, 1957.

[2] G. J. Anders, M.A. El-Kady, “Transient Rating of Buried Power Cables, Part 1: Historical Perspective and Mathematical Model”, IEEE Trans. Power Delivery, vol. 7, no.4, pp. 1724-1734, 1992.

[3] G. J. Anders, *Rating of Electric Power Cables in Unfavorable Thermal Environment*, New York, Wiley-IEEE Press, 2005.

[4] D. Tasić, *Osnovi elektroenergetske kablovske tehnike*, Niš, Elektronski fakultet, 2001.

[5] “Calculation of the Cyclic and Emergency Current Ratings of Cables, Part 1: Cyclic Rating Factor for Cables up to and Including 18/30 (36) kV”, IEC Std. 60853-1, 1985.

[6] “Calculation of the Cyclic and Emergency Current Ratings of Cables, Part 2: Cyclic Rating Factor of Cables Greater than 18/30 (36) kV and Emergency Ratings for Cables of All Voltages”, IEC Std. 60853-2, 1989.

[7] G. Mazzanti, “Analysis of the Combined Effects of Load Cycling, Thermal Transients and Electrothermal Stress on Life Expectancy of High Voltage ac Cables”, IEEE Trans. Power Delivery, vol. 22, no.4, pp. 2000-2009, 2007.

[8] G. Mazzanti, “The Combination of Electro-thermal Stress, Load Cycling and Thermal Transients and its Effects on the Life of High Voltage AC Cables”, IEEE Trans. On Dielectrics and Electrical Insulation, vol. 16, no. 4, pp. 1168-1179, 2009.

[9] M. Abramowitz, I. Stegun, *Handbook of Mathematical Functions*, Dover Publications, INC.,New York, 1972.

[10] ABB, *XLPE Land Cable Systems, User’s Guide*, rev. 5, 2010.

The Influence of the Geometry of the Inductor on the Depth and Distribution of the Inductively Hardened Layer

Maik Streblau[1], Bohos Aprahamian[2], Vladimir Shtarbakov[3], Hristofor Tahrilov[4]

Abstract – The induction hardening of ferromagnetic details is widely used for his high efficiency, versatility, quality of products and the ability to precisely control the heating process.
Disadvantage is that the majority of industrial induction systems are limited in capacity and frequency to the large variety of hardened details.
This requires the use of inductors with shape and dimensions, providing optimal application of the advantages of this method and increasing the efficiency.
To investigate the influence of the geometry of the inductor on the distribution and depth of the hardened layer is necessary to analyze the electromagnetic and thermal processes in the detail.
For this purpose a computer model of the system inductor-detail is developed. Metallographic analysis on pre-hardened specimen of ferromagnetic steel was performed and the dimensions of the hardened layer were reported.

Keywords - induction hardening, hardened layer, modeling the system inductor-detail.

I. Introduction

The hardening of internal cylindrical surfaces of ferromagnetic details require to achieve a rate of heating ensuring uniform hardened layer in depth of the detail [2-6]. To meet these requirements is necessary to create an appropriate distribution of the temperature field in the volume of the detail.

The reason for the variability is caused by the electromagnetic field distribution, which is associated with proximity and annular effects [1] and the influence of the boundary effects [7].

The purpose of this paper, based on the theoretical model proposed in [8], is to present a specific solution to design the shape of the inductor used to heat the inner surface of the cylindrical sleeve type detail - Figure 1, with a composition of the material referred in Table I.

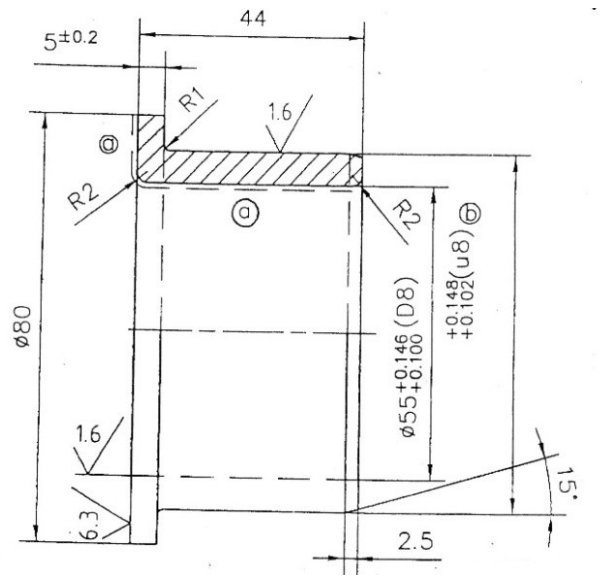


Fig.1. Overall dimensions of the detail

Table I
Chemical composition of steel type C50 EN 10 083-2

C	Si	Mn	Cr	S	P	Cu	Ni
0,47-0,55	0,17-0,37	0,50-0,80	0,25	0,04	0,04	0,25	0,25

The heating and hardening of the detail shown in Figure 1, is carried through an inductor with ferromagnetic core - Figure 2, powered by a tube generator, with duration of the process 10 s.

Fig.2. Inductor with ferromagnetic core.

[1]Maik Streblau is with the Faculty of Electrical Engineering, Technical University – Varna, 1 Studentska str., Varna, Bulgaria, e-mail: streblau@yahoo.com.
[2]Bohos Aprahamian is with the Faculty of Electrical Engineering, Technical University – Varna, 1 Studentska str., Varna, Bulgaria, e-mail: bohos@abv.bg.
[3]Vladimir Shtarbakov – Engineer in "METAL" PLC, 9000 Varna v_shtarbakov@yahoo.com
[4]Hristofor Tahrilov is with the Faculty of Electrical Engineering, Technical University – Varna, 1 Studentska str., Varna, Bulgaria, e-mail: h.tahrilov@gmail.com.

The quality of hardening is determined by metallographic analysis - Figure 3 and Figure 4.

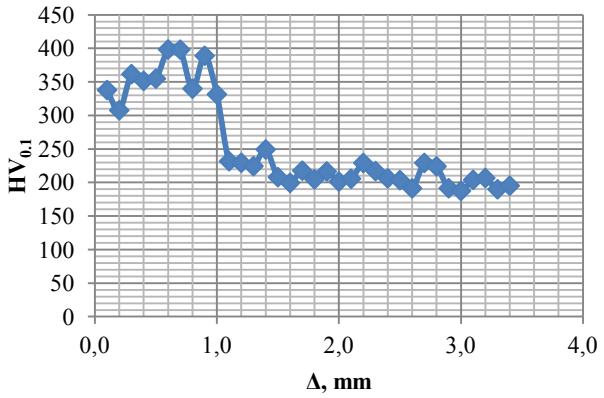


Fig..3. Change of hardness in the depth of the detail.

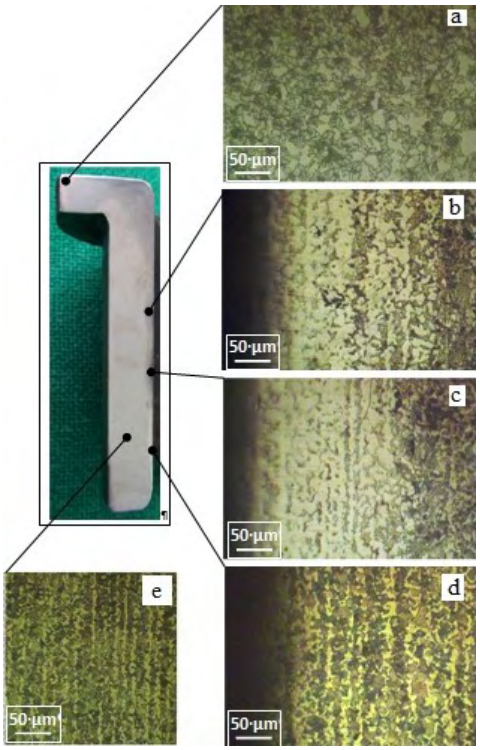


Fig.4. Structural condition of the detail after the process of heat treatment:
a - troostito-sorbite structure in the area with intensive induction and respectively thermal effects;
b, c - a transitional area between areas with concentrations of magnetic field lines and respectively with thermal effects on the material and the base;
d,e - basic textured ferrite-pearlite structure.

From the results presented in Figure 3 and 4 it is found that after the heat treatment, the resulting structure is non-heterogeneous along the axis of the detail, which determines insufficient degree of the necessary hardening, defined by the technical documentation.

This requires further research to clarify the shape and dimensions of the inductor.

II. Theoretical Investigation

A study using axial symmetric model presented in Figure 5, is performed.

The inductor Ω2 is without ferromagnetic core and is composed of four coils made of profiled wire.

The detail Ω3, subject to heat, is concentrically located to the inductor.

The size of the air field Ω1, encircling the inductor-detail system is consistent with the distribution of the magnetic field.

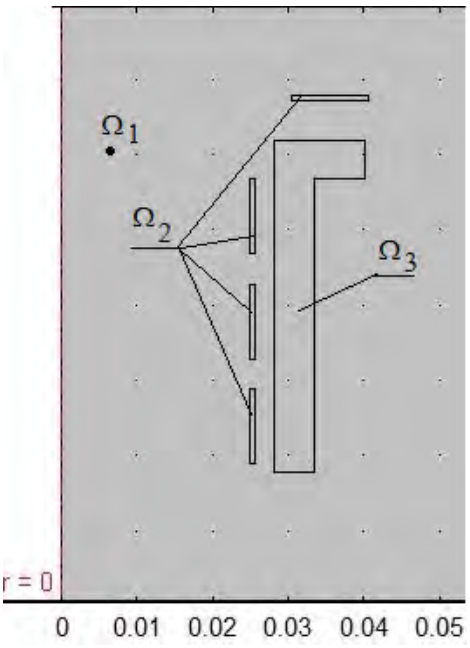


Fig.5.The system inductor-detail

With the so prepared model are conducted multiphysics model analysis – quasi steady state electromagnetic (1) and transient thermal (2) analysis:

$$\nabla \times \left(\frac{1}{\mu(H,T)} \times \nabla \times \dot{A}\right) + J \cdot \omega \cdot \gamma(T) \cdot \dot{A} = \frac{\dot{U}_{coil}}{2 \cdot \pi \cdot r} \quad (1)$$

$$\rho(T) \cdot c(T) \cdot \frac{\partial T}{\partial t} = \nabla(\lambda(T) \cdot \nabla T) + \frac{1}{2} \cdot \gamma(T) \cdot \omega^2 \cdot \dot{A}^2 \quad (2)$$

In these equations $\dot{A}$ is the vector magnetic potential, $H\,(A/m)$ is the strength of magnetic fields, $j = \sqrt{-1}$, $\mu\,(H/m)$ is the magnetic permeability, $\omega(rad/s)$ is the circular frequency, $\gamma(S/m)$ is the electric conductivity,

$\dot{U}_{coil}(V)$ is the voltage in the induction coil, T (K) is the temperature, $\rho(kg/m^3)$ the specific density, $c\ (J/kg.K)$ is the specific thermal capacity, t (s) is the time, $\lambda\ (W/m \cdot K)$ is the thermal conductivity.

The results for the distribution of the electromagnetic and thermal fields are presented in Figure 6 and Figure 7.

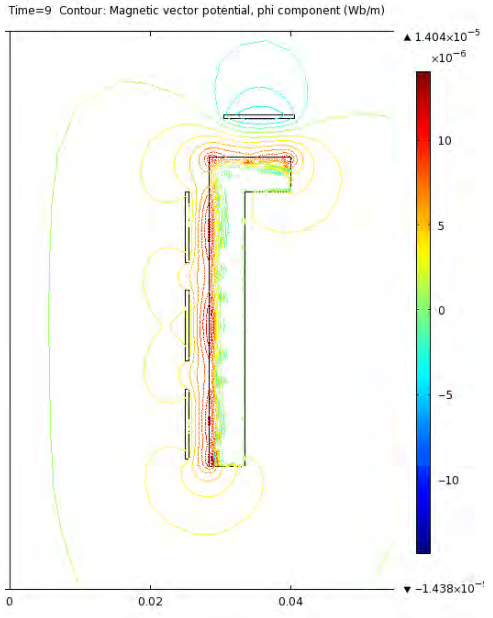


Fig.6. Distribution of the electromagnetic field.

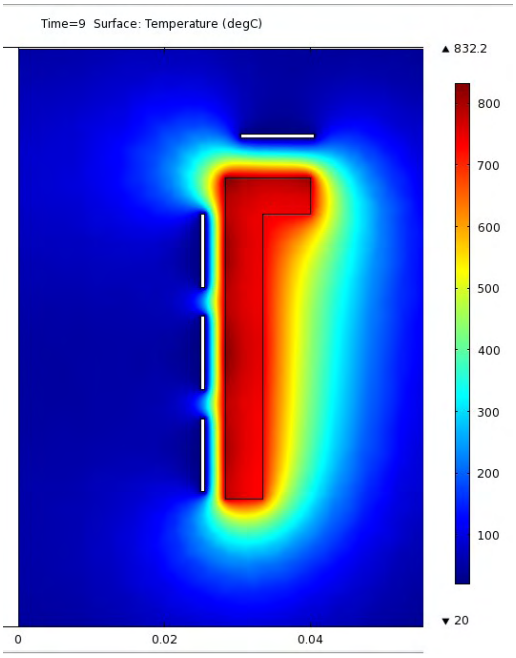


Fig.7. Distribution of the temperature field.

III. Conclusions

The following conclusions based on the presented results are found:

- The system inductor-detail with specially shaped inductor, corresponding to the hardened surface provides the necessary distribution of the electromagnetic field in the depth of the detail that determines the appropriate temperature distribution along his longitudinal surface;

- The results show that relatively small sizes and distance between the inductor and the detail require the use of profiled wire for making the inductor. This determines the relatively better electromagnetic connection and a more even distribution of the electromagnetic field opposite the wires;

- The lack of internal core of the inductor set loss reduction, respectively redistribution of the active power and increases the ability to reduce the number of turns of the inductor. Accordingly, this leads the increasing of the magnetic flux and power emitted in the detail that provides high-speed heating.

Acknowledgement

This paper is developed in the frames of project „Improving the energy efficiency and optimization of the electrotechnological processes and devices", № MU03/163 financed by the National Science Fund.

References

[1] Тодоров Т., Мечев И., Индукционно нагряване с високочестотни токове, Техника, София, 1979.
[2] Hoemberg D., Induction heat treatments – modeling, analysis and optiomal design of inductor coils, Habilitationsschrift, TU-Berlin, 2002.
[3] Альтгаузен А.П., Смелянский М.Я. - Электротермическое оборудвание., Энергия, Москва, 1967.
[4] Ставрев, Д., Ст.Янчева, Сл.Харизанова, Технология на термичното обработване, ВМЕИ- Варна, 1989.
[5] Кувалдин А.Б., Индукционный нагрев ферромагнитной стали, Энергоатомиздат, Москва, 1988.
[6] George E. Totten, Steel Heat Treatment – Equipment and Process Design, Taylor&Francis Group, 2007.
[7] Немков В.С., Б.Б.Демидович, Теория и расчет устройств индукционного нагрева, Энергоатомиздат, Ленинград, 1988.
[8] Aprhamian B., M.Streblau, Modeling of electromagnetic and thermal processes of high – frequency induction heating of internal cylindrical surfaces of ferromagnetic detail, ICEST, 2011

Examination of Frequency Controlled Asynchronous Drives at Variable Load Torque - Laboratory Simulator

Vasil Dimitrov[1]

Abstract – **Contemporary electrical transport vehicles are designed on the base of power drives controlled by highly efficient devices and microprocessor safety and control systems. A laboratory simulator, based on the newest technologies, enables the examination of both open loop motion control and vector control on a drive in case of inconstant load torque.**

Keywords – **asynchronous drives, frequency control, characteristics.**

I. INTRODUCTION

The development of the energy and transport equipment entered into a new stage during the past years. Regenerative converters (based on IGBT technology) and energy-saving motors are introduced into many modern trams, trolleybuses, underground vehicles and locomotives. The fast technological progress entails higher requirements of education quality. The bachelor and master programmes on Power Engineering and Electrical Equipment in the Todor Kableshkov University of Transport in Sofia prepares highly qualified experts in the respective fields of the transport and industry. These fields are attractive areas for Bulgarian and foreign investment for progress, modernization and expert education. The lecturers realize the necessity of training of skilled workers for electrical transport and power engineering needs. Therefore, a modern laboratory simulator built on contemporary devices was created (Fig.1). It includes an energy-saving induction motor AD (2,2 kW), controlled by a Sinamics G120 frequency converter [1]. A SG synchronous generator (30V, 60A) and resistors simulate the load of the motor. An HTL encoder can be used for closed loop motion control on the drive. A WINDOWS-based computer system is used for configuring and data storage. The necessary configuration software STARTER is installed on this computer. It can be used for setting-up the Sinamics G120 frequency converter.

II. BASIC PRINCIPLES OF THE FREQUENCY CONTROL ON ASYNCHRONOUS DRIVES

As it is known, in order to regulate the speed of asynchronous motors a change of frequency of the supply voltage (frequency control) is necessary [2]. The stator winding is fed with voltage with adjustable frequency f_1, that causes changes of the synchronous speed ω_o defined by the following equation:

$$\omega_o = \frac{2.\pi.f_1}{p} \quad (1)$$

where p is the number of the motor pole pairs.

At the same time, however, an amendment of the voltage amplitude is required, determined by flux saturation. Ignoring the relatively small voltage drop in the stator circuit, the following relation exists:

$$U_1 \sim E_1 = 4{,}44.w_1.k_{w_1}.\Phi.f_1 = c_1.\Phi.f_1 \quad (2)$$

where: U_1 – supply voltage; E_1 – electromotive voltage of the stator winding; w_1 – number of turns of stator winding; k_{w1} – coefficient of the filling of the stator winding; Φ – flux; c_1 – coefficient: $c_1 = 4{,}44.w_1.k_{w1}$.

Eq. (2) shows that the reduction in frequency while maintaining the nominal supply voltage is accompanied by a strong increase in magnetic flux Φ and saturation of the machine. Therefore, when reducing f_1 it's necessary to reduce the amplitude of the voltage to meet the condition $\Phi = const$.

There are several techniques for speed and torque control for inverters with induction motors. These techniques can be roughly classified as follows:

- V/f characteristic control (known as: V/f control);
- Field-orientated closed-loop control technique (known as: Vector Control).

The simplest speed control represents proportional V/f characteristic. In this case the stator voltage of the induction motor is adjusted proportionally to the stator frequency. This technique has proven itself for a wide range of basic applications, where the load is approximately constant.

There are several versions of the V/f characteristic:

- Linear (standard case);
- Square-law characteristic (f^2 characteristic), which takes into consideration the torque characteristic $M=f(\omega^2)$ of the motor load (e.g. fan/pump). This is an energy saving characteristic as the lower voltage also results in lower currents and losses;
- Programmable characteristic, which takes into consideration a specific torque characteristic of the driven load;
- Flux current control (FCC), which can give a more efficient and better load response than other V/f modes because the FCC characteristic automatically compensates the voltage losses of the stator resistance for static (steady-state) or dynamic loads. This is used especially for small motors which have a relatively high stator resistance.

[1] Vasil Dimitrov is a lecturer with Faculty of Communications and Electrical Equipment, Department of Electric Power Supply and Electrical Equipment in Transport at the Todor Kableshkov University of Transport, 158 Geo Milev Str., 1574 Sofia, Bulgaria. E-mail: vdimitroff@abv.bg

The goal of the V/f control is to keep the flux Φ constant in the motor. In this case, Φ is proportional to the magnetizing current I_μ and the ratio between voltage V and frequency f:

$$\Phi \sim I_\mu \sim U_1 / f_1 \quad (3)$$

The torque M, developed by induction motors, is proportional to the vectorial product of flux and current:

$$M = \overline{\Phi} * \overline{I} \quad (4)$$

In order to generate the maximum possible torque from a given current, the flux must be held constant at its nominal value. Therefore, the value of the magnetizing current must be constant even if the stator frequency changes. This can be achieved approximately if the stator voltage U is changed proportionally to the stator frequency.

Vector Control significantly improves the torque control. The Vector Control is based on the fact that for a specific load situation or required torque, the required motor current is regulated with respect to the motor flux so that the appropriate torque is obtained. If the stator current is emulated in a circulating coordinate system, linked with the rotor flux Φ, then it can be broken down into the flux-generating current component i_d in-line with the rotor flux and into a torque-generating current component i_q, vertical to the rotor flux. These components are corrected to track their set points in the current controller using their own dedicated PI controllers and are equal to the set points in steady-state operation. Then the component i_d is proportional to the flux Φ and the torque is proportional to the product of i_d and i_q.

When compared to V/f control, Vector Control has the following advantages: stable during load and set point changes, better control performance, better noise/disturbance characteristics, the motor and braking torque are controlled independently of the speed, accelerating and braking are possible with a maximum adjustable torque.

III. METHODS FOR TESTING THE CHARACTERISTICS OF FREQUENCY CONTROLLED ASYNCHRONOUS DRIVE AT VARIABLE TORQUE

Methods for examination of two aspects of frequency controlled asynchronous drives are developed:

- Determination of the static mechanical and electro-mechanical characteristics of asynchronous electric drive with frequency control;
- Determination of the dynamic characteristics at a load torque change.

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 raised with a jerk switching the circuit closers K_1 and K_2 by pushing buttons B_1 and B_2. Then the load torque can be reduced with a jerk by pushing button S.

The laboratory simulator enables examinations of many cases of open loop or closed loop motion control in inconstant load torque. The Sinamics G120 frequency converter offers several versions of the V/f characteristic control without speed feedback: it can be linear (parameter P1300=0), flux current control (parameter P1300=1), square-law characteristic (parameter P1300=2) or programmable characteristic (parameter P1300=3) [3]. The closed-loop vector control with speed encoder can be examined, too (parameter P1300=21). This frequency converter offers also many other methods of control: the user can fix the output voltage of the inverter, independently of the frequency, or there is a typical setup for textile applications. Even a sensorless vector control without feedback may be used (parameter P1300=20) if very high speed accuracy is not required and a defined torque has not to be maintained for low speeds (below approximately 10% of the rated motor frequency) [1].

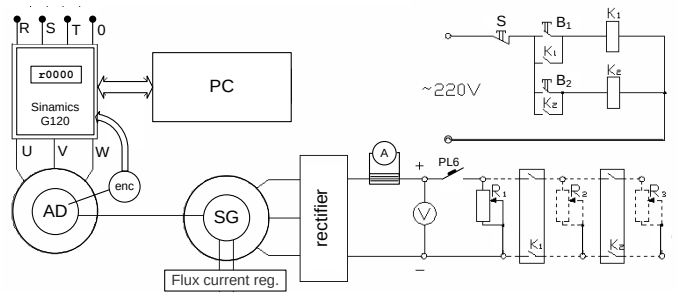


Fig. 1. Laboratory simulator

A. Determination of the static mechanical and electro-mechanical characteristics of asynchronous electric drives

The static mechanical and electromechanical characteristics of an asynchronous electric drive represent the speed as functions of the torque and the motor current:

$$\omega = f(M),\ \omega = f(I)$$

Interesting are also the characteristics that show the change of energetic parameters (power factor $\cos\varphi$ and efficiency η) of the drive at the load change:

$$\cos\varphi = f(M),\ \eta = f(M)$$

The static characteristics are examined at a set frequency by changing the load. The first point of the characteristics is obtained at the minimum load torque (PL6 on Fig. 1 is opened). During the examination, the following variables are measured by the Sinamics G120 frequency converter: actual inverter output frequency f (parameter r0024), rms voltage applied to motor U_{AD} (parameter r0025), rms value of motor current I_{AD} (parameter r0027), electrical torque M (parameter r0031), power P_{AD} (parameter r0032), power factor (parameter r0038) and the actual speed detected by encoder n (parameter r0061). At the same time, also the generator voltage U_g and current I_g are measured respectively by the voltmeter and ammeter mounted in the load circuit. The generator speed ω_g is measured by a digital tachometer. In order to get the second and further points of the characteristics, PL6 must be closed and the load torque must be changed by the above listed techniques (M_{rated} = 14,8 Nm).

After the measuring is finished, the frequency can be changed and all examinations may be repeated.

Thus a family of characteristics may be obtained (Figs 2 and 3).

The drive efficiency and the load torque may be calculated:

$$\eta = \frac{P_g}{P_{AD}} = \frac{U_g.I_g}{P_{AD}} \quad (5)$$

$$M_c = \frac{P_g}{\omega_g} = \frac{U_g.I_g}{\omega_g} \quad (6)$$

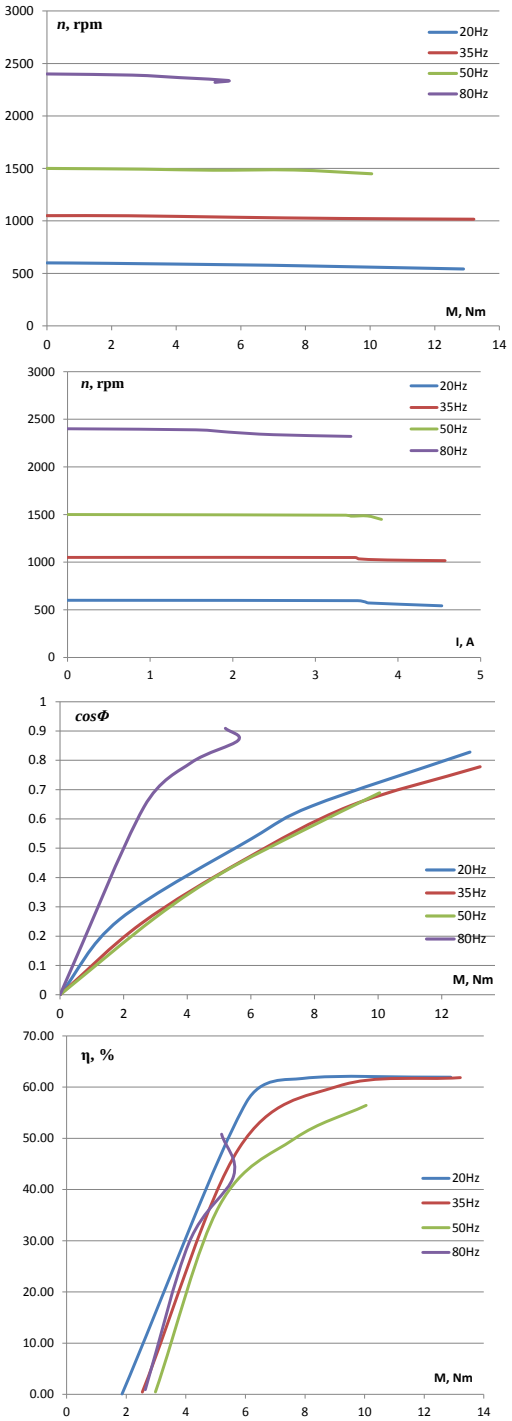


Fig. 2. Static characteristics at linear V/f control

B. Determination of the dynamic characteristics of asynchronous electric drives

The software STARTER provides additional options for control on the drive. A program which can trace the parameters of the frequency converter in the real time and write them into the computer memory is added. It can trace up to 30 parameters simultaneously, and then export the obtained data into an Excel sheet and create graphs. Many characteristics at motor starting and subsequent load changing are obtained at various control modes (Fig. 4a-f).

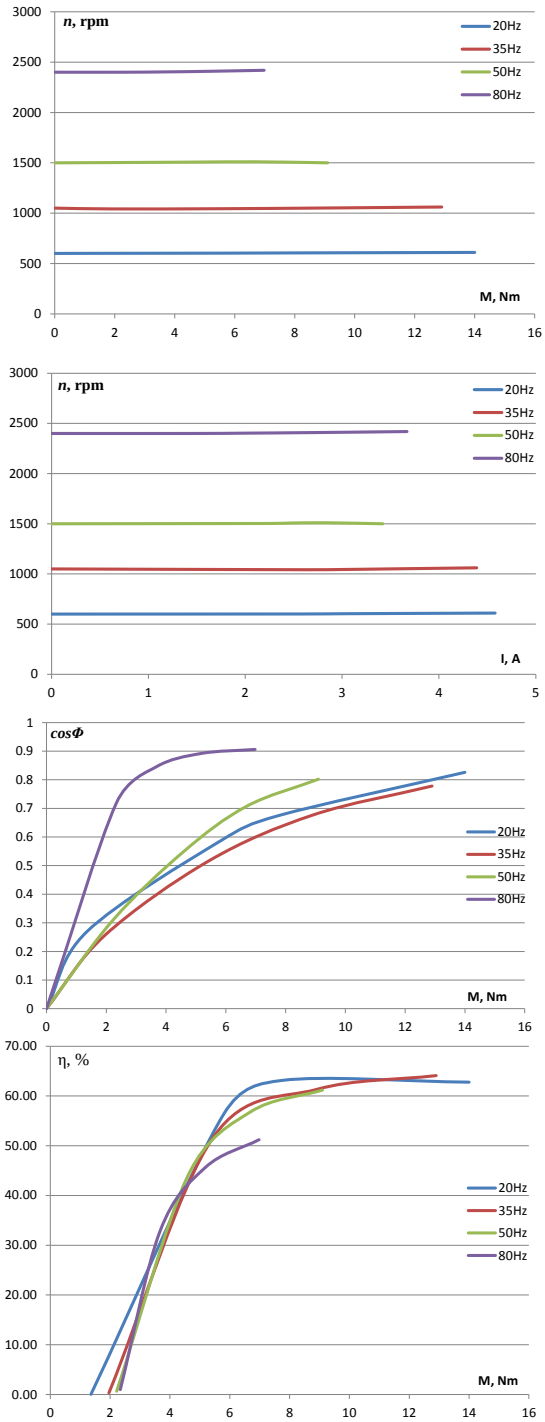
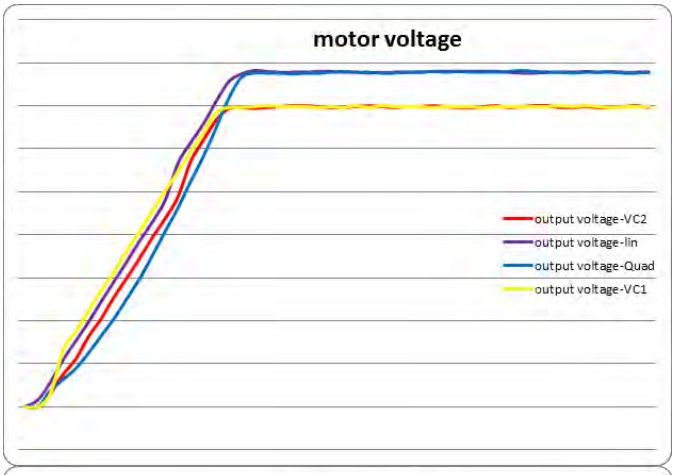
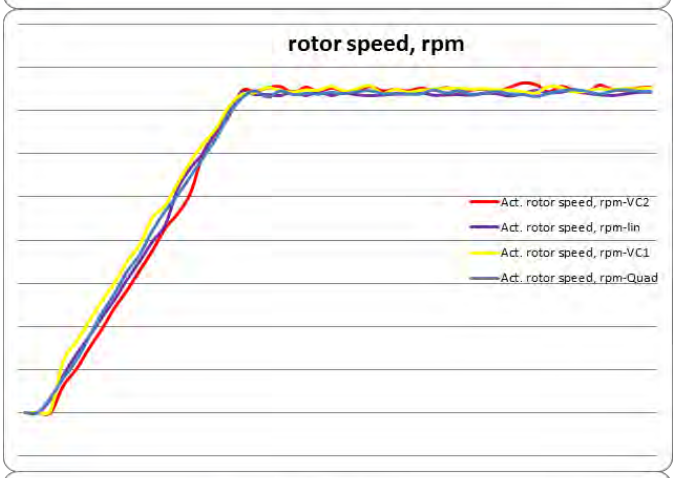


Fig. 3. Static characteristics at Vector control





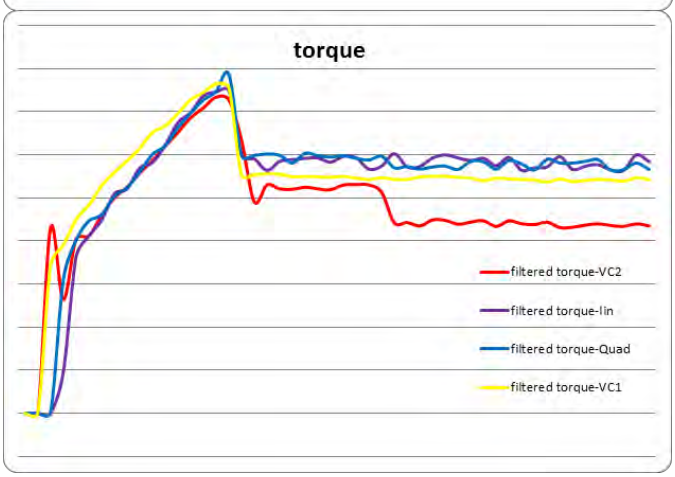
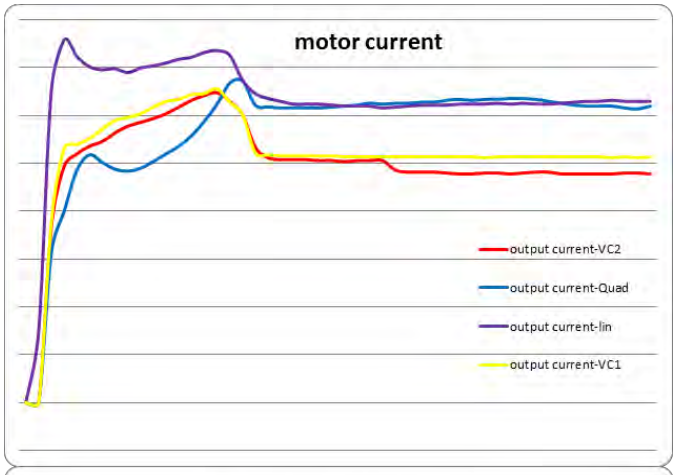


Fig. 4a, b, c. Dynamic characteristics



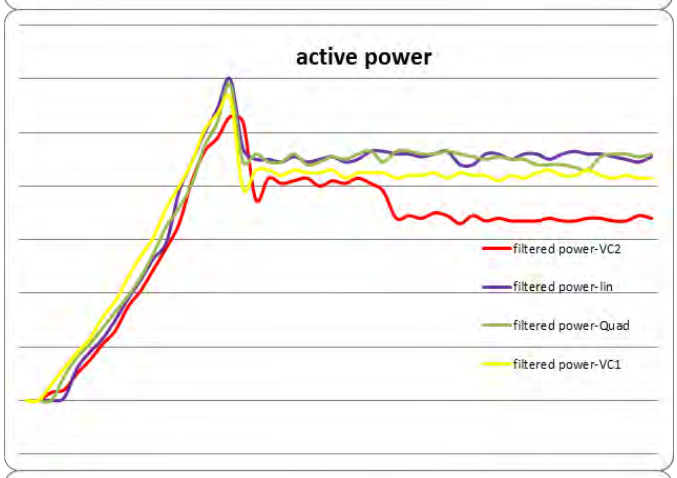


Fig. 4d, e, f. Dynamic characteristics

IV. Conclusion

In this paper an experimental verification of the principles of the frequency control on the asynchronous drives is done. The Vector Control provides a stable speed irrespective of any changes of the load. It achieves almost the same moment, but at much lower current and voltage compared to the V/f characteristic control. This leads to higher values of the power factor and the efficiency. Therefore, the vector control must be used in all applications where the load varies widely, as in electric vehicles. If the load torque is approximately constant or varies in a narrow range the V/f control gives good results.

The square-law characteristic offers higher values of the power factor at low frequencies due to less flux generating current. It should be used to drive fans, pumps and other centrifugal mechanisms.

References

[1] SINAMICS G120 - Function Manual, Siemens, 2007.
[2] Boldea I, S.Nasar, Electrical Drives, CRC/Taylor&Francis,2006
[3] SINAMICS G120 - Control Units CU240E, CU240S, Parameter Manual, Siemens, 2008.

Multi Leveled Hierarchical Approach for Monitoring and Management Information Systems Construction

Emiliya Dimitrova[1] and Galina Cherneva[2]

Abstract – **Dynamic of complex technical systems state is shaped in this working out. On this basis a model of two leveled hierarchical system for active power controlling in an energy system is proposed. Conclusions are made regarding the advantages of multi leveled management.**

Keywords – **Supervisory Control and Data Acquisition (SCADA), automatic management, hierarchical system.**

I. INTRODUCTION

Supervisory Control and Data Acquisition systems (SCADA) [1,2] in some important production sectors, such as production and distribution of electricity, transport management, etc., are crucial for their correct and reliable functioning.

Electric power system represents a complex corporative structure with varied equipment, enormous number of sub-systems and complex connections between them. Its management is impossible without quick and due decisions minded the high requirements for reliability and safety of functioning, because any failures in the energy system could have serious economic and social outcomes. All this requires development and improvement of new approaches for organization and management of the energy system.

In this paper, based on the established objective function, a model of two leveled hierarchical system for active power controlling in energy system is proposed. Conclusions are made regarding the advantages of multi leveled approach to information systems for monitoring and management.

II. MODEL OF DYNAMIC OF COMPLEX TECHNICAL SYSTEMS STATUS

Each complex technical system could be represented as a combination of n interacting in between sub-systems. Each state x_i of sub-system i is examined as a point in the vector space X_i. Let $X_0 \in X_i$ is a closed area of the space, where p. x_i is moving in the process of system functioning, i.e. X_0 is an area of the admissible working regimes, then $x_i \in X_0$ for $i \in [1, n]$ expresses the current state of the system.

Let $u(t) \in U$ and $v(t) \in V$ are respectively the entrance impacts and the exist signals of the system, which represent a function of time t.

The system transition from one state $x_i(t_1)$ into another $x_i(t_2)$ for $t_2 > t_1$ determines its dynamic. It could be realized under internal managing signals influence, as well as in consequence of internal disturbances.

Let in the initial moment t_0 the system is in state: $x_i(t_0) \in X_0$.

If the system state modification is a result an external signal influence, then it is described with the equation:

$$x_i(t) = F[t, t_0, x_i(t_0), u(t)], \quad (1)$$

where F is the operator of the transition and it is determined in accordance with the specific case.

The signal

$$v(t) = F[t, x_i(t), u(t)] \quad (2)$$

is set up at the system exit.

Depending on the specific task for management, functional dependency F is defined and equation (2) is calculated on condition that dependency (1) is performed. The problem could be solved by one – computer configuration on the basis of so called full interlocking or on the basis of multi-leveled hierarchical computer system. In the case of a hierarchical system, each higher management level is performed by radial connections with controlled systems of one type. The number of systems of higher level depends on the target functions complexity and the controlled systems. Each more complicated production process requires due decisions coordinated with other processes. The more complicated the target for management is the more information the managing body has to cover and process with the necessary precision and quick operation

III. TWO LEVELED SYSTEM FOR ACTIVE POWER CONTROLLING IN THE ENERGY SYSTEM

A classical example of multi leveled hierarchical system is the problem for active power controlling in the energy system.

The system is divided into n interacting in between sub-systems (areas) [2]. The borders of the respective areas are chosen in a way so that each of the sub-systems represent a separate company or entity (Fig. 1). Each area includes a row of generating stations and a great diversity of users. As far as only the exchange of electric power between areas is examined in this instance, we accept that each area is described with the following parameters: C_i - full loading in area i; X_i - active power, generated by elements in area i;

[1]Emilya Dimitrova is with T. Kableshkov University of Transport, 158 Geo Milev Str., Sofia, Bulgaria,E-mail: edimitrova@bitex.bg.

[2]Galina Cherneva is with T. Kableshkov University of Transport, 158 Geo Milev Str., Sofia, Bulgaria,E-mail: galja_cherneva@abv.bg.

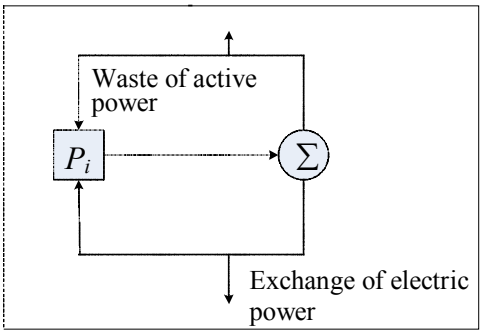


Fig. 1. Sub-system i

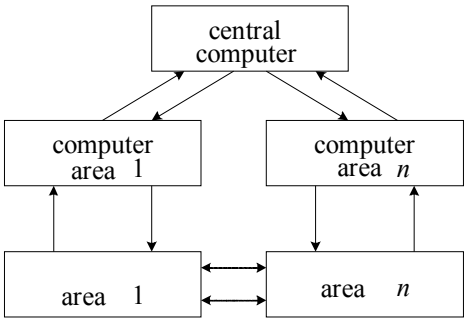


Fig. 2. Two leveled system

Y_i - wastage of active power in area i; U_i - exchange of power through the energy lines, connecting area i with the other areas.

In case of back load, the wastages in area i depend on the power, generated in the area itself, as well as on the exchange of power, i.e. $Y_i = F(X_i, U_i)$

The equation determining power balance could be written in the following mode:

$$F(X_i, U_i) + C_i - U_i + X_i = 0 \quad (3)$$

n equations of this type could be compared. Apart from that, the exchange of power between the areas has to be balanced, i.e.

$$U_1 + U_2 + ..U_n = 0 \quad (4)$$

Equation (2) describes the sub-processes and equation (4) describes their interaction.

On this basis the task for active power tracking in the united energy system is brought to definition of the power X_i and the volume of exchanges U_i, where the wastage of power Y_i is minimal.

The full wastage of power is:

$$Y = \sum_{i=1}^{n} F_i(X_i, U_i) \quad (5)$$

Then the optimal controlling of active power is brought to

$$Y = \sum_{i=1}^{n} F_i(X_i, U_i) \rightarrow \min \quad (6)$$

provided, the variables $X = (X_1, ...X_n)$ and $U = (U_1, ...U_n)$ answer the balance equations (3) and (4).

Task (6) could be worked out with the help of two leveled system, possessing organizational hierarchy. In this case, except for the central managing computer, each area has its own computer for working out the task for controlling. This organization is shown in Figure 2.

A question arises how the task for minimization to be distributed between the computers of different areas and the central computer.

An approach based on the so called principle for interactions forecast can be used for this purpose. This method is composed of the following.

The computer for *i*-area solves the task for minimization of $F_i(X_i, U_i^*)$ regarding X_i in the performance of condition (4), as value U_i^* is set by the central computer. The task of the central computer is determination of the supporting exchange level, i.e. U_1^*, ..., U_n^*.

If the exchange is performed at this condition, the local minimum will be at the same time global. When the difference between the supporting exchange level (defined by the central computer) and the real exchange come beyond the borders of the instructed limits, the new supporting exchange is determined through an iteration process between the central computer and different areas computers.

IV. Conclusion

In this paper, the objective function of the dispatching system for active power controlling in the energy system is defined. It is shown that it can be optimized by two leveled hierarchical management. Thus the system is less sensitive on changes of external influences such as changing conditions in one area causes changes in only one of solving (local) tasks.

Acknowledgement

This paper was supported by the project BG051PO001-3.3.06-0043 "Increasing, Improving and Extending the Scientific Potential of the University in Transport by Support to Development of PhD Students, Postdocs, Trainees and Young Researchers in the Field of Transport, Power Engineering and ICT in Transport" implemented within the Operational programme "Human Resources" funded by the European Social Fund.

References

[1] www.es-electro.ru
[2] Andonov A., G. Cherneva, Information systems in energetics, S., VTU, 2009

Improving Control System in the Sulfuric Acid Plant

Viša Tasić[1] , Dragan R.Milivojević[2], Vladimir Despotović[3] , Darko Brodić[4],

Marijana Pavlov[5]

Abstract – **The article describes improvments of the process control system in the sulfuric acid plant in Bor, Serbia. Manual data collection using instrumentation on command tables and panels is replaced by microcontroller based real-time control system. Monitoring and control of the process parameters were performed with the use of Programmable Logic Controller (PLC). The simplest control network consists of two nodes: one PLC and one workstation, used for visualization and interaction with the process from a remote location. Some hardware and software solutions developed for this particular control system, as well as configuration and topology of industrial network are emphasized.**

Keywords – **sulphur-dioxide, control system, industrial network, air pollution monitoring**

I. INTRODUCTION

The Municipality of Bor is located in the southeastern part of Serbia, close to the Bulgarian and Romanian borders. The area has been the major centre for mining and processing of copper and other precious metals since 1903. Air pollution is perceived as the main environmental problem in the Bor region. The main source of air pollution with SO_2 gas, heavy metals in particulate matter and aero sediments is the Copper Mining and Smelting Complex Bor (RTB Bor Company) which has been in operation for more than 100 years [1].

The technology for copper production in this copper smelter is outdated (classic pyrometallurgy with melting in furnaces and utilization of SO_2 gas in production of H_2SO_4 with relatively small degree of utilization <60%) which leads to environmental pollution caused by higher concentrations of SO_2 (as shown in Fig. 1). In the standard Pierce-Smith converter (as in the copper smelter in Bor), the off-gases are treated in electrostatic precipitator system to remove particulate matter, and in the sulfuric acid plant to remove SO_2. Control of SO_2 off-gas emissions from the smelter is performed in the sulfuric acid plant.

In order to improve the existed control system in the sulfuric acid plant, a new industrial PLC has been installed. Manual data collection using instrumentation on command tables and panels is replaced by microcontroller based real-time control system. Appropriate software application has been developed for monitoring of the process parameters in order to reduce the air pollution from the copper smelter facilities. Main objectives of the new control system are real-time data processing, data presentation (in the form of dynamic synoptic schemes, real time graphs and tables) and database management.

II. CONTROL SYSTEM HARDWARE

Department of Industrial Informatics, at the Mining and Metallurgy Institute Bor has designed an industrial PLC, named Microprocessor Measuring Station (MMS) [2], as shown in Fig. 2. Although MMS can be an autonomous system unit, it is used more frequently as a node in a simple network (entity), which contains MMS and PC workstation.

Fig. 1. Air pollution as a result of emissions from the copper smelter

[1] Viša Tasić is with the Institute of Mining and Metallurgy, Department of Industrial Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: visa.tasic@irmbor.co.rs,

[2] Dragan R. Milivojević is with the Institute of Mining and Metallurgy, Department of Industrial Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: dragan.milivojevic@irmbor.co.rs,

[3] Vladimir Despotović is with the University of Belgrade, Technical Faculty in Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia, e-mail: vdespotovic@tf.bor.ac.rs,

[4] Darko Brodić is with the University of Belgrade, Technical Faculty in Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia, e-mail: dbrodic@tf.bor.ac.rs,

[5] Marijana Pavlov is with the Institute of Mining and Metallurgy, Department of Industrial Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: marijana.pavlov@irmbor.co.rs

Fig. 2. The MMS configuration

Main characteristics of MMS (standard configuration) are: microcontroller Motorola 68HC11E, internal eight channel, 8-bit A/D converter, 64 analog inputs, 64 + 64 digital state signals (input + output) with mutual point (or independent), RS232 communication port, 48 (56) KB for data (RAM), 16 (8) KB for software (EPROM).

Applied technology in sulfuric acid plant; single way direct catalysis - single way direct absorption (even if it worked in optimum conditions) emits waste gas in concentration of SO_2 and SO_3 that are higher than the maximum permitted by regulations. Thus, reliable and timely information about process parameters are of great importance. Improvement of the control system has demanded the installation of new sensors, transmitters and actuators, as shown in Fig. 3. All output signals from the different types of transmitters (temperature, pressure, flow, vibration, electric power) had been connected to PLC.

Since the parameters to be measured were located in different factory halls, the signals from transmitters had to be concentrated at one place. The control room of the sulfuric acid plant - K2 was chosen as the appropriate location for this purpose, as shown in Fig. 4.

Fig. 3. Temperature and vibration transmitters mounted in a rack

Fig. 4. Control room in the sulfuric acid plant

Process parameters are imported to MMS as standard current (0-20 mA or 4-20 mA) or voltage signals (0-5V or 0-24V DC). MMS performs measurements and upon request, transmits results to PC workstation. The results are presented in a real-time or archived for the later analyzes. The sulfuric acid plant process parameters can be accessed from remote plants as well (i.e. Pierce-Smith converter plant and fluo-solid reactor plant) in order to have actual information necessary for their production process control. Each MMS with the associated PC is a node of the industrial computer network. Industrial computer network consists of several dislocated segments. In order to include all the required nodes into industrial computer network design, all network nodes and segments should be carefully planned and realized. Wherever it was possible, wireless connection had been realized.

III. CONTROL SYSTEM SOFTWARE

MMS can operate independently of monitoring computer and control the process itself (local control mode). It can also run as data logger, and store over 3000 data messages in local RAM, and later, when connection to a monitoring PC is established, transfers them to PC. EPROM of MMS holds residential software (firmware). It consists of executable versions of test, control, operational and communication software modules. Serial communications are performed using specially designed logic and communication protocol developed for this purpose [2]. The complex communication subsystem on both sides (MMS and PC) was developed for efficient entity functioning.

The special Supervising Control and Data Acquisition (SCADA) real-time application, named Process Control Program (PCP), is developed in order to support MMS functions and to perform data transfer, analyses and real-time result's interpretation [3]. It is based on a client/server architecture running on both master and remote stations, enabling integration in a complex distributed control and monitoring system.

PCP is developed using Microsoft Visual C++ [4]. PCP has a complex structure and consists of several modules. Main program modules are designed for:

- Communication with MMS,
- Data acquisition,
- Real-time data processing and result's presentation,
- Interaction with technological process according to appropriate algorithm and values of monitored parameters,
- Creation of reports,
- Data archiving,
- Off-line data processing,
- Database management.

The data collected are stored in the database, in an uncompressed or a compressed form. The file management class, including the functions for data archiving, is responsible for this task. Further processing consists of number of actions:

- Data processing for graphical display (dynamic screens, time trends, tabular form), see Fig. 5,
- Data processing for alarm system,
- Data processing for distribution over LAN,
- Forming of daily and monthly archives.

Interactions with the technological process are performed according to control algorithms, considering the actual values of measured parameters. The appropriate commands are generated and send to MMS, which are executed via actuators. All the data required, such as: measuring ranges, operating ranges, operation curves, working regimes, etc., are entered in the PCP algorithms. As an example, a working curve for the blower is shown in Fig. 4. This curve has been used in the control algorithm to regulate the amount of inlet gas from the smelter. This parameter (gas flow) is very important for the air pollution control. The amount of gas which can be absorbed depends of actual working regime of the smelter, meteorological conditions and available capacity of the sulfuric acid factory.

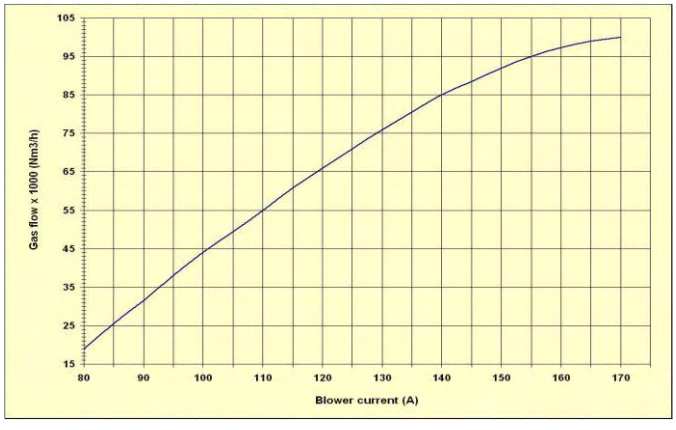


Fig. 4. Working curve for the blower in the sulfuric acid plant

Thus, the synchronization of all parts of the smelting complex is necessary. The control systems on real-time bases are of great importance for the proper control and managing of production process.

The PCP presents data using dynamic screens, graphs or tables. Fig. 5 is an example of dynamic screen for the blower. The measuring results are stored in a database for daily and monthly reports. The history of the process can be displayed in the same manner as in real-time. All data can be easily exported in the applications suitable for reports preparation (e.g. Microsoft Excel) for later analysis. The sulfuric acid factory's engaged power and spend energy is permanently measured and monitored with intention to be reduced as much as possible. The Fig. 6 presents monthly diagram of engaged power in the sulfuric acid plant.

Fig. 5. Synoptic screen for the blower in the sulfuric acid plant

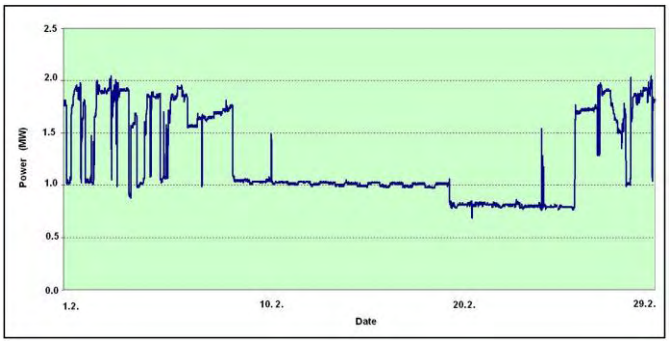


Fig. 6. Monthly diagram of engaged power in the sulfuric acid plant

IV. DISTRIBUTED CONTROL SYSTEM OVERVIEW

Realized MMS and PCP are applied mainly in metallurgy industry. Production plants are often dislocated in several kilometers wide area. To cover the main process parameters and monitor them in a real-time, it is necessary to realize a distributed control system (DCS). The complex decentralized industrial network is built for this purpose [5-6]. The basic network nodes are entities consisted of MMS and PC workstation (server in network configuration), running appropriate PCP. The remaining network nodes (clients in network configuration) acts as workstations, running a passive version of PCP named Remote Control Program (RCP), as

shown in Fig. 7. RCP is used for distant monitoring only, without having the possibility of sending commands to server (or MMS). It can only access databases stored on the server side and read the data. Data analysis, processing and presentation are performed locally, on client workstations [7, 8]. Client workstation is able to run as many RCP programs as needed at the same time (accessing servers in different plants). The network realized in practice consists of up to 10 servers and over 30 clients. Each production plant is covered by a server entity consisted of MMS and PC workstation, and interconnected with a number of clients for distant monitoring, as shown in Fig. 8.

Fig. 7. Workstation with the RCP client running (left), side by side with server executing the PCP (right)

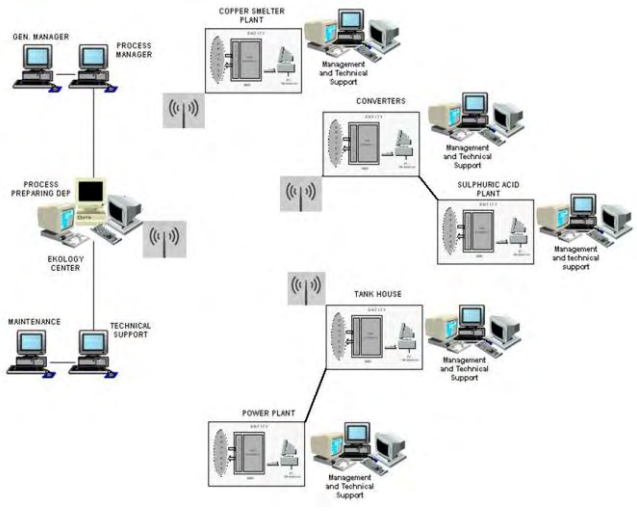


Fig. 8. Computer network in the copper smelter plants

CONCLUSION

Realized solution is a compromise between investments and expected effects of monitoring and control of the production process in the sulfuric acid factory. The realized control system showed good stability and resistance to external influences. The system response time is usually less than few seconds, depending on actual process demands. This can be considered as satisfactory response time for such kind of production process. The possibility of distant monitoring is very important for making the different business and production strategy decisions on time. Implemented network enabling managers to monitor the production process in real-time. Appropriate data analyses and creation of reports can be also performed on client side at any time.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education and Science of the Republic of Serbia under the Project TR33037 “Development and Application of the Distributed System for Monitoring and Control of Electrical Energy Consumption for Large Consumers".

REFERENCES

[1] V.Tasić, N. Milošević, R. Kovačević, N.Petrović, “The analysis of air pollution caused by particle matter emission from the copper smelter complex Bor (Serbia) ”, Chemical Industry & Chemical Engineering Quarterly, vol. 16, no. 3, pp. 219-228, 2010.

[2] D.R.Milivojevic, V.Tasic, “MMS in Real Industrial Network”, Information Technology and control, vol. 36, no. 3, pp. 318-322, 2007.

[3] D.R.Milivojevic, V. Despotovic, V.Tasic, M. Pavlov, “Process Control Program as an Element of Distributed Control System”, Information Technology and control, vol. 39, no. 2, pp. 152-158, 2010.

[4] B.Stroustrup, *The C++ Programming Language*, Addison-Wesley Pub Co, 3rd edition, 2000.

[5] V. Tasic, D. Milivojevic, M. Pavlov, V. Despotovic, “An Industrial LAN Applied in Copper Mining and Refining Complex in Bor”, ETRAN 2007, Conference Processedings CD, Herceg Novi, Montenegro, 2007. (in Serbian)

[6] D.Milivojevic and V.Tasic, “Some Software Elements of the Microprocessor Measuring Station”, Acta Electrotechnica et Informatica, vol. 7, pp. 64-68, 2007.

[7] B. Bamieh and P. G. Voulgaris, “Optimal Distributed Control with Distributed Delayed Measurement”, In Proceedings of the IFAC World Congress, 2002.

[8] D.Milivojević, V.Tasić, M.Pavlov and V.Despotović, “Synthesis of DCS in Copper Metallurgy”, ICEST 2007, Conference Proceedings, Book 2., pp.629-631, Ohrid, FYR Macedonia, 2007.

DLadder - an Integrated Environment for Programming PIC Microcontrollers

Viša Tasić[1] , Dragan R.Milivojević[2], Vladimir Despotović[3] , Darko Brodić[4],

Marijana Pavlov[5] and Vladan Miljković[6]

Abstract – **DLadder is a new integrated development environment designed for programming the Microchip PIC microcontrollers. It allows programming using ladder logic, interpretation and compilation of programs into PIC16 and PIC18 native code. Debugger and the real-time simulation are also included, which allows displaying the contents of a microcontroller ports and memory locations in the real-time. Furthermore, the state of the I/O ports and A/D conversion results can be memorized in the database. Benefit of this development environment is the easy way of 'visual' programming in low level code. This paper describes briefly some elements of DLadder development environment.**

Keywords – **microcontroller, programming, ladder, software**

I. INTRODUCTION

Ladder diagram represents a program in ladder logic. Ladder logic is a programming language that represents a program by a graphical diagram based on the circuit diagrams of relay logic hardware. When a programmable logic controller (PLC) is used primarily to replace relays, timers, and counters, it's hard to beat the simplicity and usefulness of ladder diagram programming [1]. The name is based on the observation that programs in this language resemble ladders, with two vertical rails and a series of horizontal rungs between them. The logic in a ladder diagram typically flows from left to right. The diagram can be divided into sections called rungs as shown in Fig. 1.

[1] Viša Tasić is with the Institute of Mining and Metallurgy, Department of Industrial Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: visa.tasic@irmbor.co.rs,

[2] Dragan R. Milivojević is with the Institute of Mining and Metallurgy, Department of Industrial Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: dragan.milivojevic@irmbor.co.rs,

[3] Vladimir Despotović is with the University of Belgrade, Technical Faculty in Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia, e-mail: vdespotovic@tf.bor.ac.rs,

[4] Darko Brodić is with the University of Belgrade, Technical Faculty in Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia, e-mail: dbrodic@tf.bor.ac.rs,

[5] Marijana Pavlov is with the Institute of Mining and Metallurgy, Department of Industrial Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: marijana.pavlov@irmbor.co.rs

[6] Vladan Miljković is with the Institute of Mining and Metallurgy, Department of Informatics, Zeleni bulevar 35, 19210 Bor, Serbia, e-mail: visa.tasic@irmbor.co.rs

Each rung typically consists of a combination of input instructions. These instructions lead to a single output instruction; however, rungs containing function block instructions may be more complicated.

Ladder logic has contacts that make or break circuits to control coils. Each coil or contact corresponds to the status of a single bit in the programmable controller's memory. Unlike electromechanical relays, a ladder program can refer any number of times to the status of a single bit, equivalent to a relay with an indefinitely large number of contacts. So-called "contacts" may refer to physical inputs to the programmable controller from physical devices such as pushbuttons and limit switches via an integrated or external input module, or may represent the status of internal storage bits which may be generated elsewhere in the program.

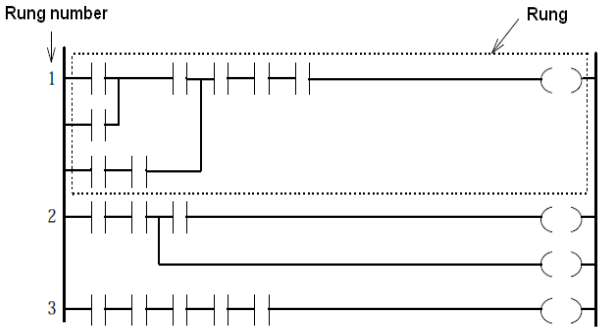


Fig. 1. The basic elements of the ladder diagram

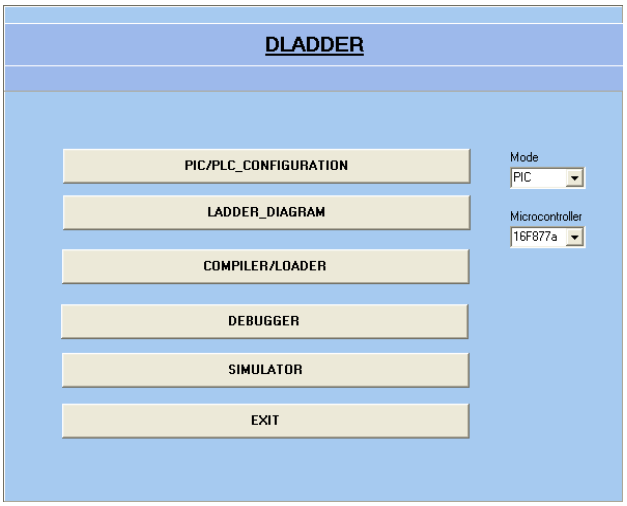


Fig. 2. DLadder main window

Ladder notation is best suited to control problems where binary variables are only required and interlocking and sequencing of binary are the primary control problems. Analog quantities and arithmetical operations are not suitable to express in ladder logic and each manufacturer has different ways of extending the notation for these problems. As microprocessors have become more powerful, notations such as sequential function charts and function block diagrams can replace ladder logic for some limited applications. Very large programmable controllers may have all or part of the programming carried out in a dialect that resembles BASIC or C or other programming language with bindings appropriate for a real-time application environment.

II. DLADDER INTEGRATED DEVELOPMENT ENVIRONMENT (IDE)

DLadder is a new integrated development environment (IDE) developed in Borland Delphi 7, and designed for certain Microchip PIC microcontrollers [2], mainly for the purpose of development of applications for control and data acquisition of industrial processes. It allows programming using ladder logic, and compilation of programs into PIC16 and PIC18 native code. However, the PIC microcontroller with this software is able to perform the PLC functions, by some additional features and predefining functional blocks. A PLC is a modular device programmable by using ladder diagrams. It internally uses a microcontroller to handle all input, output and logic scans.

DLadder allows writing programs for microcontroller basic configuration only, or for expanded configurations with a number of additional new input or output modules. Debugger and the real-time simulator are also designed, which allows displaying the state of a microcontroller ports and contents of memory locations in the real-time. Thus, DLadder offers an easy way of 'visual' programming in low level code. DLadder's main window is shown in Fig. 2. It consists of several modules that will be briefly described below.

The main difference between PLC and microcontrollers is only the way of programming. Therefore, in the very beginning the mode of operation (basic configuration – PIC mode; expanded configuration – PLC mode) should be selected. Afterwards, parameters of the PIC/PLC_CONFIGURATION are configured based on the selected mode, as shown in Fig. 3.

A. Ladder Diagram Editor

The option LADDER_DIAGRAM opens the editor for drawing ladder diagrams. Writing a ladder diagram is performed by selecting the required elements from the object toolbar (shown in Fig. 4) and setting them in the appropriate position on the screen. Each window in ladder editor consists of eight rows and 11 columns. The last column in each row is reserved for placing the coils and function objects.

The function objects supporting the arithmetic operations, logic operations, etc. Most objects do not require any special settings. However, it is necessary to enter the memory location to which the object is referenced. An example of a simple ladder program simulating the operation of the traffic lights (semaphore) is shown in Fig. 4. An appropriate electric wiring diagram for the given case realized with the PIC16F877A microcontroller [3] is given in Fig. 5.

Fig. 3. The PIC/PLC configuration window

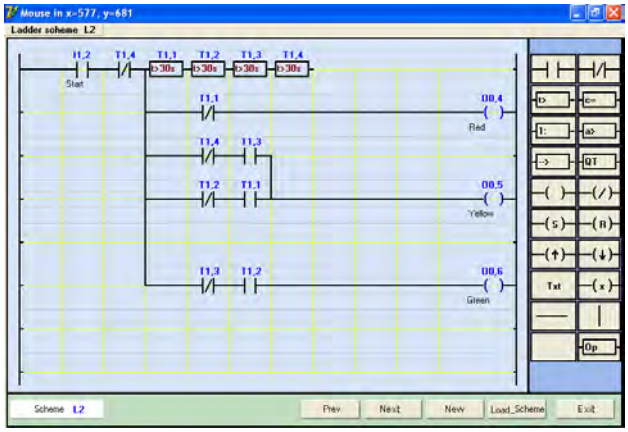


Fig. 4. The ladder diagram created in DLadder editor

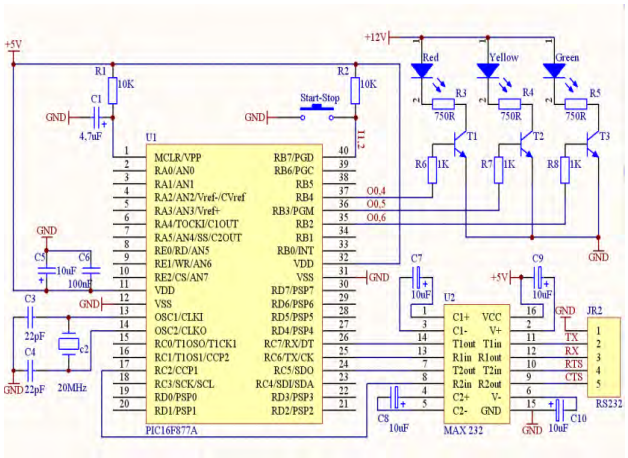


Fig. 5. The semaphore electric scheme realized with the microcontroller PIC16F877A

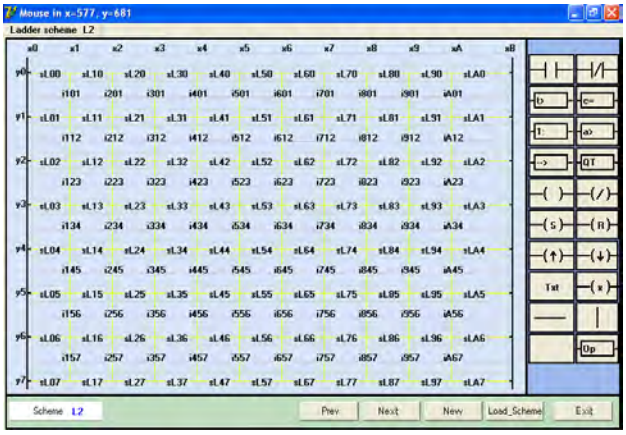


Fig. 6. The fields in DLadder database associated to the first screen of the ladder diagram

Larger ladder diagrams consist of several windows. At the bottom of the editor window there are buttons for scrolling through the ladder diagram windows, for opening of a new diagram or closing the editor. The existed ladder diagram can be loaded and modified by inserting or deleting some objects.

B. DLadder Compiler/Loader

Each ladder diagram is stored in the database as a series of symbols. The database records the position of each object on the ladder diagram, as well as their interrelations. List of fields in the database with their positions on the ladder diagram is shown in Fig. 6. The source code of the program, whether it is written in assembler or other programming language (C, Pascal, and ladder) has to be translated into machine language of the microcontroller. As a result of this operation an object file will be created in hexadecimal format (*. hex), which carries a series microcontroller instructions. Hence, in order to be ready for loading into the PIC microcontroller the ladder diagrams must be compiled and stored into the hex files in the prescribed form. Microchip has published a protocol for transmission of hex file into internal program memory of microcontrollers. An example of the structure of hex file is shown in Fig. 7.

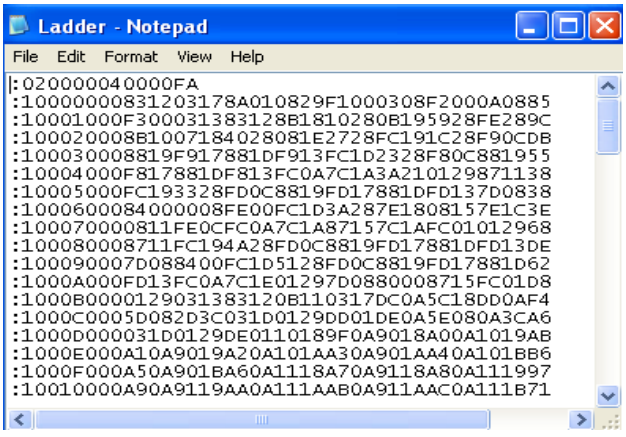


Fig. 7. The structure of the hex file prepared for programming the PIC microcontroller

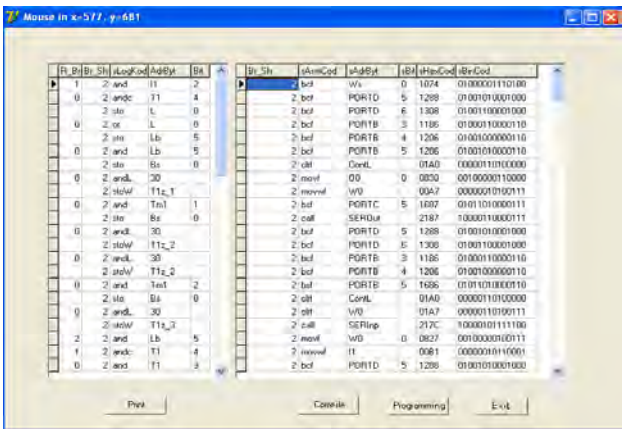


Fig. 8. DLadder's compiler/loader window

The option COMPILER/LOADER in DLadder main window opens compiler/loader window (shown in Fig. 8). Compiling of the ladder diagram starts after pressing the *Compile* button. The compilation of the ladder diagram is executed from left to right. The result of the compilation process is the code written into hex file. In a special case, when the PIC microcontroller on the development board is connected to PC, it is possible to program the microcontroller directly from DLadder. In such a case, by pressing the *Programming* button, the latter compiled program will be immediately loaded into the PIC microcontroller. DLadder uses both the USB and serial connection for file transfer, monitoring and control of the PIC program execution.

C. DLadder Debugger/Simulator

Once the code has been built and checked from the syntax point of view, it needs to be tested. Debugging is a process of finding and reducing the number of errors in a computer program. In order to test the code, we need some kind of software or hardware that will execute the PIC micro instructions [5].

DLladder's integrated debugger uses the internal in-circuit debug hardware of the target Flash PIC microcontroller to run and test the application program. When DEBUGGER option is selected in the main window, the application code is programmed into the PIC microcontroller's memory. A small "debug executive" program is loaded into the high area of program memory. Since the debug executive must reside in program memory, the application program must not use this reserved space.

The debug executive runs just like an application in program memory. It uses some locations on the hardware stack and file registers for its temporary variables. Special "in-circuit debug" registers in the target microcontroller are enabled. These allow the debug executive to be activated by the DLadder. The target microcontroller is held in reset by keeping the VPP/MCLR line low. DLadder will raise the VPP/MCLR line to allow the target microcontroller to run, starting from address zero and execute until the program counter reaches the breakpoint address previously stored in the internal debug registers.

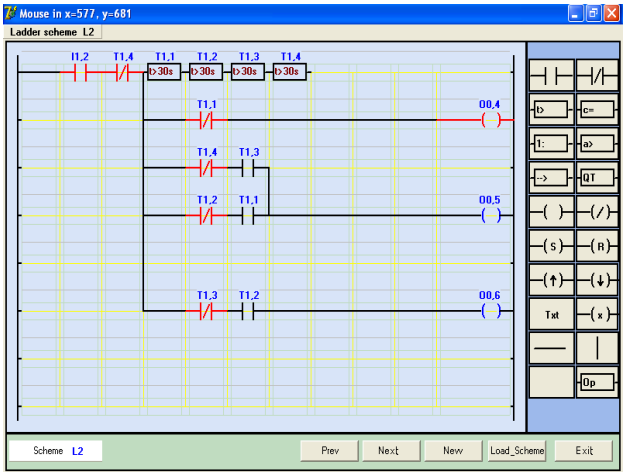


Fig. 9. DLadder debugger window

After the instruction at the breakpoint address is executed, the in-circuit debug mechanism transfers the program counter to the debug executive (much like an interrupt) and the user's application is effectively halted. DLadder IDE communicates with the debug executive via PGC and PGD lines [6, 7], gets the breakpoint status information and sends it back to the DLadder IDE. The DLadder IDE then sends a series of queries to get information about file register contents and the state of the CPU. These queries are ultimately performed by the debug executive.

Based on the described principle it is possible to monitor the status of certain elements of ladder diagram. The program appears on screen with the energized (true) branches highlighted (red color), as shown in Fig. 9, which makes it easy to debug.

The simulator is a software program that runs on the PC to simulate the state of input instructions of the PIC microcontrollers. When SIMULATOR option is selected in the main window, it allows monitoring of ladder diagram and changing (forcing) state of inputs during the program execution.

CONCLUSION

The primary motive for developing DLadder IDE is a desire to facilitate program writing for PIC microcontrollers using the concept of visual programming. Although there are a number of development environments for PIC microcontrollers on the market, not many of them allow writing the programs in ladder logic. DLadder IDE, which is still in the stage of continuous development, is shown to be reliable and easy-to-use. Thus, we believe that DLadder environment can be a useful tool for programming the PIC microcontrollers.

ACKNOWLEDGEMENT

This work was supported by Ministry of Education and Science of the Republic of Serbia under Project TR33037 "Development and Application of Distributed System for Monitoring and Control of Electrical Energy Consumption for Large Consumers".

REFERENCES

[1] W. Bolton, *Programmable Logic Controllers*, Fifth Edition, Newnes, 2009.

[2] M. Verle, *PIC microcontrollers - Programming in C*, Belgrade, Serbia, mikroElektronika, 2009. (in Serbian) http://www.mikroe.com/eng/chapters/view/14/chapter-1-world-of-microcontrollers/#c1v4

[3] Microchip Inc., *PIC16F87XA Datasheet* (10/31/2003) http://www.microchip.com/wwwproducts/Devices.aspx?dDocName=en010242

[4] V. Milanović, *PC Interfaces*, Lazarevac, Serbia, Elvod Print, 2005. (in Serbian)

[5] M. Verle, *PIC microcontrollers*, Belgrade, Serbia, mikroElektronika, 2007. (in Serbian)

[6] Microchip Inc., *PIC16F87XA Data Sheet 28/40-pin Enhanced FLASH Microcontrollers*, Microchip Technology Inc., 2003. www.datasheetcatalog.org/datasheets2/41/4109312_1.pdf

[7] Microchip Inc., *MPLAB ICD 2 In-Circuit Debugger User's Guide*, Microchip Technology Inc., 2005. http://ww1.microchip.com/downloads/en/devicedoc/51331b.pdf

Frequency Measurement Using Compact DAQ Chassis

Georgi Nikolov[1] and Boyanka Nikolova[2]

Abstract – **A CompactDAQ system consists of a chassis, C series DAQ modules, and a Windows host computer connected over USB, Ethernet, or Wi-Fi. CompactDAQ chassis control the timing, synchronization, and data transfer between a host computer and DAQ modules. The core technology in CompactDAQ chassis is known as the third generation of the system timing controller. It is possible to use counters for event counting, quadrature encoder measurement, pulse-width modulation, pulse train generation, or period and frequency measurement. These counters are advanced because they contain an embedded or onboard auxiliary counter. This is not directly accessible by the user, but it can be accessed by the DAQmx driver for some frequency measurements. The present paper is aimed to introduce the basic concepts, techniques and the underlying principles that constitute the realization of virtual universal counter. The introduced virtual device is based on counter module, which is build in the third generation of the system timing controller of CompactDAQ Chassis and graphical application development environment LabVIEW. The various methods of frequency measurement are considered in details, supported with error estimation and comparative analyze.**

Keywords – **Frequency Measurement, Graphical Programming, Reciprocal Technique, Virtual Instrumentation.**

I. INTRODUCTION

An important electrical quantity with no equivalent in direct current circuits is frequency. Frequency measurement is very significant in many applications of alternating current, for data transmission, AC power systems etc. Due to many advantages of frequency as an informative parameter of sensors, many manufactures produce different sensors and transducers with frequency, period, time interval or duty cycle output. This measurement transducer converts an input current, an input voltage or the signal from a sensor (thermocouple, Pt-100, resister, measurement bridge, etc.) into a frequency which is proportional to the magnitude of the input signal. Since the signal is in the form of a frequency, dynamic range is not limited by supply voltage and noise. Sensitivity of the device is maximized by a large detector area and precision input circuitry. Once converted to a frequency, the signal is virtually noise immune and may be transmitted over cables from remote sensors to other parts of system. Isolation is easily accomplished with optical couplers or transformers. The signals of several sensors may be easily multiplexed into one microcontroller or counter using digital logic [1, 2].

Data acquisition (DAQ) technology plays a fundamental role in a lot of virtual measurement solutions. The purpose of such a system is generally the analysis of the measured data and the improvement of the object of measurements. The data acquisition system can be divided in two main parts: hardware and software. The hardware part is made of sensors, cables, data acquisition module and computer. The software part is made of the instrumental drivers and the analysis software.

In present paper a design, development and implementation of frequency measuring system is presented. As measuring hardware a CompactDAQ system is selected. The hardware timebase for CompactDAQ system is located on the backplane of the chassis and is not specific to the DAQ modules themselves [5].

As application development environment for designing and implementing of virtual systems is chosen the LabVIEW [4]. This graphically-based programming language is optimized for test and measurement, automation, instrument control, data acquisition, and data analysis applications. Drivers and abstraction layers for many different types of instruments and buses are included or are available for inclusion in LabVIEW. The abstraction layers offer standard software interfaces to communicate with hardware devices. The provided driver interfaces save program development time and even people with limited coding experience can write programs and deploy test solutions in a reduced time frame when compared to more conventional or competing systems.

II. COMPACT DAQ MODULAR SYSTEM

Though fast, a downside to traditional measurement system is that it is a closed system, typically designed for a specific application. Therefore a new system or large modification is needed to implement a new type of measurement. New concepts in data acquisition technology incorporated into a modular production systems have quickly gained popularity primarily to the flexibility and cost-effectiveness that it brings to the final system.

National Instruments has launched a number of CompactDAQ chassis which can be used with more than fifty DAQ modules [5]. These chassis support wireless, USB and Ethernet buses, giving engineers and scientists the ability of a scalable measurement system for portable and distributed applications. There are more than fifty measurement-specific modules that feature multiple electrical and sensor connectivity options and can be combined with any chassis to create customized measurement systems specific to the needs of various applications. With such number of modules, the CompactDAQ platform eliminates the fixed functionality of traditional sensor measurement systems and gives users the ability to increase productivity while decreasing cost [5]. The metal enclosures make the chassis more resistant to environmental damage as compared to the traditional modular DAQ. An innovative signal streaming technology delivers high-bandwidth capabilities that make it possible to achieve sustained high-speed and bidirectional data streams over serial

[1]Georgi Nikolov is with the Faculty of Electronics and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gnikolov@tu-sofia.bg.

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

buses. In addition the chassis operate in a wide temperature range and can withstand shock and vibration, making them ideal for demanding measurement applications on the benchtop, on the production line or in the field.

All components of the CompactDAQ platform are supported with DAQmx drivers. With these software drivers, engineers and scientists can log data for simple experiments or develop a complete test system in various software environments. Consistent application programming interface means that an application developed for an wireless chassis will work with an or Ethernet chassis without any modifications to the software

The CompactDAQ (cDAQ) system consists of at least three parts as shown in Figure 1 – Input / Output Module (C Series I/O Module), the module interface, and the system timing controller (STC3). These components digitize signals, perform digital–to-analog conversions to generate analog output signals, measure and control digital input and output signals, and provide signal conditioning [5].

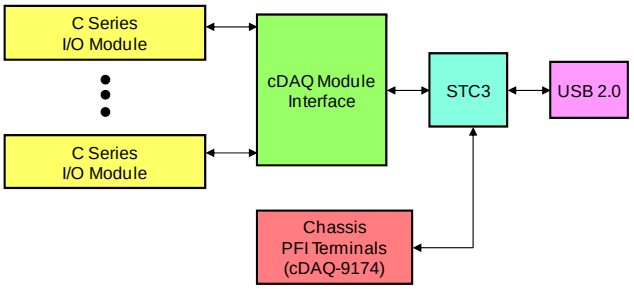


Fig. 1. CompactDAQ-9174/8 Block Diagram.

Input / Output Module (C Series I/O modules) provide built-in signal conditioning and screw terminal, spring terminal, BNC, D-SUB, or RJ-50 connectors. These modules are hot-swappable and automatically detected by the chassis. Bidirectional channels are accessible using the-DAQmx drivers. In most cases, the modules provide isolation from channel-to-earth ground and channel-to-channel.

The CompactDAQ module interface manages data transfers between the STC3 and the I/O modules. The interface also handles autodetection, signal routing, and synchronization.

All multifunction data acquisition hardware requires onboard timing circuitry to control analog, digital, and counter / timer lines. The evolution of timing application-specific integrated circuit technology occurred over decades. The National Instruments System Timing Controller (STC) an custom integrated circuit designed specifically for data acquisition applications. In comparison with the off-the-shelf counter / timer chips generally used on data acquisition devices, the STC stands alone. The third generation timing and synchronization technology STC3 delivers a new level of performance to multifunction data acquisition (DAQ) devices. This technology is the driver behind the advanced digital, timing, triggering, synchronization, counter / timer, and bus-mastering features. STC3 now equips all acquisition and generation tasks with inherent retriggerable capabilities with a single DAQmx property node

With STC3 technology, users can now accomplish more advanced analog, digital, and counter operations. In addition, applications that previously required additional onboard resources or were difficult to program can now execute independently and with less DAQmx code.

Other special functions include buffered pulse-train generation, timing for equivalent time sampling, relative timestamping, and instantaneous changing of sampling rate.

III. Methods for Frequency Measurement with Build-in cDAQ Chassis Counters

The CompactDAQ-9174 chassis has four general-purpose 32-bit counter / timers and one frequency generator. The counter / timers can be used for many measurement and pulse generation applications. All four counters on the CompactDAQ chassis are identical. In the Figure 2 the architecture of one of the counters is shown. As can be seen each counter has eight input signals, although in most applications only a few inputs are used [5].

Each counter has a FIFO memory that can be used for buffered acquisition and generation. There are also additional embedded counter. The embedded counters cannot be programmed independent of the main counter and signals from the embedded counters are not routable.

With described functionality of counters build in CompactDAQ Chassis four methods for frequency measurement can be used.

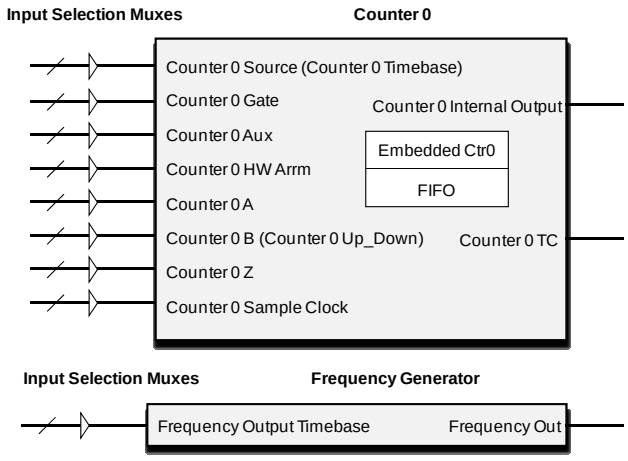


Fig. 2. CompactDAQ-9174 Chassis Counter 0 and Frequency Generator.

A. Period measurement with One Counter

The simplest method is low frequency measurement. This is accomplished by counting the rising or falling edges of a known source frequency f_{ref} between the two consecutive rising or falling edges of the unknown frequency f_X. By taking the frequency of the known source and dividing by the count, it is possible to calculate the period of the unknown signal. Typically one of the internal timebase, 20 MHz or 80 MHz, is used, but external clock source can also be used.

This type of measurement is suitable for low frequency measurement. [7]

B. Frequency Measurement with Two Counters

In this configuration, the counter will count number of unknown high frequency f_X during a period of known signal T_{ref}. The frequency can be calculated by multiplying the count by the frequency of the unknown signal [5].

In this method, a pulse of known duration must be routed to the Gate of a counter. This pulse can be generated by a second counter or can be generated externally and connect it to a PFI terminal. For second approach only one counter is needed.

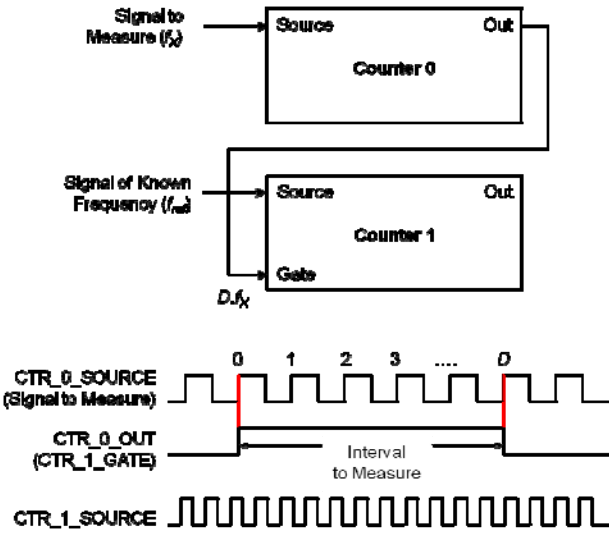


Fig. 3. Large Range of Frequencies with Two Counters.

C. Reciprocal Frequency Measurement with Two Counters

This method is used for measuring a large range of frequencies with two counters. The first counter is used to divide down the frequency of the signal to be measured, and then the second counter is used to measure the period of the divide down frequency. The actual frequency can be calculated by multiplying the resulting frequency measurement by the divide down value *D*. With this frequency method, the larger the divide down value *D*, the slower the resulting frequency, and more accurate the measurement result.

The signal to measure is routed to the Source input of Counter 0, as shown in Figure 3. NI-DAQmx automatically configures Counter 0 to generate a single pulse that is the width of *D* periods of the source input signal. Next, route the Counter 0 Internal Output signal to the Gate input of Counter 1 and configure it to perform a single pulse-width measurement.

D. Sample Clocked Buffered Frequency Measurement

As mentioned buffered counter functionality, using STC3 technology, has improved on its predecessors' capabilities in the areas of buffered period and frequency measurements. The user can now select sample clock as the timing type. When using a sample clock as the timing type, buffered frequency and period measurements are made by counting both an internal timebase (counted by embedded counter) as well as the unknown signal of interest up until the rising edge of the sample clock. However, the sample clock is a signal that must be specified and created by the user. The ideal frequency of the internal timebase is then divided by its count to find the effective frequency up to the next sample clock edge. With sample clocked frequency measurements, care must be taken to ensure that the frequency to measure is twice as fast as the sample clock to prevent a measurement overflow.

E. Choosing a Method for Measuring Frequency

Measurement errors are inherent in frequency measurements, but the effects of this error can be minimized by choosing a frequency measurement method that is most suited for given application [6]. The best method to measure frequency depends on several factors including the expected frequency of the signal to measure, the desired accuracy, how many counters are available, and how long the measurement can take.

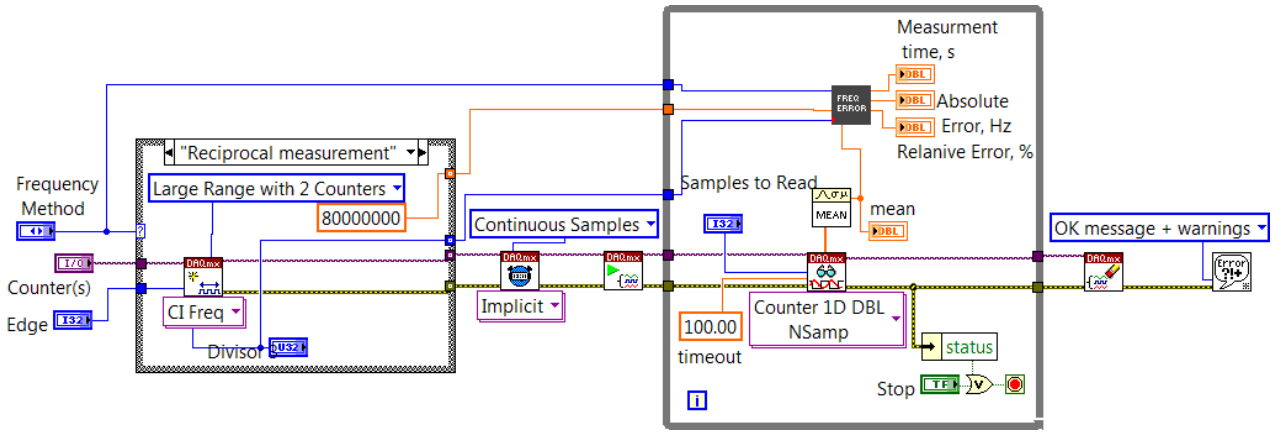


Fig. 4. Block Diagram of Virtual System for Frequency Measurements.

There are two sources of error in frequency measurements: timebase error (errors in the frequency of the crystal oscillator) and resolution error. Resolution may be limited by quantization error and trigger error. Timebase error is the maximum fractional frequency change in the timebase frequency due to all error sources (e.g., aging, temperature, line voltage).

The crystal oscillator used on the CompactDAQ Chasiss is rated at 50 parts per million (ppm) base clock accuracy.

The relative frequency measurement errors and measurement time are summarized in Table 1.

TABLE I
ACCURACY AND MEASUREMENT TIME OF FREQUENCY MEASUREMENT METHODS

Method	Parameter	
	Measure Time	**Relative Frequency Error**
Sample Clocked	$\frac{1}{f_S}$	$f_{ref}\left[\frac{f_X}{f_S}-1\right]+\frac{\Delta f_{ref}}{f_{ref}}$
Period Measurement	$\frac{1}{f_X}$	$\frac{f_X}{f_{ref}-f_X}+\frac{\Delta f_{ref}}{f_{ref}}$
Frequency Measurement	$\frac{1}{f_{ref}}$	$\frac{f_{ref}}{f_X}+\frac{\Delta f_{ref}}{f_{ref}}$
Reciprocal Method	$\frac{D}{f_X}$	$\frac{f_X}{D.\left(f_{ref}-f_X\right)}+\frac{\Delta f_{ref}}{f_{ref}}$

IV. EXPERIMENTAL RESULTS

With LabVIEW DAQ drivers and other measurement analysis tools it is easy to input real-world, time-domain signals directly from data acquisition hardware and provide results ready for charting, graphing, or further processing. Either traditional NI-DAQ drivers or newer one DAQmx can easy configure counters.

Because this is only a high-level discussion, this paper only touches on significant features of the developed source code. The part of this code or so called block diagram is shown in Figure 4. The "While loop" provides the ability to continuously execute until the conditional operator in low left corner is set to "false". "Case structure" on the left contains the DAQmx channel configuration for each of measurement methods.

The user interface or so called front panel of the created virtual system is shown in Figure 5. The front panel consists of three main sections. The first one is placed in left and is related to the configuration of channels. The next section is placed in central part of the front panel and consists of menu rings for choose of measurement method. There are and numeric digital indicator that indicated the measured value of current frequency. The last section consists of digital indicators that indicated the relative error, absolute error and measurement time according to selected method of frequency measurement. These values are calculated using equations in Table 1 with additionally developed software code.

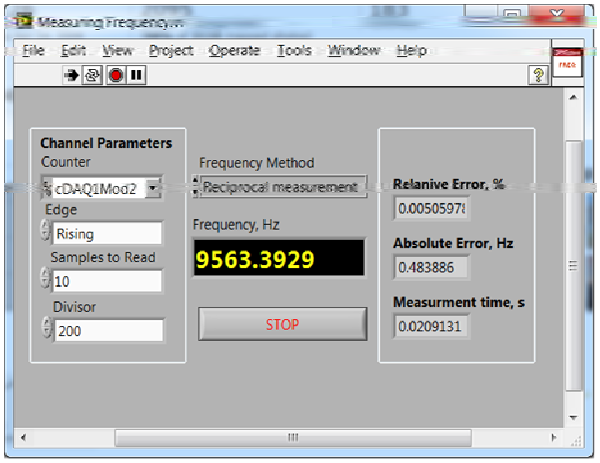


Fig. 5.Front Panel of Developed Virtual System.

V. CONCLUSION

In present paper a design and development of virtual system for frequency measurement is considered. This virtual system consists of CompactDAQ Chassis, Digital I/O C series DAQ, and graphical programming environment. The illustrated approach offers a simple solution for an application that needs the low cost frequency measurement of any sensor with frequency output and DAQs with counter / timers. With software error estimation the system can be used for both designs in the same time: for maximum resolution and accuracy as well as for maximum data acquisition rate.

ACKNOWLEDGEMENT

This investigation has been carried out in the framework of the research project № DTK 02/50/17.12.2009.

REFERENCES

[1] J. Fraden, *Handbook of modern sensors: physics, designs, and applications*, ISBN 0-387-00750-4, Springer-Verlag New York, 2004.

[2] E. Haile, J. Lepkowski, "Oscillator Circuits For RTD Temperature Sensors", Microchip Technology Inc., AN895, 2004, http://www.e-sonic.com/whatsnew/Microchip/signal/AN895.pdf

[3] John G. Webster, *Electrical measurements, signal processing, and displays*, ISBN 0-8493-1733-9, CRC Press LLC, 2004.

[4] P. Blume. *The LabVIEW Style Book*, ISBN 0-13-145835-3, Pearson Education, 2007.

[5] National Instruments, "NI CompactDAQ 9171/9174/9178 USB Chassis", User Manual, July 2011.

[6] Agilent Technologies, "Understanding Frequency Counter Specifications", Application Note 200-4, May 1997.

[7] National Instruments, "Making Accurate Frequency Measurements", Tutorial, Oct 5, 2007.

Wind Energy and Steps Towards 100 Percent of Renewable Energy Penetration

Aleksandar Lj. Malecic[1]

Abstract – **In this paper the idea of 100 percent of renewable energy penetration sometime in the future with wind energy as its important part is analysed. As a presumably more sustainable approach to energy supply than for instance fossil fuels, this concept needs to be capable to last for at least as long as the time interval from the first industrial revolution to this moment. This means the capacity to sustain future climate, economic, political, environmental and technological challenges.**

Keywords – **Wind energy, Generation mix, New economy, Grid, Research & development.**

I. INTRODUCTION

Wind energy is at the moment the fastest developing variable renewable energy technology. As such, it will have a significant role in the future transition towards the 100 percent renewable energy penetration if and when such a transition happens. During that time it will face many challenges and barriers. First of all, it is very challenging to develop scenarios for something so enormously complex spread decades in the future (the 21st century with all of its socioeconomic and technological changes and beyond). Such a radical change of an existing energy infrastructure requires a modified view of economics and coordination of many activities related to grid connection (as well as grid stability requirements and energy mix) in a way that they really provide a more sustainable system which is robust enough to last for centuries and allows improvements in technology and energy mix.

II. SCENARIOS

No matter when the goal of 100 percent of renewable energy penetration is achieved, any initiative towards a further development and implementation of renewable energy should be a step along the path that has 100 percent penetration as the goal. This goal is very abstract because:

1. The exact amount of needed energy production is uncertain. – Renewable energy needs to replace sooner or later fossil fuels and allow the system built upon relatively cheap fossil fuels working without too many disturbances. Also, when energy consuming machines and vehicles switch to electrical energy, there is a sudden increase of demand for it. National economies rising and declining can cause additional complications.
2. There is a disagreement about when exactly should the complete transition to renewable energy take place. – It will hopefully happen sometime in this century, but an uncertain timeframe of several decades is enormous in globalised business as usual.

The world we live in is the most complicated ever. The environment is being changed fast, human population is the biggest ever and there is no will at the moment to reverse the pressure towards a global growth of population and economy (both still perceived as patriotic duty among nations worldwide). In such a world renewable energy is supposed to replace fossil fuels. This is not just a technological or societal/economic challenge, but it also requires a totally new mindset: long-term thinking (towards the transition and after it) instead of short-term, abstract and hardly traceable goals such as leaving a better world and the environment in a better condition to the offspring instead of investments and profit and collective instead of self interests.

Some obstacles on the path towards renewable energy transition with wind playing an important role are:

1. There is a need for a lot of coordination across levels: transmission and distribution networks (a significantly different approach between fossil fuels being dug out of the ground and transported by trucks, ships and pipes and renewable energy on the surface demanding new transmission lines) and wind farm clusters, administration and communities of practice.
2. Already established societal structures and working places need to be modified in order to change an already established and deeply interconnected (primarily, but not only around financial interests of multinational corporations making profit on fossil fuels) system.
3. As a variable energy source becomes more present percentage-wise, there are changes in capacity credit (capability to replace energy produced by other conventional energy source) and a need for additional grid connection requirements [1].

[1]Aleksandar Lj. Malecic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: aleks.malecic@yahoo.com

Even with comparable prices between non-renewable and renewable energy (fossil fuels will be expensive and later even nonexistent), there is a friction between two systems, the one already existent with its administration and defined silos of experts, working places and responsibilities and the one that shall hopefully develop (e.g. looking for energy under vs. above the ground and highly depending on weather, short-term vs. long-term planning, oil magnates vs. communities, mining vs. recycling, transparency and competition vs. monopolies etc.). Besides new transmission lines, there is a similar need for redefined working places for people and transmission and distribution systems collectively providing energy in a very different way.

While different scenarios for renewable energy transition are being analysed, they differ in generation mix and future demand (energy efficiency, economy). There are different future projections for achievable reaching the 100 percent renewable energy supply on a national level in published papers and presented by respected authors and magazines, such as for example by 2030 only by wind, sun and water [2] or for example by 2050 with biomass as a part of the energy mix [3]. Biomass [3][4] does have a closed cycle of organic matter and CO2 emissions and can theoretically be treated as a renewable energy source, but it requires organic waste (fuel, even if it is non-fossil). Whatever the case, as this transition takes longer, there will be more risk from sudden changes in supply and demand of technologies and energy worldwide [5][6].

III. NEW ECONOMY

New economy [7] is a concept that appeared in economic theory at the end of the 20th century. It was at the first place about the internet and dot-com bubble. Even if it seems that it failed at the beginning of this century during a recession, this concept still makes sense. Future challenges such as transition towards renewable energy and sustainability especially around the world will require proper use of information and networks (getting things done instead of profit making by and for a limited number of people and companies) perhaps even more for collective focus to the problems (tasks and solutions instead of profit as the main goal) than simulations and analyses.

There are differences between wind farms and conventional power stations. Wind farms are modular, meaning they contain more individual wind turbines and technical problems on a single turbine do not necessarily reflect to the whole farm. Also, costs are distributed differently during a life cycle and they are concentrated on building of a turbine, connection to the grid and later maintenance. While comparing costs of wind and other energy technologies, one must bear in mind that during the expected life time (approximately 20 years for a wind turbine [6]) wind turbines are very little influenced by fluctuating prices of fossil fuels and risks of reliance on import from a distant destination and country.

Since the role of renewable energy is to replace already existing energy sources, we need to understand that wind power stations need to respond to future challenges and to sustain under future circumstances different from what we have at the moment, meaning that infrastructure, technologies used (including grid codes and hardware and software used for management and control) should provide solutions instead of more problems. Potential problems are climate changes and financial and resources crises [5]. Renewable energy is supposed to become an integral part of a future upgraded energy system. A significant number of people around the world don't use electricity and many of them have even never made a phone call. Complexity is, according to historical records, an exception rather than a rule. Former complex societies have faced collapses when they were incapable to deal with the law of diminishing marginal returns. Once a society becomes so complex that its status quo is unbearable (expensive infrastructure and administration, high rate of consumption of energy and materials), a collapse can be the only way forward. This is exactly a danger with developed countries at the moment. Energy consumption tends to go up and these countries demand enormous amounts of energy and natural resources just to protect the established status quo and these trends cannot be spontaneously reversed.

Let's take for example two supergrids, one for wind farms and the other one for solar energy. They are expected to transmit energy from supplying areas to demanding areas for decades. During that time, there can emerge improvements for either technology (wind and solar) and requirements for expansion of one grid (connected to the location with the superior technology) and demand can change in the region and in the world as well as penetration of variable energy (grid code requirements and congestion management). Also, there are possible changes in energy mix. There were warnings about the limits of growth [5] on a limited planet decades ago and renewable energy still has problems to reach the mainstream. Radically changed grids (the transition to renewable energy and supposedly more sustainable configuration) will face problems if they change one more time in the future. Even if they stay similar,

there will be a need for strong national (transnational for transnational grids) economies and companies (and global competition, investments abroad and off-shoring) capable to maintain big, complex and expensive grids (possible problems with diminishing marginal returns). An established grid connecting demand locations with renewable energy should be much higher on the list of priorities for investments just to protect a new status quo than it was in the past (our present).

Building and maintenance of a new energy infrastructure requires a lot of money. In order to provide energy in the future, developing countries [8] need innovative methods such as distributed generation, tighter communities, ICT, exchange of information, locally manufactured parts, and others. These very methods should also be considered in developed and at the moment rich countries. A sustainable technology also needs to be economically sustainable.

IV. GRID CONNECTION

Wind always blows somewhere. Wind farm clusters [9] spread across larger areas have a lower correlation (the smoothing effect) of power produced by individual wind turbines and resemble more to a classic more predictable and controllable power station. This effect needs to be taken into account while new transmission lines connecting present and future wind farms are planned. For example, Spain is one of global leaders in wind energy and the future of this technology depends on connections with France [10].

In order to properly plan supply, wind power production needs to be predicted as precisely as possible [11]. These forecasts are needed for congestion management (short-term planning) and planning of the energy mix (long-term planning).

Grid codes [1][12] define special requirements for variable energy sources such as frequency control, power control (active and passive) and fault ride through. They vary between countries. New wind turbines need to be designed and the old ones need to be upgraded according to these rules.

Energy mix [13] is an important part of the overall stability. For instance, some technologies, including those still not widely used such as combined heat and power (CHP) and electric vehicles can by their specific behaviour in demand, conversion, storage, and supply add more stability to the system (if they are properly combined). Different types of regulation and optimization [14][15] can help to make a choice between possible combinations by assessed costs, dispatch and capacity expansion.

V. AVAILABLE TECHNOLOGY AND ITS LIFE CYCLE

A good model resembles reality and there are aspects of changes of reality that established models and perceptions can't properly describe. Such a case is feasibility studies and choices between options when people in charge are comparing already existing energy technologies and renewable energy such as wind. According to experiences from Denmark [13], one of the world leaders in wind energy exploitation, radical changes are possible exclusively under pressure outside institutions. Radical changes are often out of reach of perception of institutions deeply rooted in one way of thinking and working. Wind technology needs to expect even more friction in countries with less transparent political activities than Denmark mentioned here. The establishment will try to create a situation that looks like there aren't possible alternatives and that a new (for instance) coal-fired power plant is the only existing choice and as such the best one. Neoclassical economics [16] sees the world as a set of rational agents capable to choose between different options and cost benefits as the most precise assessment of value. Still, fossil fuels are limited resources and their price will very likely start irreversibly growing in not so distant future (peak oil sooner than other peaks), while at the same time wind turbines will grow larger and be attached to improved technology connecting them to the grid (frequency, voltage, and power stabilization) and become cheaper even without feed-in tariffs and other regulatory measures financially supporting renewable energy and lower CO2 emissions.

Renewable energy will go through three phases [13] (some countries have already passed the first phase):

1. Introduction – Wind turbines don't influence much grid stability, but there is a lot of resistance and misunderstanding from business as usual and official politics.
2. Large-scale integration – Grid code requirements are being developed for variable renewable energy and new turbines need to obey them and the old ones need to be adapted. Analyses of what to do in this stage need to bear in mind energy efficiency and savings and investments and planning of transmission and distribution lines as well as renewable energy mix needed to properly approach the next stage. This stage in its mature form needs to be a strategic step (instead of an obstacle akin to non-renewable energy technologies) towards countries and the world relying solely on renewable energy.

3. 100 percent renewable energy systems – The entire energy system relies on renewable energy. The energy mix needs to be close to the optimum and renewable energy technologies are competing between each other.

In every phase there is a need to allow the best results from research and development [6][17].

The book Limits to Growth (analysed in [5]) warned about the future complications forty years ago. It will be more difficult to cope with them now than back then. In order to do so, global leaders need to figure out how to simultaneously [5]:

1. Rapidly reduce dependence on fossil fuels.
2. Adapt to the end of economic growth.
3. Design and provide a sustainable way of life for 7 billion people.
4. Deal with the environmental consequences of the past 100 years of fossil fuel growth.

These tasks are directly or indirectly attached to renewable energy, especially its portion based on wind (the faster growing at the moment) and sun. The first one is obvious, because it is about replacement of one limited and harmful energy source with another. The second one is about approaching to business, differentiation between socioeconomic and business economic feasibility studies (different technologies working on the common goal of energy supply), and challenges to operate and maintain the new energy infrastructure even when the global economy goes down. The third and fourth task are not so obviously linked to wind energy, but if the first two tasks are properly implemented, they will inevitably raise choice awareness [13] (especially if renewable energy gets closer to 100 percent penetration) across society necessary for building a more sustainable world.

VI. Conclusion

Wind energy will, if and where it becomes a larger part of a national energy mix, on its way as a leader towards the 100 percent renewable energy future, necessarily cause many changes in the socioeconomic and infrastructural realm. As much as it needs to deal with the resistance from established institutions and ways of providing energy, it mustn't stay too much (some friction and delay between research and implementation is inevitable) on the way to provide the optimal energy mix and technology. Also, a supposedly sustainable technology and approach to business needs to be more capable to survive environmental, political, economic, and resources crises.

References

[1] M. Comech, M. García-Gracia, S. Arroyo and M. Guillén, "Wind Farms and Grid Codes", Spain, CIRCE-University of Zaragoza, 2010.

[2] M. Jacobson, M. Delucchi, "A Plan for a Sustainable Future: How to get all energy from wind, water, and solar power by 2030", Scientific American, pp. 58-65, November 2009.

[3] H. Lund, B. Mathiesen, "Energy System Analysis of 100 Per cent Renewable Energy Systems - The Case of Denmark year 2030 and 2050", Aalborg, Denmark, Aalborg University, 2007.

[4] D. Connolly, M. Leahy, H. Lund and B. Mathiesen, "Ireland's pathway towards a 100% renewable energy-system: The first step", Ireland, 2009.

[5] R. Heinberg, "The Post Carbon Reader Series: Foundation Concepts Beyond the Limits to Growth", Santa Rosa, USA, Post Carbon Institute, 2010.

[6] "Wind Energy – The Facts", Intelligent Energy - Europe programme of the Executive Agency for Competitiveness and Innovation, 2009.

[7] M. Boudourides, "Indicators & the New Economy", EICSTES Meeting, Vienna, Austria, 2002.

[8] A. Vaccaro et al, "Reliable Electric Power for Developing Countries", Humanitarian Technology Challenge, 2008.

[9] T. Ackermann, "Wind Power in Power Systems", Chichester, UK, John Wiley & Sons, 2005.

[10] P. Sonvilla, R. Piria, E. Zane, J. Bracker and D. Bauknecht, "Integration of electricity from renewables to the electricity grid and to the electricity market – RES-INTEGRATION, National Report: Spain", Berlin, Germany, Öko-Institut, 2011.

[11] B. Ernst, B. Oakleaf, M. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken and K. Rohrig, "Predicting the Wind", IEEE Power & Energy Magazine, 2007.

[12] M. Tsili, Ch. Patsiouras and S. Papathanassiou, "Grid Code Requirements for Large Wind Farms: A Review of Technical Regulations and Available Wind Turbine Technologies", Athens, Greece, National Technical University of Athens (NTUA), 2008.

[13] H. Lund, "Renewable Energy Systems: The Choice and Modeling of 100% Renewable Solutions", USA, UK, Academic Press, 2010.

[14] "LEAP: Long-range Energy Alternatives Planning System, User Guide for Version 2011", Somerville, USA, Stockholm Environment Institute, 2011.

[15] H. Lund, "EnergyPLAN: Advanced Energy Systems Analysis Computer Model - Documentation Version 9.0", Aalborg, Denmark, Aalborg University, 2011.

[16] C. Arnsperger, Y. Varoufakis, "What is Neoclassical Economics?: The three axioms responsible for its theoretical oeuvre, practical irrelevance and, thus, discursive power", Louvain, Belgium, University of Louvain, 2005.

[17] A. Vergnol, J. Sprooten, B. Robyns1, V. Rious and J. Deuse, "Real time grid congestion management in presence of high penetration of wind energy", 13 th European Conference on power electronics and applications - EPE 2009, Barcelona, Spain, 2009.

Multitool Online Assisted Design of Communication Circuits and Systems

Galia Marinova[1]

Abstract – **The paper deals with a Multitool environment for computer-aided design assisted by a portal for online calculators. The structure of the portal is presented as well as design examples (projects of loudspeakers, LED circuits and PCBs, etc.) illustrating join application of Computer-aided design (CAD) tools and online calculators.**

Keywords – **CAD tools, Online calculators, Communication circuits and systems.**

I. INTRODUCTION

Recently there is a boom of online calculators for different engineering tasks. The paper describes an extension of the Multitool environment for computer-aided design, described in [1], with a portal for online calculators, applicable in communication circuits and systems.

The Multitool environment from [1] integrates CAD tools as: MATLAB, ORCAD/PSpice/Layout, WARP 6.2, ISE Webpack, Quartus II, Microwind, FilterCAD, PAC Designer, MMICAD, etc., which cover different aspects of realistic projects in communications (system level, analog, digital, analog-digital, programmable, microwave, microelectronic design and PCB design). The design process can be assisted by specialized online calculators which add more features to the Multitool environment. The structure of the portal for online calculators is presented further. The approach of joint application of online tools and downloaded tools for solving communication circuit design tasks is illustrated with four examples in the area of crossover design, PCB design, SMPS design and LED array and driver design.

II. PORTAL FOR ONLINE CALCULATORS FOR COMMUNICATION CIRCUITS AND SYSTEM DESIGN

The portal is in experimental phase of development and it will be connected to E-content for PSpice and other CAD tools for communications from [2]. It organizes different groups of online calculators:

- Online design tool PowerEsim for switch-mode power supply design [3];
- Online calculators for resistors and capacitors [4, 5, etc.];
- Online calculators for PCBs and transmission lines [6, etc.];
- Online matching network tools – Smith Chart diagram tools [7, etc.];
- Online calculators and wizards for LED array design [8, etc.];
- Online analog/digital/crystal filters design tools [9, etc.];
- Online 555 timer calculators [10, etc.];
- Online design tools for loudspeakers and microphones (mono and stereo) [11, etc.];
- Online antenna calculators and design tools [12, etc.];
- Online SCA based open source software defined radio – OSSIE [13, etc.].

The online calculator groups are connected to a set of online calculators. Each calculator is described shortly and a link is provided to its original web address (see Fig.1).Instructions are given for joint online and downloaded tools tasks solutions for communication circuits.

III. EXAMPLES OF ONLINE CALCULATOR ASSISTED DESIGN OF COMMUNICATION CIRCUITS IN THE MULTITOOL ENVIRONMENT.

Four examples illustrating the joint approach are presented in details.

A. Example 1. Crossover design through joint application of loudspeakers' online calculator and PSpice

Figure 2 illustrates Two way crossover design (with 6th order Linkwitz-Riley LC filters) using the online calculator [11], modeling the loudspeakers' drivers with Thile small equivalent circuit as indicated in [14] and simulating the whole design in PSpice for specification verification. The output voltages of the tweeter and the woofer are plotted separately and then the resulting sum is obtained. All three voltages are plotted in dB.

B. Example 2. PCB design through joint application of ORCAD/Layout, online calculator for transmission line impedance calculation and PSpice

Figure 3 illustrates the characteristic impedance calculation with the online tool [6] of a PCB trace, taking its geometric parameters from ORCAD/LAYOUT tool. Then the PCB trace is modeled as a transmission line and simulated in PSpice as proposed in [15]. The rise time of the input pulse is 5ns. The parasitic picks in the output voltage are attenuated with a serial resistor R4 and then a new PSpice simulation illustrates the elimination of the parasitic picks.

[1]Galia Marinova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gim@tu-sofia.bg.

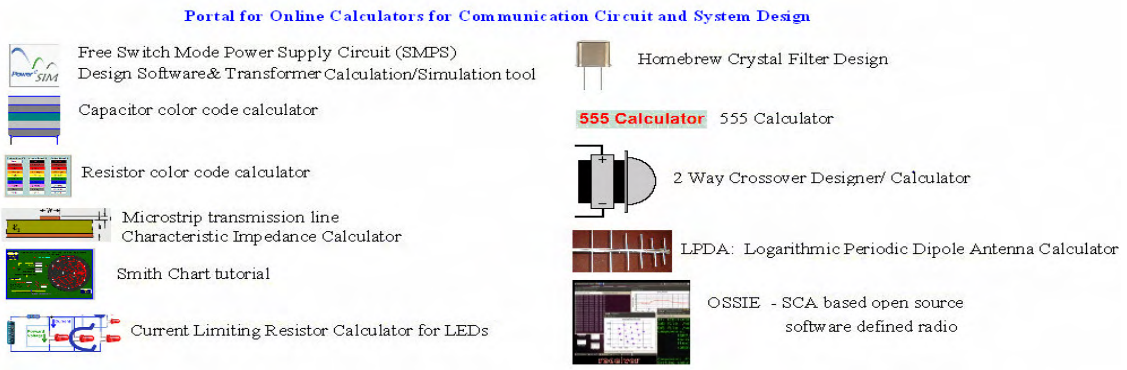


Fig.1. Portal for online calculators for communication circuit and system design

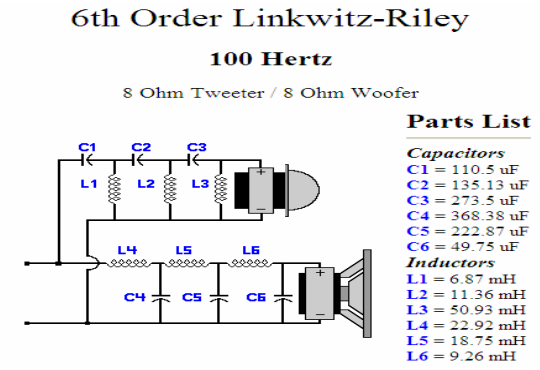


Fig. 2a. Crossover circuit generated in Loadspeakers' online calculator

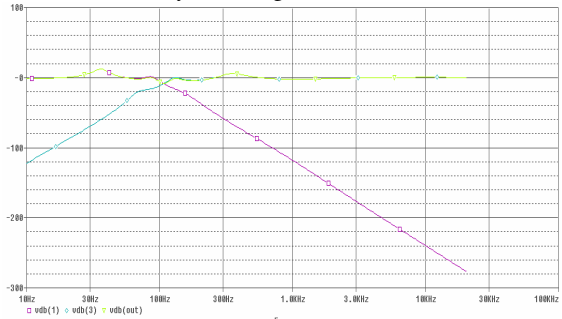


Fig. 2c. PSpice simulation of the two-way crossover circuit: Tweeter output voltage in dB; Woofer output voltage in dB; Output voltage in dB for the two-way crossover circuit

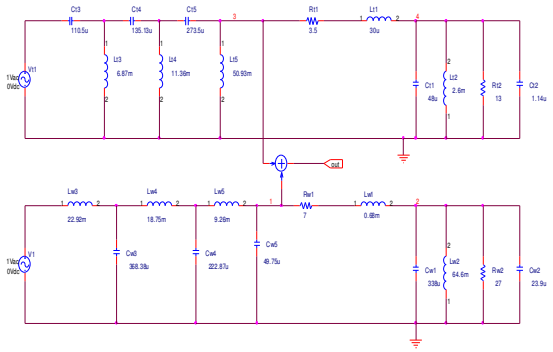


Fig. 2b. PSpice model of the circuit from Fig.2a

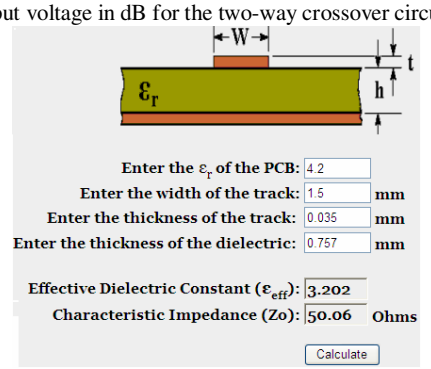


Figure 2. Two way crossover design

Fig. 3a. Characteristic impedance calculation of a PCB trace modeled as transmission line

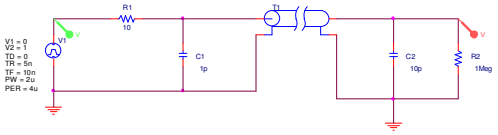


Fig. 3b. PSpice model with transmission line and pulse voltage source with 5ns raise time

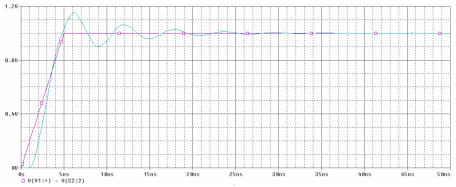


Fig. 3c. PSpice simulation of the circuit from Fig.3b.: input pulse and output voltage with parasitic picks

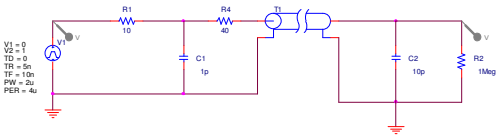


Fig. 3d. Circuit from Fig.3b with R4 for pick attenuation

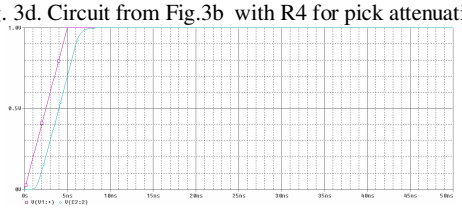


Fig. 3e. PSpice simulation of the circuit from Fig.3d.: Input pulse and Output voltage without parasitic picks

Fig. 3. Transmission line effect of a PCB trace for high speed

C. Example 3. Switch mode power supply design through joint application of PowerEsim and PSpice

Figure 4 illustrates the generation of LLC (Series Parallel) converter block in PowerEsim [3] and the possibility for simulation in PSpice of a circuit from this block for verification. A circuit from the LLC PWM (Pulse width modulation) block is simulated in frequency area.

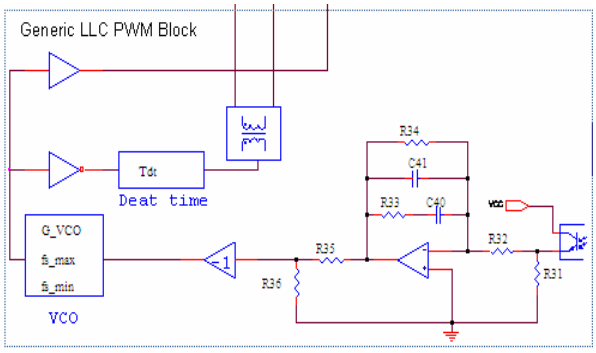


Fig. 4a.LLC PWM block synthesized in PowerEsim

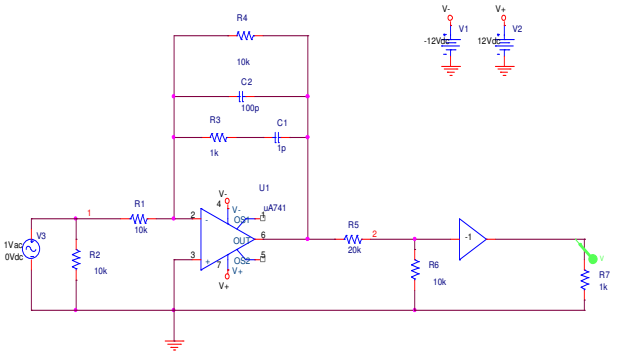


Fig. 4b. Electrical circuit from the LLC PWM block in ORCAD/Capture

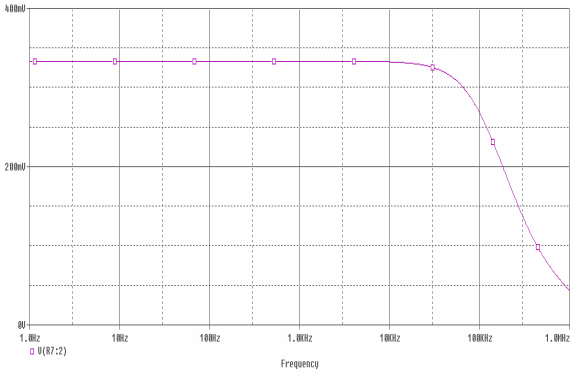


Figure 4c. Simulation in PSpice of the circuit from Fig.4b. in frequency area

Fig. 4. PWM block of LLC circuit designed in PowerEsim and PSpice simulation of a circuit in the block

D. Example 4. LED array and driver circuit design through joint application of online calculator and WARP 6.2

Figure 5 illustrates the LED array design in the online calculator [8] and the design of a Pseudo random number generator (PRNG) in VHDL and its simulation in the tool WARP 6.2 (CYPRESS). The PRNG is designed using shift register and XOR circuits.

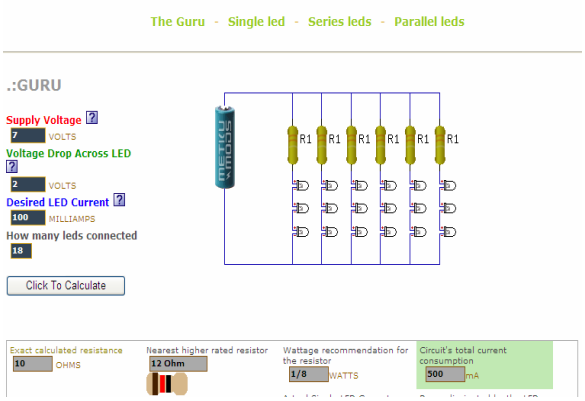


Figure 5a. LED array design in the LED online calculator

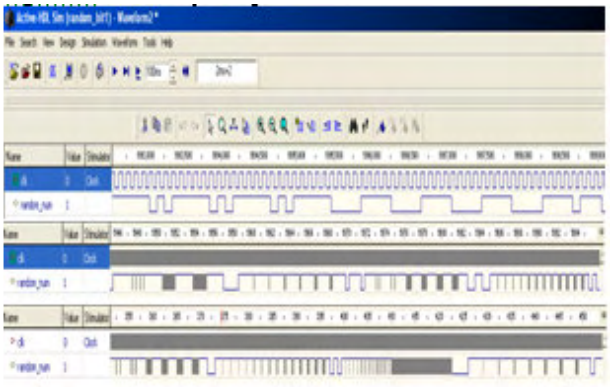


Figure 5b. Simulation waveforms of a PRNG designed in VHDL with the tool WARP 6.2(CYPRESS).

Figure 5. LED array design in online calculator and design of a driving circuit in WARP 6.2.

IV. Conclusion

The paper is an attempt to make useful the large number of online calculators available nowadays to facilitate the design of communication circuits and systems adding them to a Multitool design environment through a new portal in the web.

Four different types of online calculators for loudspeakers' design, for PCB trace characteristic impedance calculators, SMPS tool design and LED array design tool are combined with ORCAD/Capture/PSpice, Layout and WARP 6.2. tools.

The PSpice simulator is applied mainly as a verification tool. ORCAD/Capture tool is applied for equivalent circuit's edition. ORCAD/Layout is applied as a source of data in the PCB trace Characteristic impedance calculator and WARP 6.2 is applied for VHDL description and simulation of a digital driver circuit.

The four examples, illustrating the joint approach, confirm the advantages of the approach and promise plenty of applications in students' education contributing to its accessibility and attractiveness. The portal and the applications proposed can be applied in the course of Computer-aided design for the Bachelor and/or Master degree students and they can serve for a diversity of individual projects.

REFERENCES

[1] G. Marinova, Multidisciplinary approach for teaching Computer-Aided Design incommunications using multiple software tools, *Advanced Aspects of TheoreticalElectrical Engineering Sozopol '2010,* 19.09.10 – 22.09.10, Sozopol, Bulgaria, Plenarylectures, pp. 34 -42

[2] E-content for PSpice and other CAD tools for communications http://www.pueron.org/pueron/E_CADSystems/E_CADSystems.htm

[3] Free Switch Mode Power Supply Circuit (SMPS) Design Software& Transformer Calculation/Simulation tool http://www.poweresim.com/

[4] Capacitor color code calculator http://www.csgnetwork.com/capcccalc.html

[5] Resistor color code calculator http://www.csgnetwork.com/resistcolcalc.html

[6] Microstrip transmission line Characteristic Impedance Calculator http://chemandy.com/calculators/microstrip_transmission_line_calculator_Hartley27.htm

[7] Smith Chart tutorial http://www.fourier-series.com/rf-concepts/smithchart.html

[8] Current Limiting Resistor Calculator for LEDs http://ledcalc.com/

[9] Homebrew Crystal Filter Design http://www.changpuak.ch/electronics/calc_17.php

[10]555 Calculator http://freespace.virgin.net/matt.waite/resource/handy/pinouts/555/index.htm

[11] 2 Way Crossover Designer/ Calculator http://www.diyaudioandvideo.com/Calculator/XOver/

[12] LPDA: Logarithmic Periodic Dipole Antenna Calculator http://www.changpuak.ch/electronics/lpda.php

[13] OSSIE - SCA based open source software defined radio http://ossie.wireless.vt.edu/

[14] How to simulate speaker/crossover using ORCAD http://www.shine7.com/audio/orcad.htm

[15] Johnson H., High Speed Digital Design, Prentice Hall, 1993

E-learning Systems as a Behavioural Analyst

Valentin Videkov[1], Rossen Radonov[2]

Abstract – **The report discusses the application of e-learning systems as a tool for indirect survey. Some possibilities for summarizing data from the application of electronic learning as a tool for assessing the behaviour of students are presented. Some practical results of such studies are also shown.**

Keywords – **E-learning, statistics, behavioural analysis**

I. INTRODUCTION

Over the past two decades, e-learning systems have gone through various stages - from implementing electronic tests [1] Internet-based learning [2] to remote [3] and mobile [4]. There are many platforms for creating e-learning courses [5] [6] which widely use multimedia tools. The application of state-of-the-art information technology allows the learning process to be statistically traceable. Virtually any system monitors and processes certain information. Most often it is directly linked to the learning process and presents the results of the evaluation [7]. A more detailed analysis of this information can show other sides of the implied process characterizing the behaviour of its participants. These include efficiency in the completion of tasks, teamwork and others.

II. DATA IN E-LEARNING

A. The training objective

We can consider the question of the purpose of learning from different positions. These could be educational, economic, moral, and others. Achieving these goals is related to the assigning of tasks to be performed by trainers and trainees through appropriate means. On the other hand the achievement of the goals, i.e. tasks is associated with the collection and evaluation of various data. The evaluation of these data is the criterion for achieving the objective.

B. Data

Data that are associated with e-learning can be divided into the following types: data representing a database object (such as text documents, multimedia) data to evaluate the outcome of training (two types - tests and assessments), data from the process (attendance, number trained, schedules), data on participants (personal information for teachers, students).

Another division can be introduced to a state of data in time domain. Part of the data is slightly changing or constant (data of the participants in the process, quantity), others are dynamic (evaluations, test results, statistics of visits).

The typical for e-learning systems is that data are directly applicable for processing by various electronic means.

C. The statistics

Virtually all modern e-learning platforms contain a set of procedures for statistical processing. These are classic processing for average scores, distribution depending on the score, number of solved problems, distribution of wrong decisions and similar. Processing which brings interactive learning, i.e. the platform allows evaluation of the achieved level and manages training procedures is relatively rare [8].

III. STATISTICS OF BEHAVIOUR

A. The task

Solving the problem of determining the behaviour of participants in the educational process assessed using electronic systems is of interest. In this case, different demands can be placed on this assessment and the assessment of the behaviour itself. It can be viewed from the perspective of psychologists, educators, managers, users and other staff. In this paper an attempt is made to show some results from the use of e-learning systems for such behavioural assessment.

B. Database

For the purposes of this study various electronic modules and databases for them were used. They had been developed at the Technical University of Sofia. These are the university student information system (USIS), E-management, electronic survey, electronic report of student's assessment, electronic report of teacher's workload.

USIS contains data for students, faculty, courses, assessments of students and number of times, good luck. The E-management platform is used for electronic control of the learning process and is linked to USIS. It contains data on discipline (teaching materials), academic schedule, results from the conduct of classes and tests, time parameters for data entry and more. The electronic surveys collect data for evaluating the discipline by students and time related parameters of the vote. Other modules contain similar data.

[1]Valentin Videkov is with the Faculty of Electronic Technologies at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: videkov@ecad.tu-sofia.bg

[2]Rossen Radonov is with the Faculty of Electronic Technologies at the Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: radonov@ecad.tu-sofia.bg

C. A hypothesis

A number of training data for assessment of the learning process is collected in the electronic systems . However, these may be associated with additional time indicators allowing indirect evaluation of the behavior of learners, and learners well as.

For example, tracing of time spent for test training or number of trainings with different tests, the average rating of training and final assessment may be an indicator for memorizing the material. Even more interesting relationship may be the beginning of the task assignment and its completion. This can be an assessment of statistical predictions of the time for completing the work in a company. In this spirit a link between the performance of duties in curriculum and a future technological discipline could be sought.

D. Experimental results

Fig. 1 shows the relative distribution of answers (Q) in time at answers to the questionnaire and submission of reports (R).

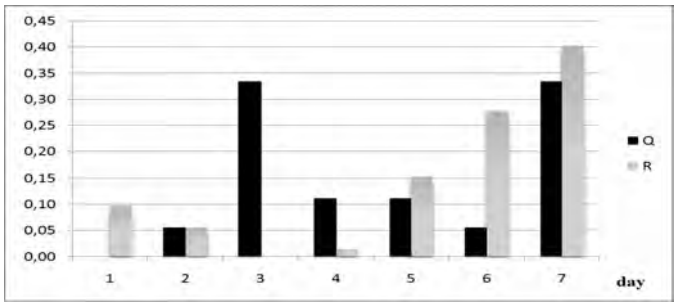


Fig. 1.

It can be seen that the statistical distributions are similar, indicating a behavioural influence.

Fig. 2 shows the distribution of assessments in two subgroups divided on the basis of their own choice.

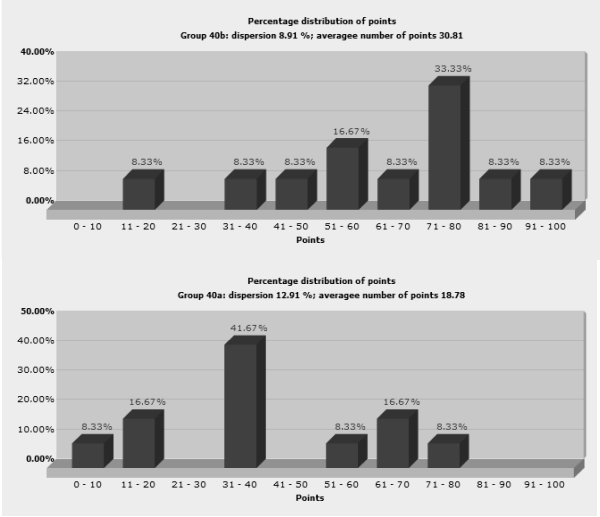


Fig. 2.

The difference in scores that could be interpreted as self-grouping of strong students in a separate subgroup can be seen.

The results from tracing the tendency for self-study or use of external results are interesting. Fig. 3 shows the data for review of other reports before the preparation of their own.

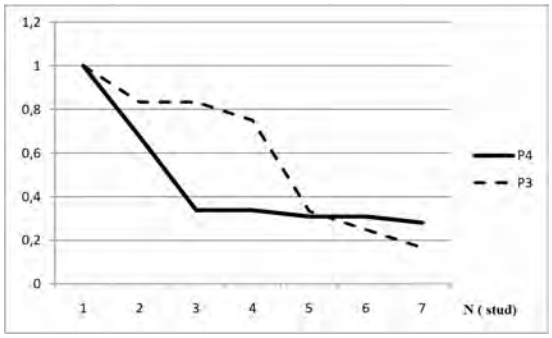


Fig. 3. Relative distribution of foreign protocols reviewed 4th year - P4 and 3 rd - P3.

The sharp decline in the rate chart related to 4th year, we can interpret as a preference for independent work. 43% of the female students have reviewed other protocols and 73% of the male students.

IV. Conclusion

Preliminary results in this work show that using a suitable algorithm, the data accruing from e-learning systems can be used for behavioural analysis. They may show the tendency of the student group to prepare team solutions, to deliver results within a specified time, a tendency to individualism and other.

References

[1] William Horton, Katherine Horton, *E-Learning Tools and Technologies: A Consumer's Guide for Trainers, Teachers, Educators, and Instructional Designers*, Wiley Publishing,2003.

[2] Deanie French at all, *Internet Based Learning: An Introduction and Framework for Higher Education and business*, Stylus publishing, 1999

[3] Mohamed Khalifa, "Remote learning technologies: effectiveness of hypertext and GSS", Decision Support Systems, Volume 26, Issue 3, September 1999, pp. 195–207

[4] Mike Sharples at all, "Towards a Theory of Mobile Learning", http://www.iamlearn.org/public/mlearn2005/www.mlearn.org.za/CD/papers/Sharples-%20Theory%20of%20Mobile.pdf

[5] Moodle, http://moodle.org/, [Accessed 7 April 2012]

[6] ILIAS E-Learning, https://www.ilias.de/docu/ , [Accessed 7 April 2012]

[7] Bloom, Benjamin S. at all, *Handbook on Formative and Summative Evaluation of Student Learning*, McGraw-Hill Book Company, 1971

[8] В.А. Разыграева, А.В. Лямин, „*ПОСТРОЕНИЕ МОДЕЛИ АДАПТИВНОГО ЭЛЕКТРОННОГО ОБУЧЕНИЯ, УЧИТЫВАЮЩЕЙ ФУНКЦИОНАЛЬНОЕ СОСТОЯНИЕ ОБУЧАЮЩЕГО*", http://rois.loiro.ru/index.php?module=articles&action=view&cid=0&id=153, [Accessed 12 February 2012]

Application of Remote Instrumentation in Learning using LabView

Ivo Dochev [1] and Liliana Docheva[2]

Abstract – **In this paper is suggested block diagram for remote control and measurement that use computer networks. Also it is discussed how the configuration of the LabView Web Server can be set, in order to realize remote control and measurement.**

Keywords – **Instrumentation, Remout control and measurement, LabView.**

I. INTRODUCTION

In the practice very often situations occur by which the signal source, whose parameters must be measured is at a great distance from the system of data collection. The reasons for this may be different: the sources of signals which are a major area of study can be operating in harmful environment or can be positioned on moving objects, inaccessible places and others. This requires the transmission of measured parameters to the system for collecting and processing data. This process must be resistant to information distortion or noise. These adverse events may be due to external interference, influence from the transmission medium, the circuits of the transmitter and receiver.

The use of computer communication networks for the transmission of measurement information is based on standard networking protocols used in local and global computer networks. In the model, TCP / IP are several major groups of records. There are possible two main approaches: the use of protocols confirming the information received (TCP / IP) or without confirmation (UDP / IP). First protocol ensure receipt of all information, but any inclusion of additional data in order to improve data transmission slows the performance of this method. The second protocol allows faster transmission of measurement data, but at the expense of losing the package if it can not be restored.

II. APPLICATION OF LABVIEW FOR TRANSMISSION OF MEASUREMENT DATA IN COMPUTER NETWORKS

LabView software product allows to build virtual instruments, which can run in a simulation or real signals [5,6]. To measure the real physical signals can be applied several approaches:

- Using specialized test modules;
- Using standard instrumentation that enables management and transmission of data over serial interfaces (RS232, USB, etc..)
- Using specialized instrumentation using other communication interfaces (GPIB, Serial, TCP/IP, UDP/IP, etc.).

III. BLOCK DIAGRAM FOR REMOTE CONTROL AND MEASUREMENT

Fig. 1 shows a block diagram for remote control and measurement that uses a computer network to realize the transmission of measurement data over long distances. It contains: object, measurement module, LabView program, LabView server, Internet (computer network) and users (students).

The physical parameters that must be measured are converted into electrical quantities using sensors. These electrical parameters are submitted to the measuring module. LabView software is necessary to build an appropriate virtual instrument designed for a specific measurement. In this way the results of measurement of the local computer can be monitored. In order to access the results through the computer network the Internet server (Web Server) must be configured.

IV. WEB SERVER CONFIGURATION

Publication of the results of the measurements are done by LabView Web Server. Access to its options settings is done from the menu Tools -> Options. Configuration includes the following steps:

A. Web Server: Configuration

Enable Web Server, it appears Root Directory, of the HTTP port and Timeout (fig. 2). These parameters can be set to default option or them can be assigned specific values if necessary.

B. Web Server: Visible VIs

Selects the option Allow Access (fig. 3).

C. Web Server: Braowser Access

[1]Ivo Dochev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: idochev@tu-sofia.bg.

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

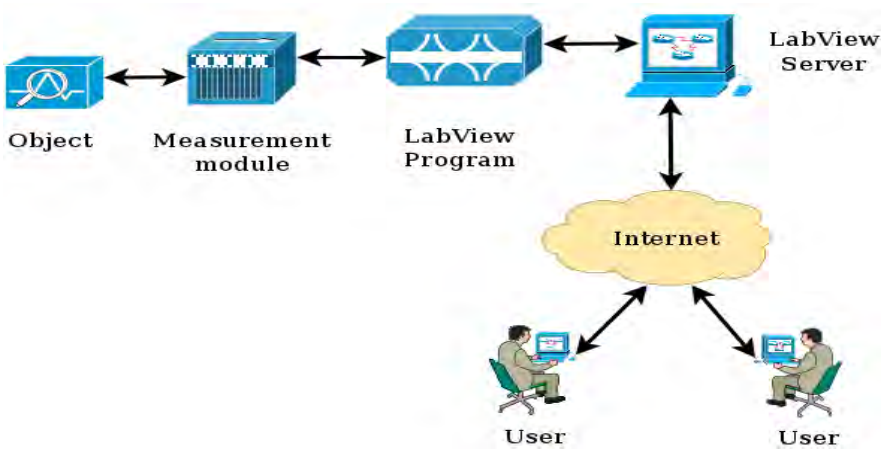


Fig. 1. Block diagram for remote control and measurement that uses a computer network.

Selects the option Allow Access and Controlling (fig. 4). Controlling option can be enabled or disabled if required users

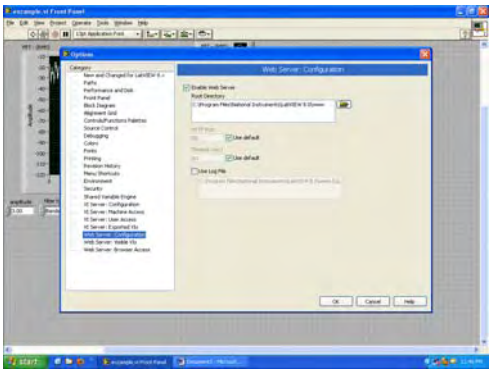


Fig. 2. Web Server: Configuration.

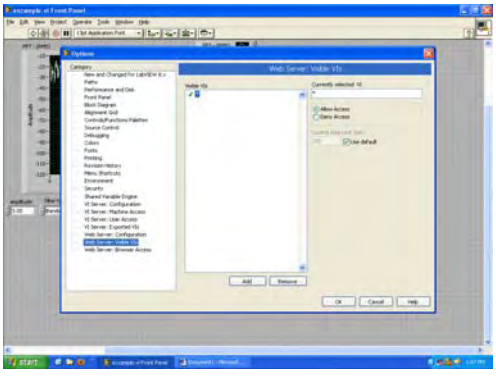


Fig. 3. Web Server: Visible VIs.

to have access to start and stop the measurement process in Web Braowser.

D. VI Server: Machine Access

selects the option Allow Access option and enter the IP address of LabView Web Serve (eg 192.168.1.1) or the name of LabView Web Serve www.example.com (fig. 5).

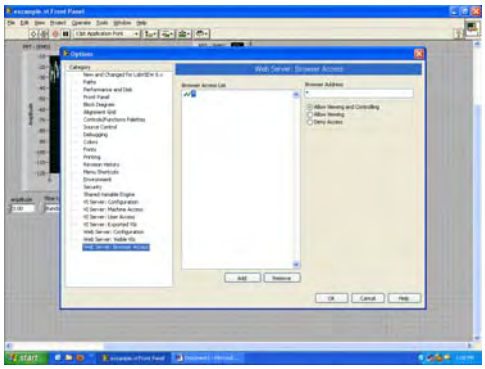


Fig. 4. Web Server: Braowser Access..

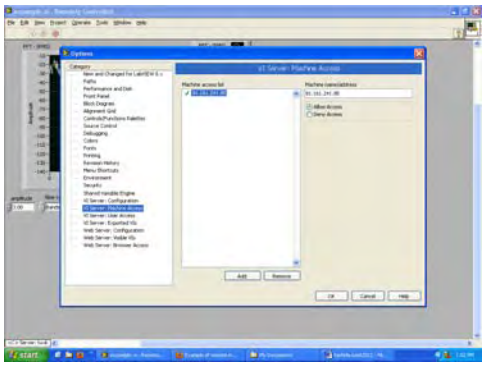


Fig. 5. VI Server: Machine Access.

V. Web Publishing Tool

After configuration of the LabView Web Server is necessary to define the parameters of the web page of the virtual instrument. These parameters can be configured from the menu Tools -> Web Publishing Tool. The configuration contains the following stages:

A. Select VI and Viewing

Choose the virtual instrument, for that the results of the measurement have to be displayed on the website. In this box set and way of displaying the results (static picture, painting renovated over a period of time or dynamic graphics for remote control via web page) (Fig. 6).

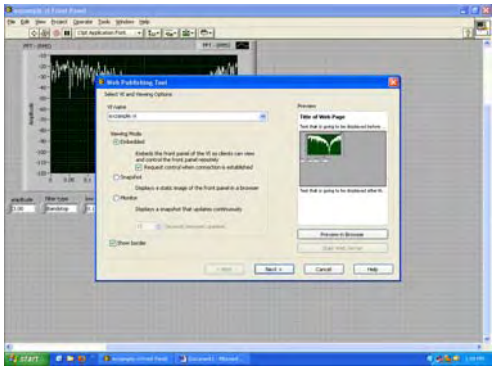


Fig. 6. Select VI and Viewing.

B. Select HTML Output

In the next box: Document title, Header and Footer, whose content is displayed on the website (Fig. 7) have to be filled.

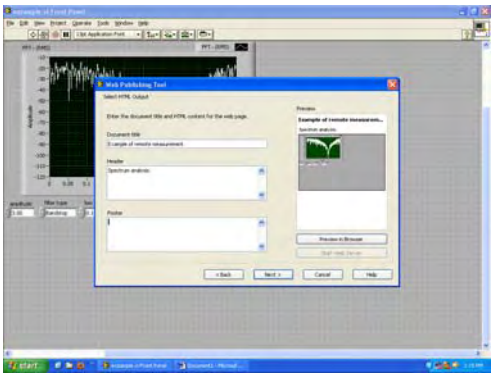


Fig. 7. Select HTML Output.

C. Save the New Web Page

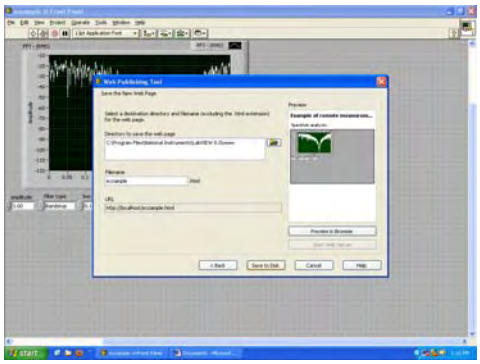


Fig. 8. Save the New Web Page.

The last box set Root Directory to the location of the html page, filename and URL. After setting all parameters examined in this new project is saved (Fig. 8).

In earlier versions of LabView is necessary to make adjustments to the html web page code to allow the data to be monitored dynamically:

```
<HTML>
<HEAD>
<TITLE>Example of remote measurement</TITLE>
<meta html-equiv="refresh" content="5">
</HEAD>
<BODY>
....
```

After the changes the page will automatically refresh every 5 seconds.

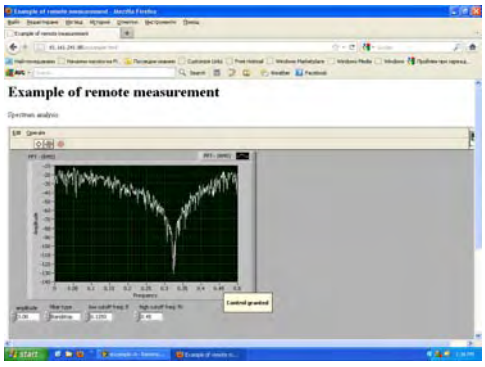


Fig. 9. The virtual instrument for remote monitoring and measurement using computer networks.

Fig. 9 indicated the appearance of the virtual instrument by login in to the page address.

VI. LABORATORY WORK

The aim of the laboratory exercise is to acquaint students with the capabilities of the software LabView for conducting remote measurement. After building a virtual instrument that is suitable for a specific measurement is carried out Web Server and Web Publishing Tool configuration. In case of problems they need to be troubleshoot. After successful completion of all steps students make remote measurements.

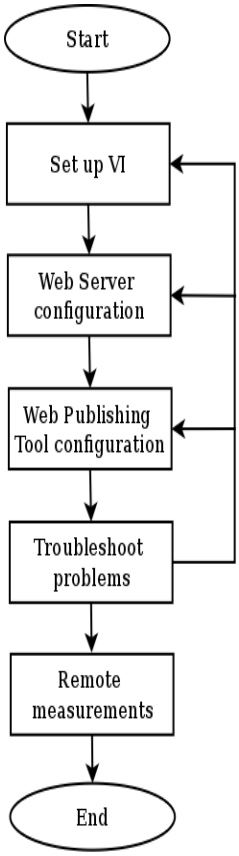


Fig. 10. Sequence of laboratory work.

VII. CONCLUSION

In this article block diagram for remote control and measurement that use computer networks is suggested. Also it is discussed how the configuration of the LabView Web Server must be set. By means of this server remote control and measurement can be realized.

This proposed system can be used in:

- control and diagnosis of household and industrial sites;
- education for students of different remote processes and functional sites;
- early warning in case of disasters.

REFERENCES

[1] LabVIEW help
[2] http://zone.ni.com/devzone/cda/tut/p/id/7350
[3] LabVIEW Measurements Manual, National Instruments Corporation, April 2003 Edition Part Number 322661B-01
[4] 4. LabVIEW Tutorial Manual, National Instruments Corporation, January 1996 Edition Part Number 320998A-01
[5] National Instruments “LabVIEW7 Express. User Manual”, National Instruments Corporation 2003.
[6] National Instruments “LabVIEW7 Express. Measurements Manual”, National Instruments Corporation 2003.

Simulation of third-order dispersion in single optical channel

Kalin Dimitrov[1], Tsvetan Mitsev[2], Lidia Jordanova[3]

Abstract – **This paper presents a simulation of the effects displayed in a single mode optical fiber. The model was created by means of the Matlab™ environment. The model is based on sequential Fourier and inverse Fourier transformations, as well as on other processing with the aim of a numerical solution of the nonlinear Schrodinger equation with acceptable accuracy.**

Keywords – **Single Mode, Fiber, Model, Fourier Transformation**

I. INTRODUCTION

The existence of contemporary civilization and its image is defined to a great extent by the progress in the methods and means of transmitting great quantities of data at great distances. To make this progress possible, there is a need for constant theoretical and practical research. Some of the research and papers aim to improve the parameters of separate system elements, while others consider the system as a whole [1,2]. Examples of this are: application of pre-transmission dispersion management, amplification and in some circumstances, using different optical carrier sources, etc. [3,4]. Some of the papers aim at wavelength division multiplexing (WDM) where multiple wavelengths on a single SM fiber are typically transmitted. Using WDM optical transmission makes possible the good usage of an optic fiber capacity. In WDM there are different approaches in the analysis depending on the wavelength spacing, the power, the types of sources, etc. However, in general, the studying of more complex systems can be done gradually, which in this case means beginning with a single optical channel system followed by an analysis of the interactions between channels. In all possible cases of studying, it is necessary to have an in-depth knowledge of the functions of the separate parts: modulation, demodulation or generation, detection of the optical signals; impairments in either electronic or photonic domains, dynamics of optical fiber, noise sources by optical amplifiers and receiver electronics; effects of electrical and optical filters.

The consideration of all the parts of an optical system in one paper is not possible. It should also be mentioned that there is variety and alternatives in the separate blocks. For example the transmission medium may consist of a variety of fiber types such as the standard SMF ITU-G.652 or the non-zero dispersion shifted fiber ITU-G.655. Another example is the receiving sub-system, which can be a single detector with direct detection or a balanced detector with the respective receiving structure.

To make it methodically possible to study and understand optical systems, it necessary to set up models which describe the actual ones in the best way. The setting up of models also has advantages for the design of much bigger systems because it is a substitute for very expensive experiments. Another advantage is the possibility of an easier quick reconfiguration of the system, as well as for a gradual upgrade.

II. BASIC THEORY

In this paper we consider the modeling of one of the basic parts of the optical communication system, that is the optical fiber.

First, we use the wave equation for light propagation in optical fiber [1]

$$\nabla\times\nabla\times E=-\frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}-\mu_0\frac{\partial^2 P}{\partial t^2}\ . \tag{1}$$

To be able to use (1) it is necessary to define the relation between P and E. For the wavelength interval $0{,}5\div 2\mu m$, where the self resonance of the medium does not go, it is possible to use the following relation

$$P(r,t)=P_L(r,t)+P_{NL}(r,t) \tag{2}$$

where

$$P_L(r,t)=\varepsilon_0\int_{-\infty}^{\infty}\chi^{(1)}(t-t').E(r,t').dt' \tag{3}$$

and

$$P_{NL}(r,t)=\varepsilon_0\iiint_{-\infty}^{\infty}\chi^{(3)}(t-t_1,t-t_2,t-t_3).E(r,t_1)\times \\ \times E(r,t_2).E(r,t_3).dt_1.dt_2.dt_3\ . \tag{4}$$

We will consider pulses with a duration range from 10ns to 10fs. With such impulses there are also dispersion and non-linear effects which affect their form and spectrum. Using (1), (2) and

$$\nabla\times\nabla\times E\equiv-\nabla^2 E \tag{5}$$

we can derive

[1]Kalin Dimitrov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kld@tu-sofia.bg.

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

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

$$\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = \mu_0 \frac{\partial^2 P_L}{\partial t^2} + \mu_0 \frac{\partial^2 P_{NL}}{\partial t^2} \quad . \tag{6}$$

Normally it is considered [1], that P_{NL} changes much less compared to P_L. It is also considered that the optical field retains its polarization along the fibers and the pulse envelope (we will mark the pulse envelope by *A*) changes slowly compared to the optical carrier

$$\begin{aligned} &\frac{\partial A}{\partial z} + \frac{\alpha}{2}A + \beta_1 \frac{\partial A}{\partial t} + \frac{i\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - \frac{i\beta_3}{6}\frac{\partial^3 A}{\partial t^3} - = \\ &= i\gamma\left(1 + \frac{1}{\omega_0}\frac{\partial}{\partial t}\right)\left(A(z,t)\int_{-\infty}^{\infty} R(t')\left|A(z,t-t')\right|^2 dt'\right) \end{aligned} \tag{7}$$

where $\beta(\omega)$ is the wave number (typically it is unknown and the Taylor series approximation is applied)

$$\begin{aligned} &\beta(\omega) = \\ &= \beta_0 + (\omega - \omega_0)\beta_1 + \frac{1}{2}(\omega - \omega_0)^2 \beta_2 + \frac{1}{6}(\omega - \omega_0)^3 \beta_3 + \ldots \end{aligned} \tag{8}$$

In (8) ω_0 is a carrier frequency where $\Delta\omega << \omega_0$ is usually fulfilled, γ is a non-linear parameter and $R(t)$ is Response function, which also depicts the Raman and electronic influences [1,5].

For pulses with duration below 5ps (7) can be simplified to

$$\begin{aligned} &\frac{\partial A}{\partial z} + \frac{\alpha}{2}A + \frac{i\beta_2}{2}\frac{\partial^2 A}{\partial T^2} - \frac{\beta_3}{6}\frac{\partial^3 A}{\partial T^3} - = \\ &= i\gamma\left(|A|^2 A + \frac{i}{\omega_0}\frac{\partial}{\partial T}\left(|A|^2 A\right) - T_R A\frac{\partial |A|^2}{\partial T}\right) \end{aligned} \tag{9}$$

where $T = t - \beta_1 z$.

We can make the following associations in (9): β_3 – third order dispersion, which appears in ultra-short pulses, the addend proportional to $1/\omega_0$ is related to the effects of self steering and shock formation; the addend proportional to T_R depicts the effects delayed Raman response and self frequency shift triggered by the intrapulse Raman scattering. Usually T_R is found experimentally. With pulses with duration over 5ps the addends proportional to T_R and $1/\omega_0$ become too small and we ignore them.

III. NUMERICAL METHODS FOR SIMULATION

The equation (9) is non-linear, partial differential and it has no common analytical solution. Because of this, we use many different numerical methods. A short summary depicting the path of the solution chosen for application is shown in (fig.1).

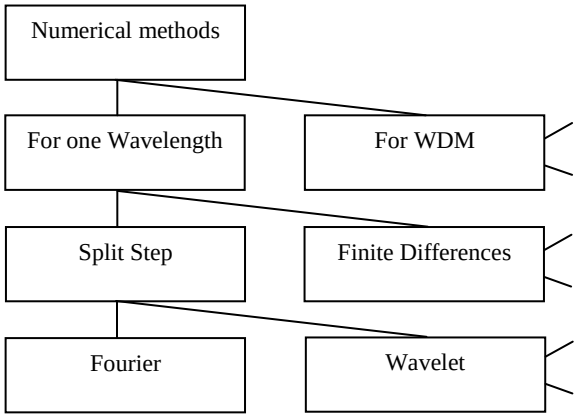


Fig. 1. Classification of the numerical methods used in the solution of equation (10)

We will consider the simulation of a single channel as basic for further simulations of WDM systems. Naturally, with WDM there are further complications resulting from effects such as FWM, XPM etc.

With one single channel the general classification is based on methods with finite differences (perturbation methods) and methods with splitting into parts. The methods with splitting into parts are used because of the quickness and the acceptable accuracy. In particular, for a non-linear dispersive medium is used the Split Step Fourier Method. One of the essential advantages of this method is that FFT is typically used in the calculations, which has been developed and introduced successfully in many systems (for example Matlab™) [5,6,7]. The main idea of this method is that (9) is represented as

$$\frac{\partial A}{\partial z} = \left(\hat{D} + \hat{N}\right)A \quad , \tag{10}$$

where

$\hat{D}$ and $\hat{N}$ are operators respectively:

$$\hat{D} = -\frac{i\beta_2}{2}\frac{\partial}{\partial T^2} + \frac{\beta_3}{6}\frac{\partial^3}{\partial T^3} - \frac{\alpha}{2} \quad , \tag{11}$$

$$\hat{N} = i\gamma\left(|A|^2 + \frac{i}{\omega_0}\frac{1}{A}\frac{\partial}{\partial T}\left(|A|^2 A\right) - T_R\frac{\partial |A|^2}{\partial T}\right) . \tag{12}$$

The optical fiber is split into small lengths h. For its part, h is split into two parts: in the first part functions $\hat{N}$, and $\hat{D} = 0$, while in the second part functions $\hat{D}$, and $\hat{N} = 0$.

We can derive that

$$A(z+h,T) \approx \exp\left(h\hat{D}\right)\exp\left(h\hat{N}\right)A(z,T) \qquad (13)$$

$\exp\left(h\hat{D}\right)$ is calculated in the frequency domain

$$\exp\left(h\hat{D}\right)B(z,T) = F_T^{-1}\exp\left[h\hat{D}(i\omega)\right]F_T B(z,T) \quad .$$

The using of FFT has the advantage that the differential operators in (11) are substituted by powers of $i\omega$.

The accuracy of the calculations improves if we use a modified version of (13)

$$A(z+h,T) \approx$$
$$\approx \exp\left(\frac{h}{2}\hat{D}\right)\exp\left(\int_z^{z+h}\hat{N}(z')dz'\right)\exp\left(\frac{h}{2}\hat{D}\right)A(z,T) \qquad (14)$$

(14) is also known as a symmetrical method with splitting (h is split into three parts). The integral from (14) can be calculated using the method of the trapeziums

$$\int_z^{z+h}\hat{N}(z')dz' \approx \frac{h}{2}\left[\hat{N}(z)+\hat{N}(z+h)\right] \qquad (15)$$

In (15) the problem is that, $\hat{N}(z+h)$ is not known yet, when we have reached $h/2$. This is solved by additional iterations or additional simplification to $\hat{N}(z)$.

IV. Results

For the possibility of the future research, on the basis of the expressions (11)-(15) and what is published in [6] we have developed a simulation in the programming environment of Matlab™. We have created suitable vectors for time and angular frequency. We have considered the case for transmitting of rectangular pulses with different relative durations and the presence of third order dispersion. The choice of the rectangular pulse is not accidental. It is made to show what is the fiber effect using OOK modulation.

From the obtained results it is clear that at a high relative pulse duration dispersion effects occur, but they can easily be avoided. When reducing the length in full agreement with our expectations it is evident that the impulse is significantly influenced both in the time and the frequency domains.

When testing the model with Gaussian pulse are derived results very close to those shown in fig.4.14 in [1].

u0(nt/2:nt/2+50)=1;.

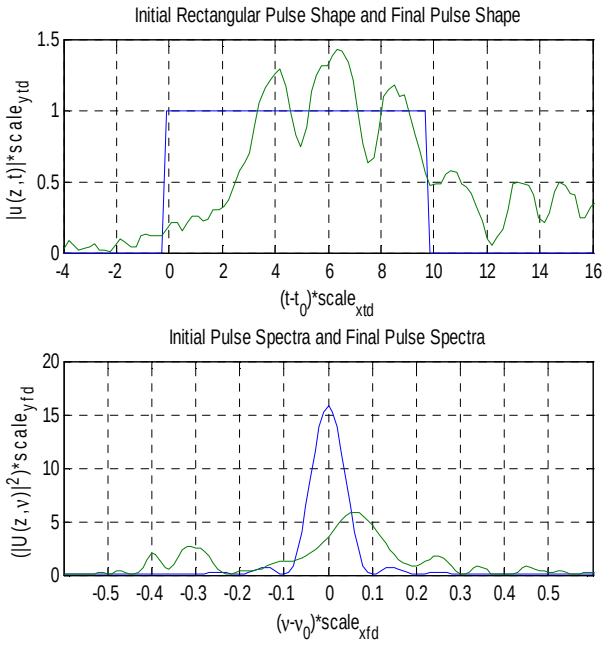


Fig. 2. Testing model with "long" pulse (u0=zeros(nt,1); u0(nt/2:nt/2+50)=1;);

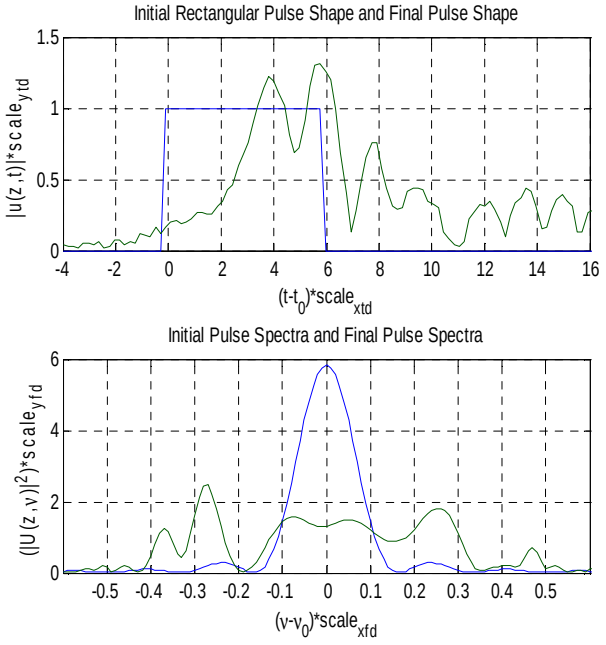


Fig. 3. Testing model with "medium" pulse (u0=zeros(nt,1); u0(nt/2:nt/2+30)=1;);

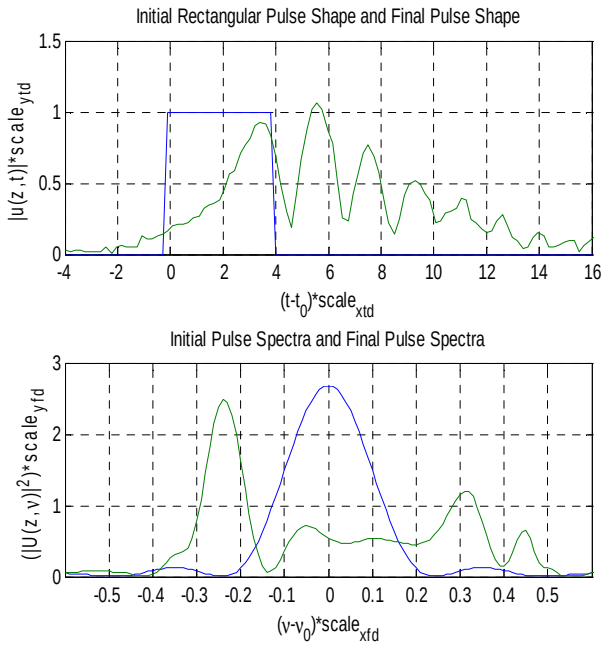


Fig. 4. Testing model with "short" pulse (u0=zeros(nt,1); u0(nt/2:nt/2+20)=1;);

V. Conclusion

The developed model is the first part of a larger system of models. Models of source, receiver, amplifier, etc. are yet to be developed. In the future we also plan to adapt them from the Matlab environment to the environment of Matlab/ Simulink.

Except for the study of distortions in the optical channel of cable communications systems, this model is very suitable for use in training students of a master's course. The model is also suitable for doctoral students.

Acknowledgement

The research described it this paper is supported by the Bulgarian National Science Fund under the contract No ДДВУ 02/74/2010.

References

[1] P. Agrawal, *Nonlinear fiber optics*, Academic Press, 2001.

[2] J. Proakis, M. Salehi, *Contemporary Communication Systems Using Matlab*, PWS Publishing, 1998.

[3] P. Kaewplung, T. Angkaew, K. Kikuchi, Simultaneous suppression of third-order dispersion and sideband instability in single-channel optical fiber transmission by midway optical phase conjugation employing higher order dispersion management, Vol.21,6, pp.1465 - 1473, IEEE, Lightwave Technology, Journal of, 2003.

[4] J. Bel Hadj Tahar, Theoretical and Simulation Approaches for Studying Compensation Strategies of Nonlinear Effects Digital Lightwave Links Using DWDM Technology, Journal of Computer Science 3 (11), pp.887-893, Science Publications, 2007.

[5] L.N. Binh, MATLAB Simulink Simulation Platform for Photonic Transmission Systems, I. J. Communications, Network and System Sciences, 2, pp.97-117, 2009.

[6] L.N. Binh, *Optical Fiber Communication Systems. Theory and Practice with Matlab and Simulink Models*, CRC Press, 2010.

[7] D. Silage, Digital Communication Systems Using Matlab and Simulink, Bookstand Publishing, 2009.

Developing of a Video Information System for the Technical University of Sofia

Kalin Dimitrov [1], Rumen Mironov[2], Alexander Bekjarski[3]

Abstract – **The work described in this paper was carried out at the department of Radiocommunications and Videotechnologies at the Technical University of Sofia, Bulgaria. The major goal of the project was to design, develop, and evaluate a centralized system, based on mini computers to display dynamic information on digital displays located at many locations. Our target displays were located at the entrances of buildings and students offices of faculties.**

Keywords – **Digital Signage System, Video Monitor, Video Information System, IPTV.**

I. INTRODUCTION

There are many commercial and also open source video information and presentation systems described in [1], [2], [3] and [4]. The commercial systems are based on professionally developed software systems. Their main disadvantages are the following:

- the high prices;
- the necessity to adapt the software to the concrete application;
- to use a service provider to which it is necessary to pay monthly or to accept their advertisements;
- to store the client data to the server provider, which is connected with the security of client data and the dependence of the client on the provider.

The open source video information and presentation systems are composed of the necessary open source software modules. The main advantages of the open source video information and presentation systems are the following:

- easy adaptation and expansion according to the concrete video information and presentation system application;
- suitable for academic and education applications, in which it is typical to make the extensions, experiments, practical student works and projects, using different communication technologies, research, etc.

The goal of this article is the development of a video information and presentation system for academic and education applications at the Technical University of Sofia, based on open source programming environment, given the possibility on future expansion and low cost for development in comparison with similar professional video information and presentation systems.

State of the Art. Today, digital media is the most compelling platform to effectively reach employees, students, customers, and partners. This digital media is used to convey important information and messages such as news, training material, and information about upcoming or current events. Digital media is effective because it brings familiarity and closeness to modern communications ([3], [4]). In today's dynamic world digital media has an important role especially for organizations which wish to spread their business throughout the world. For educational organizations digital media plays an important role in informing students about the events, such as: seminars, lectures, meetings, registration deadlines, schedule changes, exams, and sports activities. Digital media represents an emerging new communication technology; in particular digital signage is rapidly gaining popularity today. Digital signage is emerging as a new communication technology. A digital sign is defined as an electronic display that shows information, advertising, or other messages as is shown in [5], [6] and [7]. Digital signage can be implemented using liquid crystal displays, light emitting diodes, digital projection, plasma displays, etc. ([8], [9]). Such digital signage can be used in airports, research organizations, shopping malls, railway stations, and restaurants to dynamically deliver information, graphics, animations, videos, text, and other web contents on a (high quality) display to targeted viewers at a specific time ([10],[11],[12]). Fig. 1 shows the general structure of a digital signage system, including a Web server with a database server, managing server and local client computer to accept and display digital materials, transmitted over the Internet.

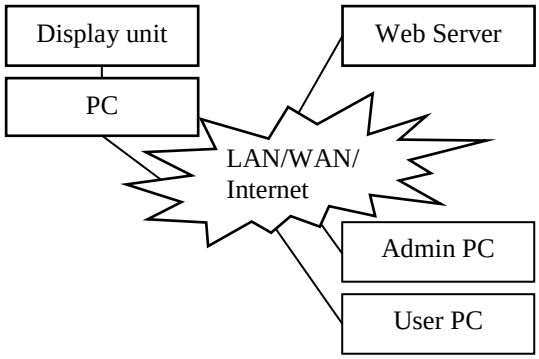


Fig.1. General structure of a digital signage system

The main challenge today for many organizations is the successful deployment and integration of a digital signage network system. In large organizations, such as educational, and research organizations, lots of events are taking place each day. Thus there is a constant need to inform people about what activities are taking place, where they are taking place,

[1]Kalin Dimitrov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: kld@tu-sofia.bg.

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

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

and how to get from where the sign is to this place. Informing the potential audience in these organizations through conventional signage has many disadvantages and drawbacks. Some of these drawbacks and disadvantages are the financial costs of printing, distributing, and removing paper posters; the environmental costs of these posters over their lifecycle; the fixed contents of such printed posters (i.e., any change is expensive as it requires a physical change to the poster - once it is printed); printed posters can only convey static text and pictures. In this research project we designed and developed a prototype system, based on using a PC directly attached to a digital display in order to display web contents. Our target displays were the large screens located at the TU-Sofia buildings.

II. System Description

The main focus and objective of the project was to design, develop, and evaluate a hardware/software solution based on mini computers, to control and display different contents on digital displays. Instances of this system could be located at different locations. The developed Video Information System (VIS-TU) displays different kinds of information, based upon a "playlist" that can be dynamically updated. A secondary goal of this project was to remove the single point of failure of a television (be it analog or IPTV) based system, since each display has an attached processor and local storage containing both the information to be displayed and the local playlist of what is to be displayed. In existing TV based systems when the main server crashes the whole system stops working. More generally this project will explicitly consider how to generate a cycle of information to be displayed, where the information can be adapted based on time, location, and viewers. The granularity of the cycle's schedule (i.e., playlist) is much shorter than that of existing digital signage systems – leading to a more visually dynamic experience for viewers.

The common block scheme of the developed Video Information System is shown on Fig. 2. The main structured blocks, included in the system are:

- Server system for information control (*VIS-TU Server System*). The system includes Content Server, based on Windows Server 2008 OS and an open source digital signage solution (Video Editing system - Xibo), centrally managed via a web administration panel (PHP/MySQL Database) and distributed over a local network or the Internet to one or more clients (.NET 3.5 Framework for Windows XP OS) connected to display hardware (TV Monitors, Projector, IP Cameras etc).
- Client's subsystems for visualization on the Samsung TV monitors in the TU buildings (*VIS-TU Client System*). The systems are based on Windows XP mini computers LEC-7000, configured with Intel Atom N270 1.6 Ghz two cores processors, 512 MB DDR2 RAM, 160 GB local hard disks, two 1 GBps Lan controllers, SVGA video controllers.
- Internet IP Cameras for real time streaming of video information. The live streaming of videos, slide show presentations, advertisements, screen shots etc. are performed using a Windows Media Encoder 9 tool. This is powerful for content producers who want to take advantage of the many innovations, including high-quality multichannel sound, high-definition video quality, new support for mixed-mode voice and music content, and more. For the adoption of broadcasting materials in the client's computers are used Media Player or VLC Player.
- Client's subsystems for remote control and administration of the VIS-TU system. For the purpose of remote administration, management and control of the local computers in a VIS-TU, two software systems - UltraVNC and LogmeIn are used. This approach guarantees the reservation and the security of remote administration and management of local computers in the VIS-TU, in the event that one of the two systems making improper connection with any of the managed computers. As a result of the need for a tool for quick access to configuration settings of the LCD monitors a software system Samsung Monitor for management and control the local computers by the RS232 serial port was developed. The software system was developed with Visual Studio 2010 program environment, in two versions - SamsungONF in command mode for use in remote mode by the programming environment UltraVNC and SamsungMonitor in graphics mode for self-management and control of a separate monitor features (brightness, contrast, sound, volume, power on/off and etc.). In Fig. 3 the starting screen of a SamsungMonitor GUI program is shown.

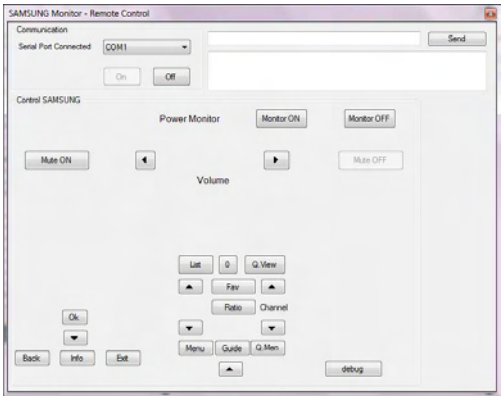


Fig.3. Starting screen of the program SamsungMonitor

- Connection with the university network and possibilities for visualization in computer classes and conference halls.

III. Experimental Results

Digital signage has a very important role in educational organizations to inform their employees and students about upcoming events and to distribute information. It improves campus communications and facilities emergency notifications. It can inform students about upcoming seminars, registration deadlines, exams, and sports activities.

The appropriate layout information shown on video monitors, makes it possible to display current messages, static images (photographs) and moving images (video clips) of the life at TU was created. The created layout is consistent with the fact that the monitors are with plasma screen technology and it is no longer necessary to display static graphic elements. This information was created and edited with the

possibilities of editorial environment in the VIS-TU system that is shown in Fig.4. In Fig. 5 the general appearance of the unit for displays control is shown.

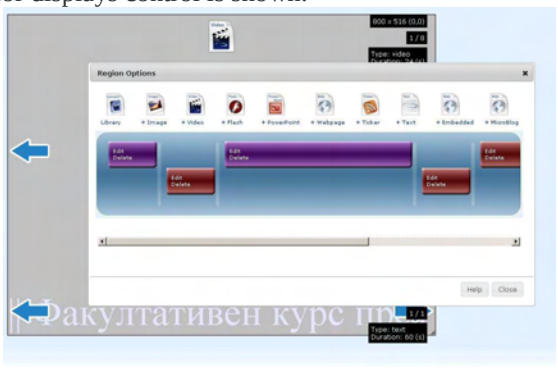


Fig.4. Video Information System editing unit.

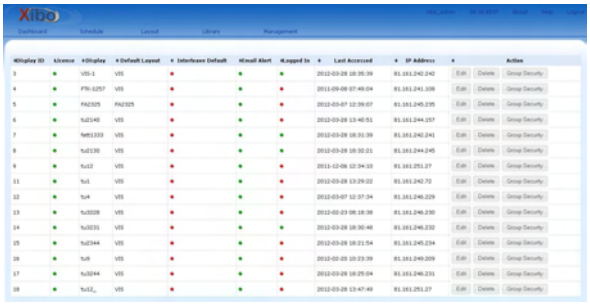


Fig.5. Control displays subsystem.

The result of the correct work of the developed Video Information System is presented in Fig.6.

Fig.6. Example of the work of VIS-TU.

IV. Conclusion

There are many benefits of a web based approach to digital signage. These benefits include efficiency, scalability, low system cost, and low complexity. The system is also scalable in the sense that we can easily add more displays to the system.

In this project we have successfully implemented a complete digital display information system. We have designed, developed, and evaluated a hardware & software solution based on a PC, to control and display different contents or other dynamic information on digital displays located at different locations.

In addition there are a number of areas for future work that would make the system more attractive and competitive with commercial digital signage systems.

There should be an easy means of setting up a backup server and offering hot-failover to this alternate server – should the main server fail (for example, using a network dispatcher – as it is often used to support load balancing across a set of web servers).

Acknowledgement

The acknowledgements are expressed to the colleagues from International Relations Section and Center for Information Resources.

This work was supported by Technical University under Contract № 111TPF001-07: "Development of a Video Information System for the Technical University of Sofia".

References

[1] Schaeffler J., Digital Signage: Software, Networks, Advertising, and Displays: A Primer for Understanding the Business, Focal Press, 2008.

[2] Lundström L.I., Digital Signage Broadcasting: Content Management and Distribution Techniques, Focal Press, 2008.

[3] Grant A., J. Meadows, Communication Technology Update and Fundamentals, Focal Press, 2010.

[4] Solomon M., D. Kim, Fundamentals of Communications and Networking, Jones & Bartlett Publishers, 2011.

[5] Huang Y.T., Tsou C.W., Adoption of digital signage and application technology in medium-sized enterprises in Taiwan, IEEE, Computers and Industrial Engineering (CIE), 2010 40th International Conference on, pp.1-5, 2010.

[6] Inoue H., K. Suzuki, K. Sakata, K. Maeda, Development of a Digital Signage System for Automatic Collection and Distribution of Its Content from the Existing Digital Contents and Its Field Trials, Applications and the Internet (SAINT),IEEE/IPSJ 11th International Symposium on, pp. 463-468, 2011.

[7] Sato H., M. Urata, K. Yoguchi, N. Arakawa, N. Kanamaru, N. Uchida, Linking digital signage with mobile phones, IEEE, Intelligence in Next Generation Networks (ICIN), 2011 15th International Conference on, pp. 86-91, 2011.

[8] Maeda K., M. Nishi, T. Yoshida, K. Suzuki, H. Inoue, Digital Signage with Audience Detection Using TV Broadcasting Waves, Applications and the Internet (SAINT), 2010 10th IEEE/IPSJ International Symposium on, pp. 225-228, 2010.

[9] Dennis C., A. Newmanb, R. Michon, J.J. Brakus, L.T. Wright, The mediating effects of perception and emotion: Digital signage in mall atmospherics, Elsevier, Journal of Retailingand Consumer Services, pp.205-215, 2010.

[10] Storz O., A. Friday, N. Davies, Mobile Computing and Ambient Intelligence. Supporting content scheduling on situated public displays, Elsevier, Computers & Graphics, 30, pp.681–691, 2006.

[11] Morrison A., T. Arnall, Visualizations of Digital Interaction in Daily Life, Elsevier, Computers and Composition, 28, pp.224-234, 2011.

[12] Dupin F., Digital signage: the right information in all the right places, ITU-T Technology Watch Report, Switzerland, Geneva, 2011.

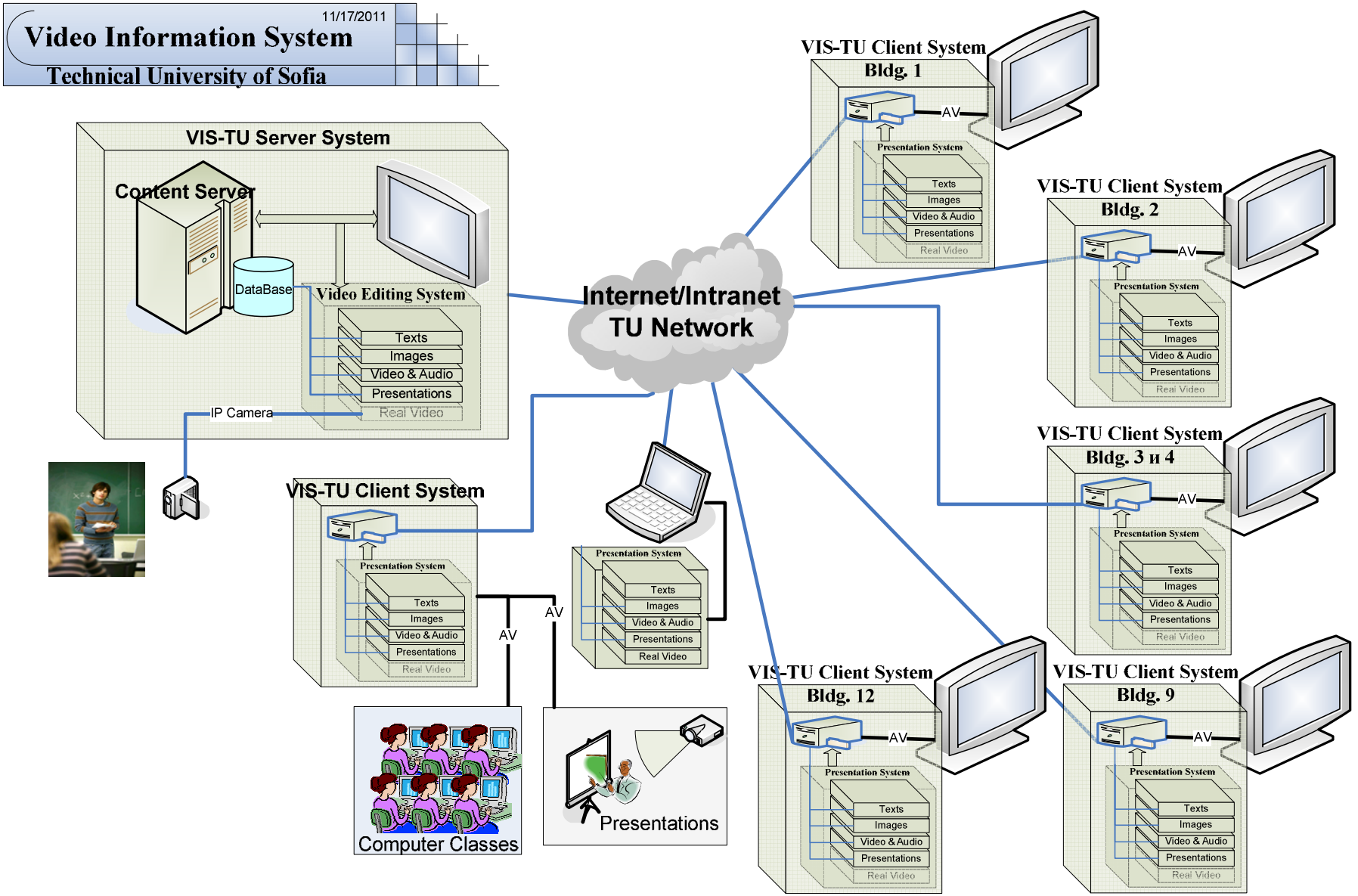


Fig. 2. Block scheme of the developed Video Information System

AUTHOR INDEX

H

I

J

K

L

M

N

O

P

R

S

T

V

Y

Z