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

www.icestconf.org

ICEST 2010

23 - 26 June, Ohrid, Macedonia

Proceedings of Papers

VOLUME 2

Bitola, 2010

XLV INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION,

COMMUNICATION AND ENERGY SYSTEMS AND TECHNOLOGIES

- ICEST 2010 -

Proceedings of Papers	**– Volume 2 of 2 volumes**
Editor:	**Prof. Dr. Cvetko Mitrovski**
Technical Editor:	**Jove Pargovski**
Published by:	**Faculty of Technical Sciences – Bitola**
Printed by:	**OFFICE 1 - BITOLA**
Print run:	**50**
ISBN:	**978-9989-786-58-7**

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

621.39(063)
004.383.3(063)
004.73/.78(063)
006.91:[004.85:004.7(063)
621.313(063)
681.51(063)
621.38(063)
620.97(063)

XLV International Scientific Conference on Information, Communication and Energy Systems and Technologies (45 ; 2010 ; Ohrid) ICEST 2010 : proceedings of papers / XLV International Scientific
Conference on Information, Communication and Energy Systems and Technologies, 23-26 Juni, Ohrid, Macedonia ; [editor Cvetko Mitrovski]. - Bitola : Faculty of Technical Sciences, 2010. - 2 св. (956 стр.) : илустр. ; 29 см

Фусноти кон текстот. - Библиографија кон трудовите. - Регистри

ISBN 978-9989-786-57-0 (вол. I)
ISBN 978-9989-786-58-7 (вол. II)
1. Mitrovski, Cvetko [уредник]
а) Телекомуникации - Собири б) Дигитална обработка на сигнали - Собири в) Компјутерски мрежи - Интернет - Собири г) Метрологија - Системи за далечинско управување - Собири д) Електрични машини - Собири ѓ) Автоматски контролни системи - Собири ђ) Електроника - Собири е) Обновливи извори на енергија - Собири
COBISS.MK-ID 84633354

TABLE OF CONTENTS

VOLUME 2

POSTER SESSION II - Telecommunications Systems and Technology

POSTER SESSION III - Radio Communications, Microwave Technique and Antennas

POSTER SESSION IV - Signal Processing

POSTER SESSION V - Robotics

POSTER SESSION VI - Power Transmission and Distribution Systems

POSTER SESSION VII - Computer Systems and Internet Technologies I

POSTER SESSION VIII - Computer Systems and Internet Technologies II

POSTER SESSION IX - Computer Systems and Internet Technologies III

POSTER SESSION X - Computer Graphics

POSTER SESSION XI - Electronic Components, Systems and Technologies

POSTER SESSION XII - Industrial Electronics

POSTER SESSION XIII - Renewable Energies

POSTER SESSION XIV - Education Quality

POSTER SESSION PO II

PO II - Telecommunications Systems and Technology

Hybrid ARQ Schemes Using Diagonally Interleaved Turbo Product Codes

Nikolay Kostov[1], Slava Yordanova[2], Ginka Marinova[3]

Abstract: **In this paper new hybrid automatic repeat request (HARQ) schemes using interleaved two-dimensional single parity check turbo product codes are proposed. Throughput and bit error rate of the proposed schemes are studied through simulations and are compared to the conventional forward error correction schemes. It is shown that using these HARQ schemes a significant energy gain over the corresponding forward error correction schemes is obtained.**

Keywords: **Hybrid automatic repeat request, interleaving, turbo product codes, simulation.**

I. INTRODUCTION

Automatic repeat request (ARQ) schemes are used to achieve near error-free transmission and to reduce link margins when the channel characteristics are poorly predictable [1]. A hybrid ARQ (HARQ) scheme uses a forward error correction (FEC) code in conjunction with a retransmission scheme (ARQ). Typically a cyclic redundancy check (CRC) code is used for frame error detection and this is an example of the so-called two-code approach since two different error control schemes are used for HARQ purpose [2]. Another HARQ method is the so-called one-code approach since only one error control code is used for both error correction and error detection [3]. A class of interleaved single parity check turbo product codes was introduced in [4], [5]. Typically a pseudorandom interleaver is used with the examined turbo product codes.

In this paper, new HARQ schemes based on interleaved two-dimensional single parity check turbo product codes (2D-ITPC) are proposed. The above mentioned one-code approach is used along with a diagonal interleaver. As it will be shown, these HARQ schemes produce a significant energy gain over the corresponding forward error correction schemes.

II. SYSTEM MODEL

First consider the encoding/decoding processes of the conventional interleaved 2D-ITPC [4], [5]. In case of parallel concatenation, the data frame to be transmitted is first encoded by a single 2D single parity check product code.

[1]Nikolay T. Kostov is with the Technical University of Varna, Department of Radio engineering, Studentska Street 1, Varna 9010, Bulgaria, E-mail: n_kostov@mail.bg

[2]Slava M. Yordanova is with the Technical University of Varna, Department of Computer Science and Technologies, Studentska Street 1, Varna 9010, Bulgaria, E-mail:slava_y@abv.bg

[3]Ginka K. Marinova is with the Technical University of Varna, Department of Computer Science and Technologies, Studentska Street 1, Varna 9010, Bulgaria, E-mail: gin_kaleva@abv.bg

Then it is interleaved end re-encoded by the same 2D single parity check product code. In case of serial concatenation, the data frame to be transmitted is first encoded by a 2D $(n-1, n-2)$ single parity check product code. Further the data and parity bits are interleaved end re-encoded using a 2D $(n, n-1)$ 2D single parity check product code. Thus, the code rate of this serially concatenated 2D-ITPC is

$$R_c = \left[\frac{n-2}{n}\right]^2 \quad (1)$$

while the data block length and the interleaver size are $(n-2)^2$ and $(n-1)^2$, respectively. In both cases the overall interleaved product code is composed of the original data bits and all parity bits from the single parity check codes. There is no noticeable performance difference between these serial and parallel concatenated 2D-ITPC. Therefore, in this paper, only serially concatenated shames will be examined.

In order to improve the performance of the considered HARQ schemes a diagonal interleaver is proposed. The interleaver is of size N ($N = m^2$) and has the following property: any two or more bits in the initial non interleaved block, located in the same raw or column, do not lie in a raw or column in the interleaved block.

Let us now consider binary phase shift keying (BPSK) transmission of the coded data via an additive white Gaussian noise (AWGN) channel. For an AWGN channel model the input/output relationship can be expressed as

$$r_i = \sqrt{E_s}(2b_i - 1) + n_i \ , \quad (2)$$

where r_i is the received symbol, E_s is the energy per code symbol, $(2b_i - 1)$ is the binary phase shift keying modulated code symbol and n_i is a zero-mean Gaussian variable with variance $\sigma^2 = N_0 / 2$.

The decoding process is the following. The original (non interleaved) noisy frame is first decoded using a soft-input/soft-output (SISO) decoding method [6] for a single decoding cycle. Then the interleaved data frame is also SISO decoded for a single decoding cycle and so on. The two constituent decoders exchange the so-called extrinsic information (the error correction term gained from the decoding) at each full iteration and the decoding iterations are executed until a predetermined stopping criteria is satisfied.

A simplified diagram of the turbo decoder is shown in Fig.1.

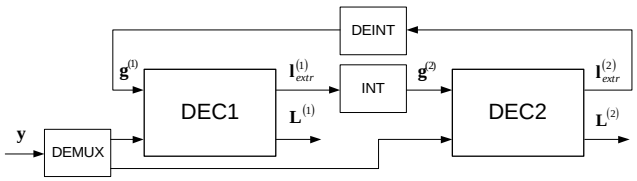


Fig.1. Turbo decoder block diagram.

The log-likelihood ratio (LLR) L_i at the output of the SISO decoder can be represented as [5, 6]

$$L_i = y_i + g_i + l_i, \tag{3}$$

where y_i is the weighted channel observation, g_i is the *a priori* information and l_i is the so-called extrinsic information gained by the current stage of decoding. Both decoders (DEC1 and DEC2) in Fig.1 use the algorithm described in [6] form an estimate of the LLR of each bit encoded by the single parity check equations. The extrinsic information l_i of the *i*th bit associated with a single parity check equation is computed as [6]

$$l_i = (-1) \cdot (\min_{\substack{j=1,\ldots,n \\ j \neq i}} |l_j|) \cdot \prod_{\substack{j=1 \\ j \neq i}}^{n} \operatorname{sign}(l_j). \tag{4}$$

According to (4) the magnitude of the extrinsic information for a particular code element is equal to the minimum magnitude of all of the other parity elements. The sign of the extrinsic information for a particular code element is equal to the sign of the element itself if the parity of the overall equation is satisfied, and opposite to the sign of the element if the overall parity fails. The final soft decision Λ_i of the *i*th data bit is given by

$$\Lambda_i = L_i^{ch} + L_i^1 + L_i^2, \tag{5}$$

where L_i^{ch} is the noisy channel observation for the *i*th data bit and L_i^1, L_i^2 are the extrinsic information terms from the first and second two-dimensional decoder, respectively. The corresponding hard decision a_i of the *i*th data bit is

$$a_i = \begin{cases} 1, \ if \ \Lambda_i > 0 \\ 0, \ if \ \Lambda_i < 0. \end{cases} \tag{6}$$

Now consider the retransmission algorithm associated with the proposed HARQ schemes. The algorithm is as follows:

1. Apply SISO iterative decoding (with a predetermined maximum number of iterations) on the received encoded frame, checking all parity equations of the component single parity check codes after each full iteration. If all parity equations are satisfied, (e.g., the stopping criterion is fulfilled) go to 3. Otherwise proceed to 2.

2. Request a retransmission of the decoded frame by sending a negative acknowledgement to the transmitter via the feedback channel. If the maximum number of retransmissions is reached, proceed to 3.

3. Output hard quantized data.

It can be observed that with this retransmission algorithm the frame error detection (and consequently the retransmission request) is based on the checks of all parity equations of the overall interleaved product code. Because of the interleaving, a significant reduction of undetected erroneous frames can be expected for the considered error control scheme as compared to the conventional two-dimensional product codes. Consequently, an efficient ARQ mode operation could be implemented for this interleaved turbo product codes. As performance results show further improvement is possible if the conventional pseudorandom interleaver is replaced with the proposed diagonal interleaver.

III. THE SIMULATION CODE

MATLAB-based Monte Carlo simulations were executed in order to estimate the performance of the proposed HARQ schemes. The simulations are based on BPSK signaling over the AWGN channel with at most two retransmissions allowed and an error-free feedback channel. Soft equal gain combining of the retransmitted frames was applied in the receiver.

Below are given the maim functions in the simulation, e.g. **encode_msg.m** (performing 2D single parity check encoding of data), **diag_perm.m** (performing a diagonal permutation) and **dec_2D** (performing soft decoding of a 2D data block).

```
function  PC = encode_msg(A)

M = length(A);
m = sqrt(M);
n = m + 1;
A = reshape(A,m,m)';
PC = zeros(n);
PC (1:m,1:m) = A;
 for k = 1:m  % row
    PC (k,n) = mod(sum(A(k,:)),2);
 end

 for k = 1:n  % column
    PC (n,k) = mod(sum(PC (1:m,k)),2);
 end

function f = diag_perm(mm)
 m=sqrt(mm); a= 1:mm;
b = reshape(a,m,m);
b= b'; c = [b b];
d = zeros(m,m);
e = zeros(m,m);
f = zeros(m,m);
 for i = 1:m
   e = c(:,i:i+m-1);
   d(i,:) = diag(e)';
 end
f(1,:) = d(1,:);
 for i = 1:m-1
    f(i+1,1:m-i) = d(i+1,i+1:m);
   f(i+1,m-i+1:m) = d(i+1,1:i);
 end
```

For a data block of 25 elements with one dimensional indexing given below

```
 1   2   3   4   5
 6   7   8   9  10
11  12  13  14  15
16  17  18  19  20
21  22  23  24  25
```

the function **diag_perm.m** produces

```
 1   7  13  19  25
 8  14  20  21   2
15  16  22   3   9
17  23   4  10  11
24   5   6  12  18
```

```
function  [Le_hor, Le_ver] = dec_2D(Ch_LLR)

CodeLLR = Ch_LLR;
    for k = 1:n  % decode rows
      C10 = CodeLLR (k,:);
      for j = 1:n
        C1 = C10;  C1(j) = [];
         Le_hor(k,j) = prod (sign (C1))*min(abs(C1));
      end
    end
    CodeLLRtmp = CodeLLR + Le_hor;
    for k = 1:n  %decode columns
      C10 =  CodeLLRtmp (:,k);
      for j = 1:n
        C1 = C10;  C1(j) = [];
        Le_ver(j,k) = prod (sign (C1))*min(abs(C1));
      end
    end
```

IV. PERFORMANCE RESULTS

MATLAB-based Monte Carlo simulations were executed in order to estimate the performance of the proposed HARQ schemes. The simulations are based on BPSK signaling over the AWGN channel with at most two retransmissions allowed and an error-free feedback channel. Soft equal gain combining of the retransmitted frames was applied in the receiver.

Two basic HARQ schemes with the above mentioned diagonal interleaver were studied: HARQ1 with a parent code rate $R_c \approx 3/4$ and HARQ2 with a parent code rate $R_c \approx 4/5$. In this paper the throughput is defined to be the effective code rate R_{eff} of the corresponding HARQ schemes. Obviously $R_{eff} < R_c$ and $R_{eff} = R_c$ only in case of no retransmissions.

Throughput versus signal-to-noise ratio (SNR) performance of HARQ1 and HARQ2 schemes is shown in Fig.1. The required effective SNR's for the HARQ schemes, necessary for achieving typical bit error rates (BER's), are shown in Table 1 along with performance of the parent diagonally interleaved turbo product codes (DITPC).

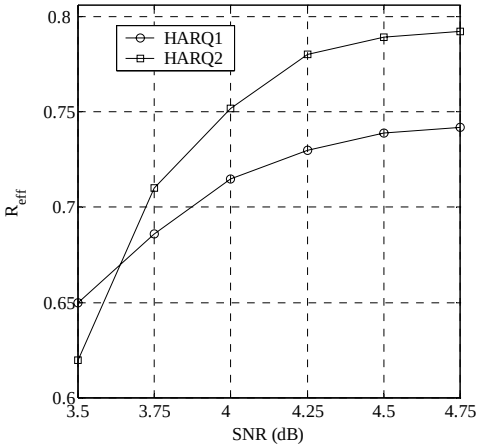


Fig.2. Throughput versus SNR performance of HARQ1 and HARQ2 schemes.

Table 1. Estimated HARQ schemes performance

	HARQ1	DITPC1 $R_c \approx 3/4$	HARQ2	DITPC2 $R_c \approx 4/5$
SNR_{eff} (dB) for $BER = 10^{-5}$	4.2	5.1	4.2	4.95
SNR_{eff} (dB) for $BER = 10^{-6}$	4.6	5.45	4.55	5.3

According to the data in Table 1, an energy gain of at least 0.75 dB is obtained with the considered HARQ schemes over the parent DITPC schemes. Further, an energy gain of approximately 0.25 dB is obtained when the considered diagonal interleaver is used instead of the conventional pseudorandom interleaver. At $BER \le 10^{-5}$ the effective code rate R_{eff} of both HARQ schemes is close to code rate of the parent schemes (e.g., $R_{eff} \approx R_c$) because of the relatively small overall number of retransmissions.

V. CONCLUSION

New simple HARQ schemes based on interleaved two-dimensional single parity check product codes have been proposed in this paper. It is shown that energy gain of 0.75 dB or even higher is obtained over the parent forward error correcting schemes with practically no data rate reduction. Further, better performance results in terms of frame error rate and bit error rate were obtained when the conventional pseudorandom interleaver is replaced with the considered diagonal interleaver. Finally, it should be mentioned that for moderate-to-high SNR's the BER performance of the considered "one-code approach" HARQ schemes is

dominated by the undetectable erroneous frames and no significant performance improvement can be expected with more than two retransmissions allowed.

REFERENCES

[1] S. Lin, D. Costello, M. Miller, "Automatic-repeat-request error control schemes", IEEE Communications Magazine, No12, pp. 5-16, 1984.

[2] S. Lin, D. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice-Hall Inc., N.J., 1983.

[3] H. Yamamoto, K. Itoh, "Viterbi decoding algorithm for convolutional codes with repeat request", IEEE Transactions on Information Theory, No 5 pp. 540-547, 1980.

[4] D. Rankin, *Single parity check product codes and iterative decoding*, Ph.D. thesis, University of Canterbury, Christchurch, New Zealand, 2001.

[5] D. Rankin, T. Gulliver, D. Taylor, "Parallel and serial concatenated single parity check product codes", EURASIP Journal of Applied Signal Processing, No 6, pp. 775-78, 2005.

[6] B. Sklar, "A primer on turbo code concepts", IEEE Communications Magazine 12, pp. 94-102, 1997.

Comparison Studies on Path Recovery Schemes in MPLS Network

Veneta P. Aleksieva[1]

***Abstract* – In this paper are discussed the problems in the existing MPLS/Multi Protocol Label Switching/ recovery mechanisms. The simulation-based experiment studies and compares the performance of existing MPLS recovery models, based on the system of parameters. It uses three criteria of the comparison.**

***Keywords* – MPLS recovery mechanisms, MPLS network**

I. INTRODUCTION

Internet is currently requiring a means for providing different users with different service levels. Traffic with high requirements for example delay, jitter and bandwidth has to be treated with a certain priority, while the traditional best effort services are still available. Many types of multimedia such as video conference or voice conference have become more important and have been widely used nowadays. Generally network operators aim to provide the fastest, most stable and the best protection mechanism that can be provided at a resource consumed.

Diffserv enables network traffic to be classified into different priority levels and then applies different scheduling and queuing mechanisms at the nodes according to the priority level [1].

The ToS field in the IP header is used to mark a packet and then is used as an indication of how the packet is forwarded.

MPLS is used as a traffic engineering tool to direct traffic in a network in a more efficient way then original IP shortest path routing. Path in the network can be reserved for traffic that is sensitive, and links and router that are more secure and not known to fail can be used for this kind of traffic. If MPLS is used, it sets up LSPs /Label Switched Paths/ along links with available resources, this ensures that bandwidth is always available for a particular flow to avoid congestion [2].

Most of the techniques used for recovery in MPLS are already available at other network layers. It seems that it is not necessary to implement these mechanisms again in another layer, but the main reason is: MPLS operates between layer 2 and layer 3 in the OSI model. This gives us now possibilities in network recovery with new functionality. MPLS is designed to work with many different network technologies and has its own mechanisms for recovery, independent of other layer mechanisms. Here they are faster and non-visible into higher layers.

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

A number of recovery schemes for MPLS have been proposed in recent years and most of the current schemes focus only on wire network and rarely there are solutions for multiple failures recovery in dynamic topologies based on wireless. The main goal of path recovery schemes is to minimize time of service disruption, which depends on the time to detect failure, time to notify, time to compute backup and time to switch traffic to the new one. MPLS recovery must have the ability to ensure recovery from a link or node failure with minimal disruption to the data traffic. Often the limit of 50 ms is considered because this has been set as the longest acceptable disruption to voice traffic and is used as a limit for recovery time in SONET/SDH networks.

In this approach is presented comparison of some MPLS recovery technique and their advantages and disadvantages.

II. NETWORK RECOVERY OF STATIC AND DYNAMIC MPLS NETWORK

Actually, to fully provide QoS in the network there also needs to be a guarantee for what happens with traffic in the case if congestion is caused by link or node failures:

First, the network must be able to detect the failure.

Then the nodes that detect this failure must send a message to the certain nodes in the network of the failure. Which nodes will be notified about the failure depends on used recovery technique.

Next, the backup path must be computed.

Finally, a node must send traffic on the backup path instead the previous path.

MPLS recovery provides different levels of service, based on their service requirements. It should give the flexibility to select the recovery mechanism, choose the granularity at which traffic is protected and choose the specific types of traffic that are protected in order to give operations more control over that tradeoff. [3]

If a failure occurs in a network there must be a way to detect this so that the recovery operation can start. But failure detection depends on the type of failure and may be done by the failing node, at a node adjacent to the failure or at a configured point of repair in the network. MPLS recovery techniques are 2 main types [4,5]:

- **Protection Switching** – This is pre-establishing a recovery path based on network routing policies and requirements of the traffic.
- **Rerouting** – This is establishing new paths or path segments on demand for restoring traffic after the failure.

Both have advantages and disadvantages, which are shown in Table 1 and Table 2.

TABLE I
COMPARISON OF MPLS RECOVERY MODELS

Recovery model	Protection Switching		Rerouting	
Recovery Path Setup	Before fault		After fault	
Restoration of service	Lacks efficient use of network resources as the recovery path is setup		Optimizing the recovery path	
	Fast restoration of resources		Slower restoration of resources	
Reservation of resources	Before a failure occurs		Does not reserve any resources in the network, but they may not be available at the time of recovery path	
Recovery path type	1+1	1:1, 1:n, m:n	Pre computed	Established on demand
Recovery time	fourth	third	fastest	Second fast
Resource utilization optimization	first	second	third	fourth
Comparison with non MPLS based recovery mechanism	Like lower layer recovery		Like rerouting at the network (IP) layer	

In the table below is presented different comparison:

TABLE II
COMPARISON OF MPLS RECOVERY MODELS

Restoration and repair method	Resource requirement	Speed of repair	Packet loss	Length of protection path
Dynamic Local Repair	No	Slow	Minimum	Might not be theSP available
Dynamic Global Repair	No	Slow+ FIS	High	Path is shortest available
Fast rerouting local	Yes, if not shared	Fast	Minimum	May not be the optimal
Fast rerouting global	Yes, if not shared	Fast, depends of FIS	High	Better than fast rerouting global

There are two techniques to set-up recovery path in MPLS[6], but in both the recovery path is set up from the PSL /Path Switch Label Switch Router/ to the PML /Path Marge Label Switch Router/, but in different ways put the labels:

1. **Splicing** – the PSL change its forwarding table when the recovery path shall be used to forward packets on the recovery path instead of the failed working path. A new outgoing interface and a new label are used by the PSL to forward packets on the recovery path.

2. **Stacking** – the PSL also update its forwarding table to use a new outgoing label and a new outgoing interface for the affected LSP, but the new label for the recovery path is pushed on top of the label that would have been used for the failed working path. The next LSR /Label Switch Router/ in the recovery path pops the label stack revealing the old working path label before it forwards the packet to the PML, which doesn't know about the failure because it has the same label as a packet from the working path. Actually, this technique works if PML uses global label space and PSL has to know the label it would have used on the working path for PML.

Each of these techniques has advantages and disadvantages and it is one significant requirement for traffic management to support QoS guaranteed tunnels, according to link or node failure or topology changes. There are some proposed models for MPLS recovery, which is presented follow.

All models can protect a working path end-to-end in one MPLS domain, but there is no protection for node failures on the ingress or egress LSR. This will be a problem if it must recover the paths cross multiple MPLS domains.

In this research is used Network Simulator version 2 [7] for MPLS. With this tool are compared seven models:

1. **Makam's model** [8] – This is global recovery with protection switching, because it builds a global recovery path between the ingress and egress routers. This model has proposals for both a pre setup (protection switched) recovery path and a dynamically established (rerouted) recovery path. When a failure is detected anywhere along the working path, a fault indication signal /FIS/ is used to travel information about the occurrence of the failure to the PSL. Then the PSL is responsible for switching traffic over to the recovery path. The traffic will be sent down the failed working path until this FIS has been received by the PSL. This will result in dropped packages at the LSR that is upstream of the failure, as this node does not have any forwarding information for these packages since the downstream node is not reachable. If the failure is situated far away from the point of repair and the transmission rate is high, the number of packets dropped can be very high.

2. **Huang's model** [9] –This model develops a notification tree in a global or segment protected environment using 1:1 protection. The reverse notification tree is a point to multipoint tree rooted at the PML along which a FIS can be sent to the PSLs affected by a failure. In a case of global protection, the node that detects the failure has to communicate from the point of failure to the PSL upstream in the working path. As LSP are setup unidirectional there has to be information of how the FIS shall be sent upstream.

3. **Haskin's model** [10]- The idea of reverse backup is to reverse traffic at the point of failure in the working path, back to the PSL (ingress LSP). As soon as a LSR detects a failure on the working path, it redirects the incoming traffic on to an alternative LSP that is setup in the reverse direction of the working path. When the reversed traffic reaches the PSL, it forwards this traffic on to a global protection path. Both the reverse path and the global protection path are pre reserved. When a failure is detected, the traffic will be switched onto an alternative path by protection switching directly. In this model

both 1:1 protection and 1:N protection can be achieved. But this model has one disadvantage- until the PSL receives any of the reversed traffic, packages will be forwarded on the broken working path. When the PSL receives traffic from the reversed path, it will start to forward incoming traffic onto the global backup path. For a short period the incoming traffic will be mixed with the reversed traffic as it is forwarded on the recovery path.

4. **Hundessa's model** [11]- When a failure is detected by a LSR, the packets that would have been forwarded on the failed path are returned to the PSL via a reversed backup path as in Haskin's model. But when the first packet, acting as a FIS arrives in the reverse direction at an upstream LSR, that LSR tags the next packet, it sends out on the working path. The next packets it receives, which belong to the same working path, are buffered. The last packet that an LSR sends out on the broken working path is tagged by setting a bit in the EXP field in the MPLS header.

5. **Fast reroute** - End-to-end recovery paths needs to be pre-setup for each link or node in the working path. In extensions to the RSVP-TE protocol are defined to establish an LSP with end-to-end fast reroute backup tunnels. Two techniques are described in this method - the one-to-one backup model and the facility backup model.[12] Here is used one-to-one backup model.

6. **Sa-Ngiamsak's model** [13] – This recovery technique focus on multiple point of failure which may frequently occur on such dynamic network. Its name is Modified Flexible MPLS Signaling (MFMS). If multiple mobile nodes failed, the FMS may become malfunction or can not find a feasible recovery path. This work in five phases: LSP setup, Failure detection, Failure recovery, Ingress coordination and LSP Refresh and Recovery abort. In the simulation of this model it is accepted fixed transmission rate among mobile network nodes.

7. **Nagarajan's model** [14]- This model is created for recovery in dynamic network topologies. It uses new flexible signalling protocol for LSP rerouting in dynamic network environments. The signalling protocol recovers from node and link failures reactively, taking a local approach to LSP reestablishment.

III. Results of MPLS Recovery Models Comparison

The performance of the different MPLS recovery models is evaluated through simulations. In this research is used Network Simulator version 2.26 [7] for MPLS. The network topology and settings used in the following simulations are the same for all simulated cases. The propagation delay between two nodes is set to 1ms and the bandwidth is set to 100Mbps. Each simulated model is setup to use end-to-end recovery, so the models can recover from a single link break anywhere on the working path. The models can also recover from node failures anywhere on the working path, apart from the ingress or egress LSR. The hello mechanism is used for failure detection. The mechanism is activated for all nodes that have a RSVP-TE agent attached and start time it set to 0.01s. The hello interval is set to 5ms and the multiplier set for the failure detection interval is set to 3.5ms. So a failure check will be performed in 15ms intervals. For each simulation it is wrote data for how many packets that are dropped when the link breaks. As the settings for the hello mechanism are set to the same values for all simulations, the packet dropped during the failure detection interval is the same in all simulations.

The results are presented below. There are used 3 criteria of comparison for the network recovery models:

1. **The packet loss during the recovery operations-** The fig.1 shows the number of dropped packages for each model. It is not presented, but the number of dropped packets decreases when the failure occurs closer to the egress LSR, because the backup path to setup becomes shorter.

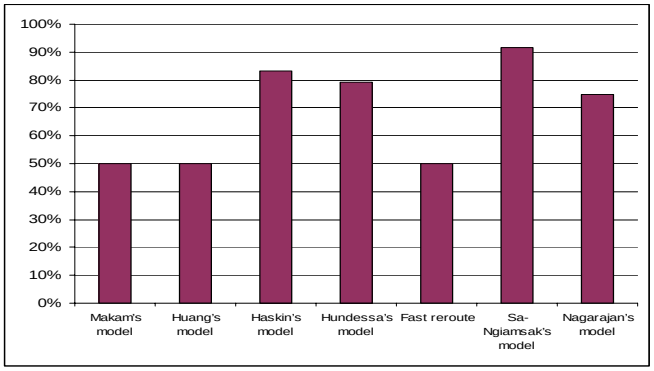


Fig. 1. Dropped packets

2. **The service disruption time** - The fig.2 shows the service disruption time, measured from the last packet that was sent over the link before it breaks is received by egress node, until the first package that is using the backup path is received by this node.

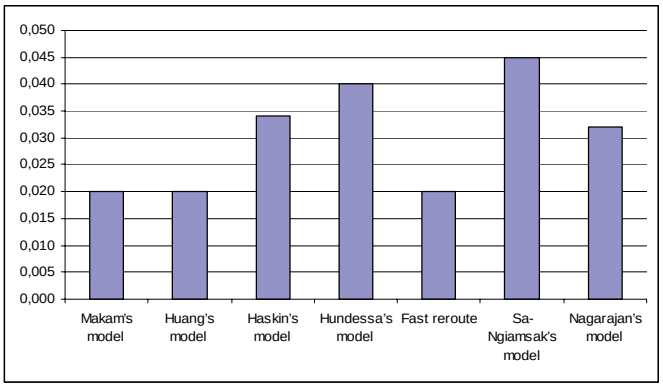


Fig. 2. Service Disruption Time

The service disruption time depends just like the number of dropped packages on the time for failure detection, failure notification, recovery path calculations and recovery path setup. This means that the rerouting mechanism will have higher service disruption time then protection switching, because with rerouting time is used for path calculations and path setup and this is not needed in protection switching.

3. **The number of pre-reserved resources used for the recovery operation –** Observe that the number of resources reserved depends on the topology of the network. Both Makam's and Haskin's model depends on a global recovery path, Haskins model will always use more resources then Makam's because it needs the reverse backup path in addition to this global recovery path. For most topologies fast

reroute will use more resources then Haskin's model, but this example is used to show that when the topology is right, fast reroute will use the same amount of backup resources as Haskins model. Both the best effort and rerouting model setup the backup path on demand after the failure has occurred, and therefore no backup resources are reserved before the failure in those models. The fig.3 presents these results.

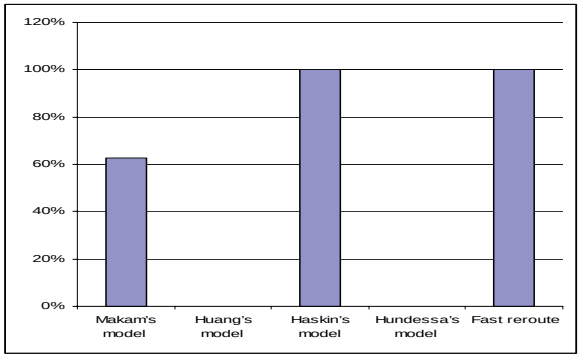


Fig. 3. Pre-reserved resources

When the multiple failure is occurred in wireless network Nagarajan's model and Sa-Ngiamsak's model are used and they recover LSP for different time. It is presented in the fig.4.

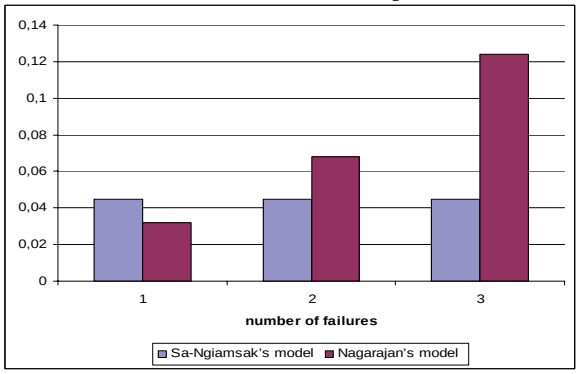


Fig. 4. Rerouting response time /ms/

IV. CONCLUSION

If recovery by re-routing is used the recovery time depends on the time to discover the fault, the time to notify the PSL of the failure, the time to calculate a recovery path (if it is not pre-calculated) and the time for a new recovery path to be set up. This can be slow and can take up to several seconds which is unacceptable for many real time applications.

If recovery by protection switching is used, the recovery time can be decreased because recovery path calculations are not needed. When the recovery path is pre-established, there is no need to signal the recovery path and recovery time, and then only depends on fault detection time and the time for the FIS to travel to the PSL.

Results indicate that the flexible signalling protocol for LSP in mobile wireless networks can effectively and efficiently handle rerouting in dynamic networks with a low protocol signalling overhead as compared to contemporary MPLS rerouting protocols. This would enable the MPLS based IP-QoS support mechanisms to extend to dynamic network topologies.

For further work it is planed to simulate the difference between the shortest paths after failure compared to the paths set up by each rerouting technique, then measure the time between occurrence of failure and instance of traffic resumption and observe the response time for each technique and finally measure and compare the total rerouting overhead for each technique. Based on this research, new model of recovery will be created, which evaluate these existing rerouting schemes via simulations. It is thinking about a low protocol overhead compared to the existing rerouting schemes and with the low response time when traffic travel on recovery LSPs.

ACKNOWLEDGEMENT

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

REFERENCES

[1] S. Blake, D. Black, etc. "An Architecture for Differentiated Services ", http://www.ietf.org/rfc/rfc2475.txt

[2] J.Martin,O. Petersson "MPLS Based Recovery Mechanisms", http://folk.uio.no/johanmp/MPLS%20Based%20Recovery%20Mechanisms.pdf

[3] V. Sharma, Ed. Metanoia, Inc. ,etc. "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", http://www.faqs.org/rfcs/rfc3469.html

[4] J. Ash , M. Girish ,etc. "Applicability Statement for CR-LDP", http://www.ietf.org/rfc/rfc3213.txt

[5] D. Awduche , L. Berger , etc. "RSVP-TE: Extensions to RSVP for LSP Tunnels", http://www.rfc-editor.org/rfc/rfc3209.txt

[6] G. Swallow "MPLS Advantages for traffic engineering", IEEE Communications Magazine December 1999

[7] http://nsnam.isi.edu/nsnam/index.php/User_Information

[8] S. Makam, etc. "Protection/ Restoration of MPLS Networks", http://zinfandel.levkowetz.com/html/draft-makam-mpls-protection-00

[9] C.Huang, V.Sharma, K.Owens, S.Makam "Building Reliable MPLS Network Using a Path Protection Mechanism"IEEE Communicaion Magazine March 2002

[10] D. Haskin, R. Krishnan "A method for setting an Alternative Label Switched Path to Handle Fast Reroute" , http://tools.ietf.org/html/draft-haskin-mpls-fast-reroute-01

[11] L.Hundessa, J.Pascual "Fast rerouting mechanism for a protected label switched path" Proceedings of the IEEE International Conference on computer Communication 01, October 2001

[12] P. Pan, G. Swallow, A.Atlas,"Fast Reroute Extensions to RSVP-TE for LSP Tunnels", http://www.rfc-editor.org/rfc/rfc4090.txt

[13] Sa-Ngiamsak, etc. "A Recovery Sheme for Qos Guaranteed Mobile IP Over MPLS Network" Wireless Pervasive Computing, 2006 1st International Symposiumon 16-18 Jan. 2006, http://ieeexplore.ieee.org/iel5/10746/33870/01613583.pdf

[14] Ramprasad Nagarajan, Eylem Ekici1 "An efficient and flexible MPLS signaling framework for mobile networks" Wireless Networks, Volume 14, Number 6 / December, 2008, http://www.springerlink.com/content/772114375010t173/?p=337b181cd2634c2ebccb7c663cc7d40e&pi=8

An Application for Geospatial Visualization of Optical Networks

Boban Joksimoski[1], Tomislav Lazarevski[2], Dejan Konevski[3], Ilija Jolevski[4], Dragan Mihajlov[5]

***Abstract* – Optical based communicational networks have contributed for a major boost in human communications. Increasing demands have led to complex network bindings that are difficult to maintain. Here we present a system for storing, maintaining and visualization of optical based communication networks.**

***Keywords* – telecommunication networks, infrastructure GIS, spatial data infrastructure, topology visualization.**

I. INTRODUCTION

Modern society is based on using different types of resources and distributing them to the users that need them. For such purposes, complex infrastructural objects have been built and connected in some manner. In modern age, information and data transfer have become a necessity. For transferring data, communication systems and networks (e.g. cable television service, wireless networks and other types of communication networks) have been developed. Usually such infrastructures have specific network topology and have tendency to change and grow. Maintaining such complex networks takes a lot of time, effort and money, especially if proper organization is missing.

For such uses, network managers make high benefits from using an application for storing and managing the huge quantity of data or each network. Additionally the use of geographic information system has proven to be of an imminent value. A whole section from geospatial computing (Spatial Data Infrastructures) is used for such purposes [1].

[1]*Boban Joksimoski* is student at the Faculty of Electrical Engineering and Information Technologies. Rugjer Boshkovik bb, PO Box 574, 1000 Skopje, Macedonia, E-mail: joksim@feit.ukim.edu.mk

[2]*Tomislav Lazarevski* is a student at the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: tomce5@yahoo.com

[3]*Dejan Koneski* is a student at the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: dkonevski@yahoo.com

[4]*Ilija Jolevski* is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: ilija.jolevski@uklo.edu.mk

[5]*Dragan Mihajlov* is with the Faculty of Electrical Engineering and Information Technologies. Rugjer Boshkovik bb, PO Box 574, 1000 Skopje, Macedonia, E-mail: dragan@feit.ukim.edu.mk

II. ORGANIZATIONAL PROBLEMS AND COMPLEXITY IN MODERN COMMUNICATION NETWORKS

Basically, various types of communication networks can be viewed as "*a set of equipment and facilities that provide a service: the transfer of information between users located at various geographical points*" [2]. Modern communication relies on advanced equipment (routers, switches) responsible for correct data transfer and the communication media (either wired or wireless).

Entities that are responsible for developing and maintaining communication networks constantly face a challenge to satisfy user demands. Users want to use new technologies like live streaming, HDTV, video sharing, software as a service (SaaS), Voice over IP (VoIP), on-line storage and many other services, to make their jobs easier and their lives better. All of the previously mentioned technologies, for optimal work, demand more bandwidth than projected 10 years ago. Also, adding new users requires exact knowledge of the deployed infrastructure. Thus, communication providers face more pressure to scale and expand their network, in order to keep up with the demands of the customers.

Today, such challenges are mainly upon providers of cable television, public switched telephone networks (PSTNs), and internet service providers (ISPs).

Our focus is to model a system that would help communication providers, providing them with the necessary information for effective work. Our main focus is on optical based communication networks, although applying this solution to other wired communications wouldn't be a problem.

Backbone networks are the most important parts of the network. Usually they are made of optical connections, mainly because of the advantage of having more throughput, better reliability and less attenuation than standard copper or satellite communications. As all modern networks, they consist of active equipment (i.e. routers, fiber optic media converters) and passive equipment (i.e. optical splitters, patchcords, and patch panels), interconnected with optical fibers, serial or Fast Ethernet links.

A simplified diagram of such type of network (an ISP) is given in Fig. 1. ISPs have multiple connection points that are commonly connected with fiber optic links. Connection points serve either as an end office or as an interconnection point between two or more connection points. An ISP has at least

one outer connection to the backbone network or another ISP. Other communication providers can have similar structure.

To ease network maintenance, systems like the one presented here, store and display important network properties. The scope of these systems is mainly about managing common problems with networks like querying the resources available at a connection point, tracing active and inactive lines, visualizing the connections, measuring performances and suggesting new deployment sites and ways of their interconnection.

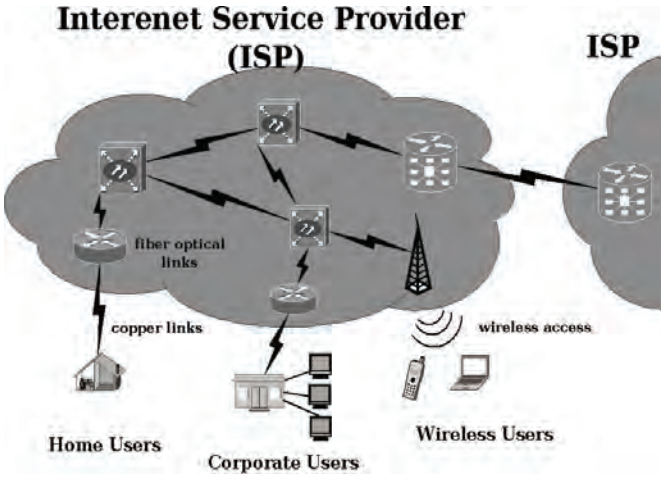


Fig. 1: Simplified diagram that shows the main parts of an ISP internal network

III. Requirements and similar solutions.

As with any software development model, the basic guidelines consist of preparing the user requirements. Network engineers have been consulted for better understanding of the problems they face and possible solutions that can be implemented to resolve the problems. Here is a list of the most important features that are required to be covered:

Managing the available resources – it is required to track used and spare resources in the network. Every cable or optical fiber must be mapped and stored in the database with its according properties and spatial parameters. The ports and other resources are also mapped, so that at any time the available and used ports can be seen. Brief information of the equipment configuration can also be stored.

- **Signal tracing** - another important part is the feature to trace the signal from one point to another and measure the degradation of the signal. Special equipment is available for such purpose, so the most obvious solution is to implement a log of the performance for each link. Integration with the underlying software is a part for future development.
- **Scalability and Reconfiguration** - because the network needs to grow, it is possible that network reconfigurations will occur, introducing new connections, cutting and splicing old connections, introducing new connection points, etc. These requirements lay down serious designer questions, mostly regarding the database scalability and maintainability.
- **Maintenance and Security** – most of the personnel that use the software are the on-site network workers. Network engineers are the other users of the application. There should be a distinction in permissions, clearance and features for each category of people involved. Also, the information stored is highly confidential and security is of primary concern.
- **Topology mapping and visualization** – in order to achieve better usability, it is required to include visualization of the network topology. Accordingly, the mapping should have a graph like structure, where connection offices and connection points are represented as vertices and the links represent the edges. Also, a feature that is required is visualisation of the connection types, the slice types and visualisation of the port availability (e.g. which ports are free, to which port the specific fiber is connected)
- **Independent** – the main principle is to make the application "manufacturer independent", but to provide common plug-in interfaces for most operations and compatibility with widespread standards.

Similar applications have been around for quite some time, and most of them are developed by telecommunication equipment manufacturers, and have features specific to the hardware (e.g. Nortel's Network Management Software, JDSU's ONMS & OFM-500, GRIS' Geoinformation System of Fiber-Optical Network Engineering).

IV. Our Idea and Implementation

As mentioned, there are already developed and workable solutions that can achieve this task. If we discuss about their disadvantages, two major problems appear. First they are difficult to obtain for smaller communication providers, mainly because of the high price and they provide little or none customization for different modules. We believe that these two features should be available to communication providers so they can apply the software to their structure.

The whole application is built for web usage. The implemented application logic allows adding new resources and modifying them, concretely adding new connection points and administration centres, adding new equipment (e.g. patch panels, switches, routers, media convertors), adding new optical fibers, cutting and splicing newly created pig tails and many more. Optical links deployment and signal tracing logs are implemented, including tracing loose connections, storing signal attenuations and pinpointing broken fibers. Security is integrated as a part of the used framework and it is further implemented for granularity. Network engineers have full access to all the equipment and admin centres, assign tasks, while network workers obtain and execute tasks. Algorithms are implemented for connection point placement and calculating various measure types for the used resources [5], to ease the engineering tasks.

Known and proven technologies are being used as building blocks of our system. The whole storage engine is based on

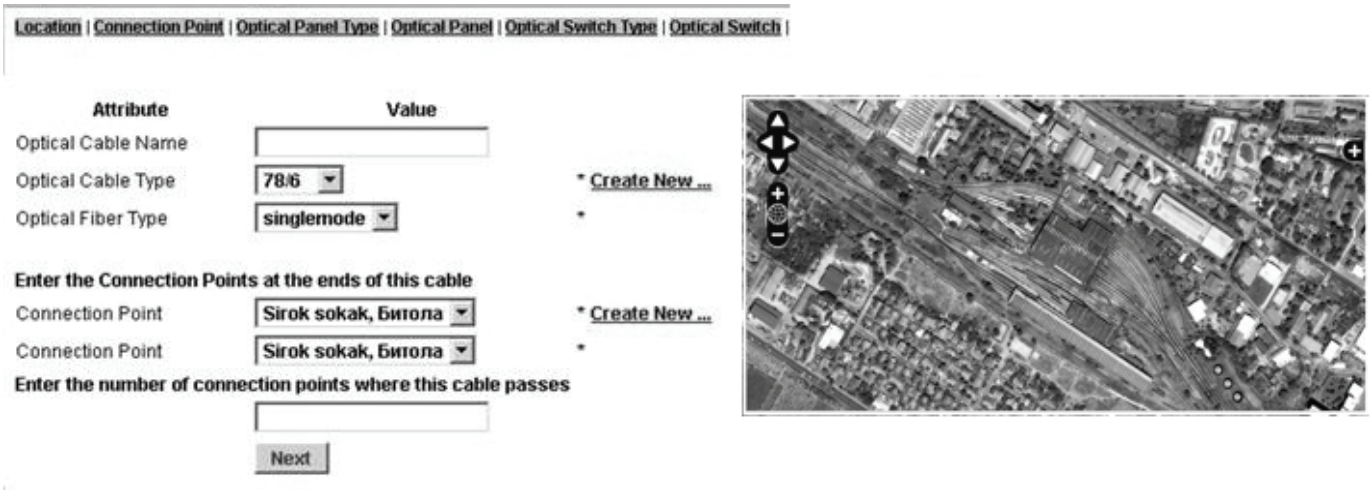


Fig 3: Early development of the web mapping

PostgreSQL and its PostGIS extension for managing spatial data. UMN Mapserver is used as support for various Open Geospatial Consortium (OGC) standards. For web interaction, Python is used for web programming (using the Django web framework) and OpenLayers is used for map manipulation (Fig. 2). More information on various platforms for GIS development can be found in [3].

The PostGIS extension allows usage of spatial database integrated queries for calculating geometric problems of proximity, adjacency and containment of spatial data. Recommendations for modeling and storing spatial data are implemented [4], according to the best practices of other spatial data infrastructure systems.

Implementation of the mapping is done via well known libraries for geographical data manipulation, and the use of Python as a language for development is because of its flexibility and good integration and usage of C and C++ libraries.

The implementation of the application is partial, with other modules besides basic functionality in plan.

The web mapping feature (at the moment of writing) is still in the development phase, but so far allows mapping of the basic resources (optical cables and connection point). Various ideas and guidelines are being implemented for the user interface design in accordance to various recommendations and experiences [6][7]. A sample screenshot of the application is given on Fig. 3.

Fig. 2: Overview of the architecture, main components and typical usage of the application

V. Evaluation and Guidelines for Further Development

As for further development, there are a lot of ideas that can be implemented. Growing networks demand implementing algorithms for suggestions for connection point placement and topology estimation, estimation of network load and prevention of network failures. Further refinement of the methods and parameters for the decision support module is planned.

Implementation issues as a typical web application are limiting user interaction, and it is considered leveraging the user interface to a RIA application. Thus, it shall include advanced AJAX and JavaScript functionalities. Other efforts of visualizing are planned for boosting the user experience. The applications can also be used for mapping non-optical communication networks or mixed optical and electrical networks, mapping wireless networks, natural resource mappings, etc. Adding support for GPS tracking is also planned.

VI. Conclusion

In this paper we introduce an application for maintaining optical networks on a regional SDI level. It faces the challenges of information storing and retrieval in a specific domain. The ideas implemented are not revolutionary, but the approach is. Open source and free technologies are used, with no need for expensive licenses from the major software vendors. Geospatial mapping is of key interest for better productivity and management, because it provides native presentation and excellent overview of the data.

References

[1] Chan T. O. and Williamson I. P. 1999b ‘Spatial data infrastructure management: lessons from corporate GIS development’. Proceedings of AURISA '99, November 1999, Blue Mountains, NSW, AURISA 99: CD-ROM

[2] Leon-Garcia, A., Widjaja, I.: Communication Networks, Fundamental Concepts and Key Architectures, McGraw-Hill, 2003
[3] Rinaudo, F., Agosto, E., Ardissone, P.: Gis and Web-Gis, Commercial and Open Source Platforms: General Rules for Cultural Heritage Documentation. In. XXI International CIPA Symposium, Athens (2007)
[4] Friis-Christensen, A., Tryfona, N., Jensen, C. S.: Requirements and Research Issues in Geographic Data Modeling. In: 9th ACM international symposium on Advances in geographic information systems, pp. 2-8. ACM New York, NY, USA (2001)
[5] Wheeler D. C., O'Kelly M. E. Network Topology and City Accessibility of the Commercial Internet,
[6] ESPRIT Programme: Guidelines for Best Practice in User Interface for GIS. The European Commission (2000)
[7] McCurley, K. S.: Geospatial mapping and navigation of the web. In: 10th international conference on World Wide Web, pp. 221- 229, ACM New York, NY, USA (2001)

Analysis of the Physical Layer in IEEE 802.16(e) Standard

Grigor Y. Mihaylov[1], Teodor B. Iliev[2] and Georgi V. Hristov[3]

***Abstract* – WiMAX (IEEE 802.16e) is a broadband wireless solution that enables convergence of mobile and fixed broadband networks through a common wide area broadband radio access technology and flexible network architecture. In this paper we explains the steps in channel coding stage and make comparative analysis of the forward error correction codes implemented in IEEE 802.16e standard. The main function of the channel coding is to prevent and to correct the transmission errors of wireless systems and they must have a very good performance in order to maintain high data rates.**

***Keywords* –WiMAX, channel coding, turbo codes**

I. INTRODUCTION

Worldwide Interoperability for Microwave Access (*WiMAX*) is a standards-based wireless technology for providing high-speed, last-mile broadband connectivity to homes and businesses and for mobile wireless networks. WiMAX is similar to Wi-Fi but offers larger bandwidth, stronger encryption, and improved performance over longer distances by connecting between receiving stations that are not in the line of sight. WiMAX uses Orthogonal Frequency Division Modulation (*OFDM*) technology, which has a lower power consumption rate. WiMAX can be used for a number of applications, including last-mile broadband connections, hotspots and cellular backhaul, and high-speed enterprise connectivity for business. It supports broadband services such as VoIP or video [1], [2].

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

[1]Grigor Y. Mihaylov is with the Department of Communication Systems and Technologies, 8 Studentska Str., 7017 Ruse, Bulgaria, E-mail: gmihaylov@uni-ruse.bg

[2]Teodor B. Iliev is with the Department of Communication Systems and Technologies, 8 Studentska Str., 7017 Ruse, Bulgaria, E-mail: tiliev@ecs.uni-ruse.bg

[3]Georgi V. Hristov is with the Department of Communication Systems and Technologies, 8 Studentska Str., 7017 Ruse, Bulgaria, E-mail: ghristov@uni-ruse.bg

II. WIMAX PHYSICAL LAYER

The WiMAX physical layer is based on orthogonal frequency division multiplexing. OFDM is the transmission scheme of choice to enable high-speed data, video, and multimedia communications and is used by a variety of commercial broadband systems, including DSL, Wi-Fi, Digital Video Broadcast-Handheld (*DVB-H*), and MediaFLO, besides WiMAX. OFDM is an elegant and efficient scheme for high data rate transmission in a non-line-of-sight or multipath radio environment.

Apart from the usual functions such as randomization, forward error correction (*FEC*), interleaving, and mapping to QPSK and QAM symbols, the standard also specifies optional multiple antenna techniques. This includes space time coding (*STC*), beamforming using adaptive antennas schemes, and multiple input multiple output (*MIMO*) techniques which achieve higher data rates. The OFDM modulation/demodulation is usually implemented by performing fast fourier transform (*FFT*) and inverse FFT on the data signal. Although not specified in the standards, other advanced signal processing techniques such as crest factor reduction (*CFR*) and digital predistortion (*DPD*) are also usually implemented in the forward path, to improve the efficiency of the power amplifiers used in the base stations. The uplink receive processing functions include time, frequency and power synchronization (ranging), and frequency domain equalization, along with rest of the decoding/demodulation operations necessary to recover the transmitted signal [3].

One of the ambitious design goals of future wireless systems, including 4G, IEEE 802.11n/802.16 standards, is to reliably provide very high data rate transmission in hostile environments: hundreds of Mb/s or more for downlink transmission with a low frame error rate (*FER*), typically less than 5.10^{-4}. Therefore, efficient equalizers and decoders are required in order to mitigate inter-symbol interference (*ISI*) and residual interference, respectively. OFDM modulation is particularly suited for transmissions over multipath channels. An OFDM system transforms the frequency selective channel into a set of narrowband Gaussian orthogonal subchannels. Since the frequency selectivity implies that some subbands are strongly weakened, a powerful receiver is needed. Several methods such as power allocation or channel coding have been used [4], [5].

Like all other standards, only the components of the transmitter are specified; the components of the receiver are left up to the equipment manufacturer to implement.

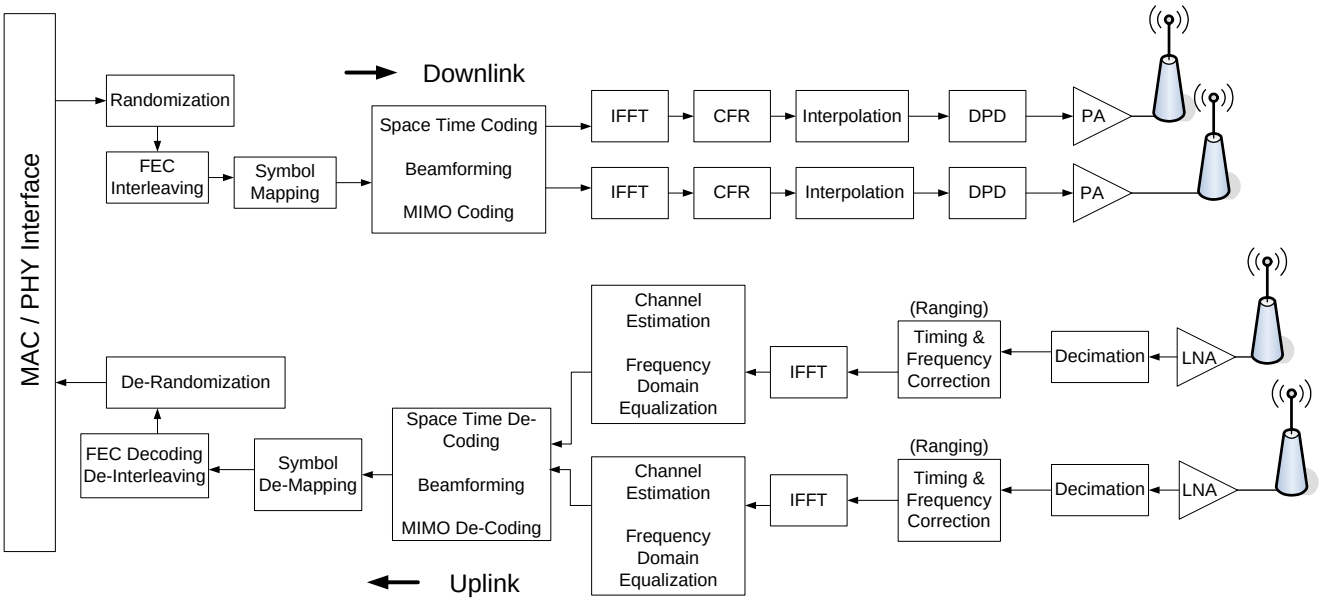


Fig.1 PHY Layer functions in a typical WiMAX base station

III. Channel Coding

In IEEE 802.16e-2005, the channel coding stage consists of the following steps: data randomization, channel coding, rate matching, HARQ, if used, and interleaving. Data randomization is performed in the uplink and the downlink, using the output of a maximum-length shift-register sequence that is initialized at the beginning of every FEC block. This shift-register sequence is modulo 2, added with the data sequence to create the randomized data. The purpose of the randomization stage is to provide layer 1 encryption and to prevent a rogue receiver from decoding the data. When HARQ is used, the initial seed of the shift-register sequence for each HARQ transmission is kept constant in order to enable joint decoding of the same FEC block over multiple transmissions.

Channel coding is performed on each FEC block, which consists of an integer number of subchannels. A subchannel is the basic unit of resource allocation in the PHY layer and comprises several data and pilot subcarriers. The exact number of data and pilot subcarriers in a subchannel depends on the subcarrier permutation scheme. The maximum number of subchannels in an FEC block is dependent on the channel coding scheme and the modulation constellation. If the number of subchannels required for the FEC block is larger than this maximum limit, the block is first segmented into multiple FEC subblocks. These subblocks are encoded and rate matched separately and then concatenated sequentially, to form a single coded data block. Code block segmentation is performed for larger FEC blocks in order to prevent excessive complexity and memory requirement of the decoding algorithm at the receiver. [3]

A. Convolutional Coding

The mandatory channel coding scheme in IEEE 802.16e is based on binary nonrecursive convolutional coding (CC). The convolutional encoder uses a constituent encoder with a constraint length 7 and a native code rate 1/2. The output of the data randomizer is encoded using this constituent encoder. In order to initialize the encoder to the 0 state, each FEC block is padded with a byte of 0x00 at the end in the OFDM mode. In the OFDMA mode, tailbiting is used to initialize the encoder, as shown in Fig. 2. The 6 bits from the end of the data block are appended to the beginning, to be used as flush bits. These appended bits flush out the bits left in the encoder by the previous FEC block. The first 12 parity bits that are generated by the convolutional encoder which depend on the 6 bits left in the encoder by the previous FEC block are discarded. Tailbiting is slightly more bandwidth efficient than using flush bits since the FEC blocks are not padded unneccessarily. However, tailbiting requires a more complex decoding algorithm, since the starting and finishing states of the decoder are no longer known. In order to achieve code rates higher than 1/2, the output of the encoder is punctured, using the puncturing pattern.

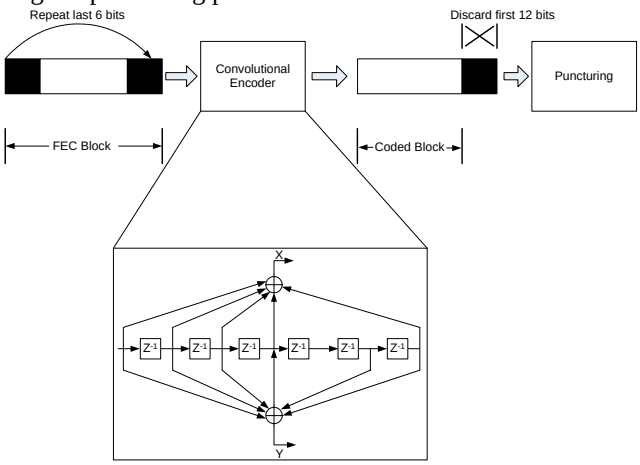


Fig.2 Convolutional encoder and tailbiting in IEEE 802.16e

B. Turbo Codes

WiMAX uses duobinary turbo codes with a constituent recursive encoder of constraint length 4. In duo binary turbo codes two consecutive bits from the uncoded bit sequence are sent to the encoder simultaneously. Unlike the binary turbo encoder used in HSDPA and 1xEV-DO, which has a single generating polynomial for one party bit, the duobinary convolution encoder has two generating polynomials, $1+D^2+D^3$ and $1+D^3$ for two parity bits. Since two consecutive bits are used as simultaneous inputs, this encoder has four possible state transitions compared to two possible state transitions for a binary turbo encoder. [6]

Duobinary turbo codes are a special case of nonbinary turbo codes, which have many advantages over conventional binary turbo codes:

✓***Better convergence***: The better convergence of the bidimensional iterative process is explained by a lower density of the erroneous paths in each dimension, reducing the correlation effects between the component decoders;

✓***Larger minimum distances***: The nonbinary nature of the code adds one more degree of freedom in the design of permutations (interleaver)-intrasymbol permutation-which results in a larger minimum distance between codewords;

✓***Less sensitivity to puncturing patterns***: In order to achieve code rates higher than 1/3 less redundancy, bits need to be punctured for nonbinary turbo codes, thus resulting in better performance of punctured codes;

✓***Robustness of the decoder***: The performance gap between the optimal MAP decoder and simplified suboptimal decoders, such as log-MAP and the soft input soft output (SOVA) algorithm, is much less in the case of duobinary turbo codes than in binary turbo codes.

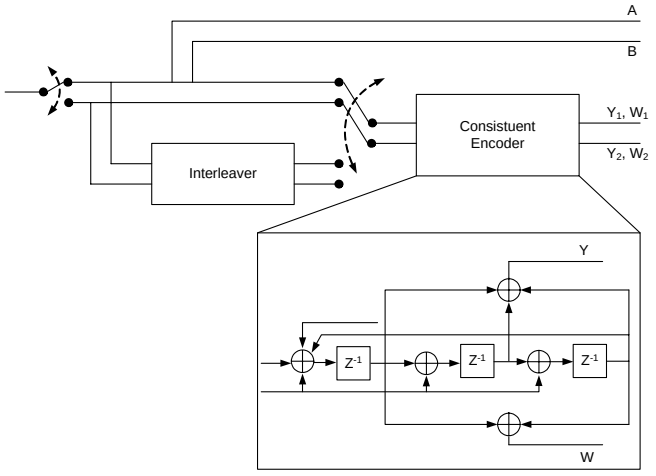


Fig.3 Turbo Encoder in IEEE 802.16e

C. Block Turbo Codes and LDPC Codes

Other channel coding schemes, such as block turbo codes and LDPC codes, have been defined in WIMAX as optional channel coding schemes. The block turbo codes consist of two binary extended Hamming codes that are applied on natural and interleaved information bit sequences, respectively. A BTC codeword is a simple product code, usually formed by a serial concatenation of two block encoders separated by a block interleaver. Let (n_i, k_i, δ_i), $i=\{1, 2\}$, be the length, dimension, and minimum distance of the constituent codes, respectively. Then the parameters for the product code are $n_p=n_1 \bullet n_2$, $k_p=k_1 \bullet k_2$, $\delta_p=\delta_1 \bullet \delta_2$. The IEEE 802.16 standard offers several different options for encoding, but the longest constituent block code is a (64, 57, 4) extended Hamming code.

Low-density parity-check (*LDPC*) codes have recently attracted tremendous research interest because of their excellent error correction performance. LDPC codes have been adopted in many standards such as DVB-S2, 10GBase-T, 802.16e (*WiMAX*) and 802.11n. However, designing an LDPC code that has superior performance and can be mapped efficiently into hardware, is still a challenge. [7]

LDPC codes have a large degree of freedom in both code and decoder design. The datapath of the decoder is generally simple, and the operations can be easily parallelized. However, because of the interconnection complexity, the fully parallel LDPC decoder is huge for large block sizes. The partial parallel decoder which makes use of small block matrices with ordered structure is highly preferred. Several LDPC codes with ordered structures based on algebraic constructions have been proposed. These codes make use of algebraic properties that achieve good bit error rate (*BER*) performance. [8], [9]

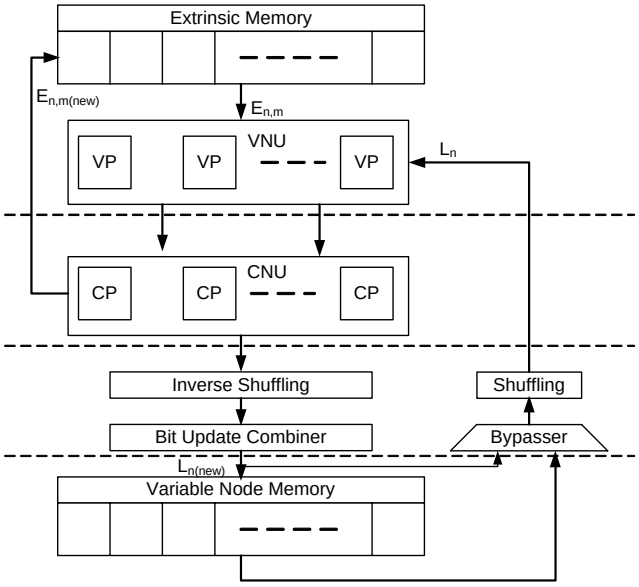


Fig.4 LDPC decoder architecture

The conventional TPMP SPA (the standard two-phase message passing Sum-Product algorithm) is commonly regarded as the standard LDPC decoding algorithm and is generally implemented in log domain. The check-to-variable messages R_{cv} are computed as Eqs. (1) and (2).

$$R_{cv} = S_c \times sign(L_{cv}) \times \Psi\{\sum_{n \in N(c)} \Psi(L_{cn}) - \Psi(L_{cv})\} \quad (1)$$

$$S_c = \Pi_{n \in N(c)} sign(L_{cn}) \quad (2)$$

where $N(c)$ denotes the set of variable nodes connected to the check node c, and $\Psi(x) = -\log(\tanh(|x|/2))$ is a nonlinear function. The variable-to-check message L_{cv} is computed as Eqs. (3) and (4).

$$L_{cv} = L_v - R_{cv} \tag{3}$$

$$L_v = \sum_{m \in M(v)} R_{mv} + I_v \tag{4}$$

where L_v is the LLR message of variable node v and $M(v)$ denotes the set of check nodes connected to the variable node v. The intrinsic message corresponding to variable node v is $I_v = 2r_v / \sigma^2$, for binary input (mapping 0 to +1 and 1 to -1) and AWGN channel, where r_v and σ are the received soft value and the standard deviation of noise, respectively. The sign of L_v is taken as the estimated codeword bit c_v (mapping +1 to 0 and -1 to 1). The check-sum P_c of parity equation corresponding to check node c is computed by Eq. (5). [10]

$$P_c = \oplus_{v \in N(c)} c_v \tag{5}$$

where $\oplus$ represents binary addition. If P_c = 0 for any check node c, a valid code is found and the decoding process can be terminated.

IV. Simulation and Results

We have simulated and compared LDPC codes and convolutional turbo codes intended for the WiMAX (*IEEE 802.16e*) forward error correcting schemes. For the CTC, iterative decoding was stopped after 10 iterations. Concerning the LDPC decoder, the maximum number of iterations of belief propagation decoding was limited to 100. The simulations were carried out for different code rates, lengths and modulation schemes in additive white Gaussian noise (*AWGN*) channel. Simulations were run to determine the performance of CTC and LDPC in AWGN channel with BPSK modulation [6]. For each simulation, a curve showing the bit-error rate (*BER*) versus E_b/N_0 was computed. The E_b stands for energy per bit and the N_0 stands for the noise power spectral density ratio.

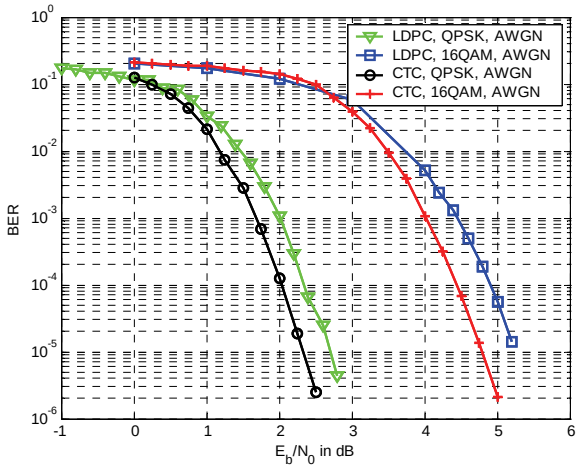


Fig.5 Comparison between LDPC and CTC with code rate R=1/2

On Fig. 5 we shows the comparison between LDPC codes and CTC codes with code rate R=1/2, two modulation schemes (QPSK and 16QAM) and N=576 bits.

V. Conclusion

The contribution of this paper has been a study into WiMAX forward error correcting codes. It presents a validation and a discussion of these types of codes. Secondly, this paper presents an implementation of convolutional turbo codes and LDPC codes developed in Matlab. The performance gain using advanced coding techniques like CTC and LDPC is quite small for rate 1/2 codes. One reason for this is that the standard only provides short to moderate code lengths (N≤2304) which is the most crucial parameter for this class of codes. The performance of CTC and LDPC is about the same and by changing some decoding parameters the small advantage of one of them can be interchanged. Nevertheless, LDPC decoding is less complex than CTC decoding.

Acknowledgement

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

References

[1] G. S. V. Radha, K. Rao and G. Radhamani, "*WiMAX A Wireless Technology Revolution*", Auerbach Publications, 2008

[2] Jeffrey G. Andrews, A. Ghosh and R. Muhamed, "*Fundamentals of WiMAX*", Prentice Hall, 2007

[3] S. Ahson and M. Ilyas "*WiMAX: Technologies, Performance Analysis, and QoS*", CRC Press, 2008

[4] M. Ma, "*Current Technology Developments of WiMax Systems"*, Springer, 2009

[5] Y. Zhang, H.-H. Chen, "*Mobile WiMAX*", Auerbach Publications, 2008

[6] T. Iliev, G. Hristov, P. Zahariev and M. Iliev, Performance of the Duo-Binary Turbo Codes in WiMAX Systems, Novel Algorithms and Techniques in Telecommunications and Networking, Springer, 2010, pp. 161 – 165

[7] L.-H. Lin and K.-A. Wen, "A Novel Application of LDPC-Based Decoder for WiMAX Dual-mode Inner Encoder", Proceedings of the 9th European Conference on Wireless Technology, Manchester, UK, 2006

[8] Y. Zhu and C. Chakrabarti, "Aggregated CIRCULANT Matrix Based LDPC Codes" white paper

[9] Z. Cui, L. Chen, and Z. Wang, "An Efficient Early Stopping Scheme for LDPC Decoding" white paper

[10] T. Iliev, G. Hristov, P. Zahariev and M. Iliev, Application and evaluation of the LDPC codes for the next generation communication systems, Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics, Springer, 2008, pp. 532 – 536

Analysis of Reed-Solomon Codes: Application to Digital Video Broadcasting Systems

Teodor B. Iliev[1] and Georgi V. Hristov[2]

***Abstract* – Accuracy of information in any communication system is very critical. Use of Forward Error Correction (FEC) to lower the probability of error and increase transmission distance has become widespread. Reed-Solomon is a block FEC, capable of correcting multiple errors, specifically focusing on burst errors, making it popular for storage devices, wireless and mobile communication units. In this paper we have presented, discussed and analyzed the Reed Solomon encoder and decoder structures implemented in the Digital Video Broadcasting systems.**

***Keywords* – DVB, Reed – Solomon codes, FEC**

I. INTRODUCTION

Error correcting codes (channel coding) are one of the solutions available to improve the digital communication quality. The purpose of channel coding is to introduce, in a controlled manner, some redundancy in the binary information sequence to overcome the effects of noise and interference encountered during the transmission through the channel. Reed-Solomon codes are block error correction codes with burst error-correcting capabilities that have found widespread use in storage devices and digital communication systems [1, 8]. In particular, concatenated coding employing an inner convolutional code combined with a Reed-Solomon outer code constitutes an attractive scheme that is commonly encountered in many applications, and in particular in the digital video broadcasting systems (DVB) [9, 10].

RS codes are Maximum Distance Separable (MDS), known to be the most powerful linear codes for their class, and have the ability to correct both errors and erasures, defined as errors with identified error locations.

For a typical channel, the addition of RS coding allows the system to operate within approximately 4 dB of the Shannon capacity. The resulting benefit translates into higher data rates, lower bit-error rates (BER), greater transmission distance, and greater immunity to interference effects [2].

II. PROPERTIES OF REED – SOLOMON ENCODER

Reed Solomon encoding is a block encoding scheme and partially is specified as an RS (n, k), where n refers to the output codeword length and k refers to the input word length, as shown in Fig. 1. The RS code is based on the Galois field GF (2^8), and therefore has a symbol size of 8 bits. The difference $n–k=2t$ is the number of parity symbols have been appended to make the encoded block. An RS decoder can correct up to $(n–k)/2$ or t symbols, i.e. any t symbols can be corrupted in any way and the original symbols can be recovered. The RS (255, 239) was chosen which processes a data block of 239 symbols and can correct up to 8 symbol errors by calculating 16 redundant correction symbols. As an MPEG-2 packet is 188 bytes long, the code was shortened, i.e. the first 51 information bytes were set to zero and not transmitted at all. In this way the RS (204, 188) is generated [4]. Thus the DVB code splits the message into blocks 188 symbols long. The parity symbols ($2t$=204 – 188=16) are then appended to produce the full 204 symbol long code. Up to t=16/2=8 symbol errors can be then corrected.

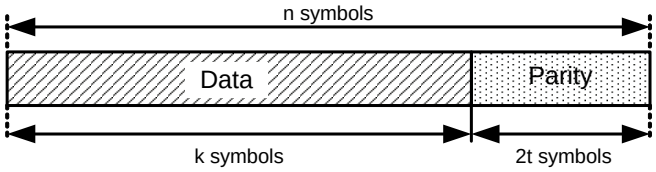


Fig.1 Code word of RS (n,k) code

The encoder forms a code word $x^{n-k}m(x) + r(x)$ by means of the following equation:

$$\frac{x^{n-k}m(x)}{g(x)} = q(x) + \frac{r(x)}{g(x)}, \qquad (1)$$

where the divisor, $g(x)$ is known as the generator polynomial. It is a polynomial of degree (n-k) and which is a factor of (x^n+1). To maximize the minimum distance between codes, the roots of this polynomial should all be consecutive. This is a direct consequence of the BCH bound, which states that the minimum distance is always larger than the number of consecutive factors of $g(x)$. The system used adapted a generator polynomial with roots from α^1 to α^{32}.

The term x^{n-k} is a constant power of x, which is simply a shift upwards n-k places of all the polynomial coefficients in $m(x)$. It happens as part of the shifting process in the architecture below. The remainder after the division $r(x)$ becomes the parity. By concatenating the parity symbols on to the end of the k message symbols, an n coefficient polynomial is created which is exactly divisible by $g(x)$.

Eq. (1) also implies that the generator polynomial is a factor of all possible code words.

The symbols in Reed Solomon coding are elements of a Galois Field (finite field). Encoding is achieved by appending the remainder of a Galois field polynomial division into the

[1]Teodor B. Iliev is with the Department of Communication Systems and Technologies, 8 Studentska Str., 7017 Ruse, Bulgaria, E-mail: tiliev@ecs.uni-ruse.bg

[2]Georgi V. Hristov is with the Department of Communication Systems and Technologies, 8 Studentska Str., 7017 Ruse, Bulgaria, E-mail: ghristov@uni-ruse.bg

message. This division is done by a Linear Feedback Shift Register (LFSR) implementation. The RS encoder with internal feedback connection corresponding to $g(x)$ is shown on Fig. 2.

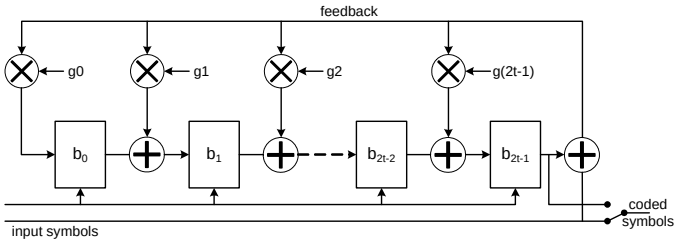


Fig.2 RS encoder with LFSR

The encoder shown on Fig. 2 is a 2t tap shift register, where each register is m bits wide. The multiplier coefficients $g0$ to $g(2t\text{-}1)$ are coefficients of the RS generator polynomial. The coefficients are fixed, which can be used to simplify the multipliers if required. The only hard bit is working out the coefficients, and for hardware implementations the values can often be hard coded.

At the beginning of a block all the registers are set to zero. From then on, at each clock cycle the symbol in each register is added to the product of the feedback symbol and the fixed coefficient for that tap, and passed on to the next register. The symbol in the last register becomes the feedback value on the next cycle. When all n input symbols have been read in, the parity symbols are sitting in the register, and it just remains to shift them out one by one.

III. Reed – Solomon Decoder

Decoding RS codes involves the extraction of two information entities from the received word. These are the set of Partial Syndromes, and the Error Locator Polynomial. From these two parameters, Error Locations and Error magnitudes are extracted, and are directly applied upon the received word to extract the original, corrected message [6].

A Generator Polynomial defines each RS code configuration. The specific configurations of the encoder and decoder are determined by this polynomial. In the decoder, the partial syndromes are computed by treating the received information as a polynomial and evaluating it at each of the roots of the Generator Polynomial. This operation is realized in hardware as a set of shift-register assemblies connected in parallel. The Partial Syndromes are computed simultaneously and are passed on afterwards to the Error Locator Polynomial unit. The Error Locator Polynomial is computed directly from the Partial Syndromes and a popular algorithm used to extract this is the Berlekamp-Massey Algorithm. The magnitudes of the corresponding errors are solved using the Forney Algorithm. An error-correcting block uses the computed error magnitudes to adjust the received codeword [5].

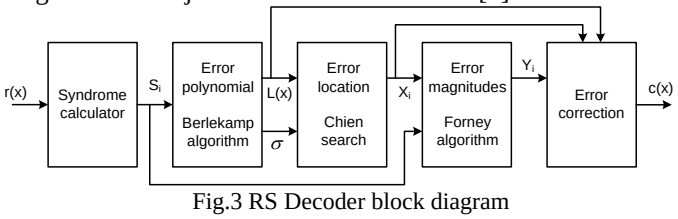


Fig.3 RS Decoder block diagram

The RS decoder consists of five major blocks as shown in Fig. 3.

The first step in decoding the received symbol is to determine the data syndrome. A codeword's syndrome $s(x)$ is the remainder of the division of the received word $r(x)$ by the generator polynomial, as implied by the following equation:

$$\frac{r(x)}{g(x)} = q(x) + \frac{s(x)}{g(x)}. \tag{2}$$

The received word, as seen in Eq. 3, can be decomposed into the code word $c(x)$ and the error component $e(x)$. Further manipulation of Eq. 2 leads to Eq. 5.

$$r(x) = c(x) + e(x) \tag{3}$$

$$\frac{c(x)+e(x)}{g(x)} = q(x) + \frac{s(x)}{g(x)} \tag{4}$$

Since all code words are divisible by the generator polynomial, only the error component will yield a remainder. Eq. 5 and 6 illustrates this:

$$\frac{c(x)}{g(x)} + \frac{e(x)}{g(x)} = q_c(x) + q_e(x) + \frac{s(x)}{g(x)} \tag{5}$$

$$\frac{e(x)}{g(x)} = q_e(x) + \frac{s(x)}{g(x)} \tag{6}$$

It can now be seen in Eq. 6 that the syndrome is independent of the message information and depends only on the error component.

In most systems, partial syndromes are computed instead of the syndrome, for reasons of simpler hardware implementation. In the computation of a partial syndrome, the divisor is no longer the entire generator polynomial, but only one of its factors, as seen in Eq. 7. There will be n-k partial syndromes for every received word, since the generator polynomial has n-k factors.

$$\frac{r(x)}{(x-a_k)} = q(x) + \frac{s_k}{(x-a_k)} \tag{7}$$

Note that the remainder s_k in Eq. 7 can be obtained by evaluating $r(x)$ at a_k, as:

$$s_k = r(a_k) \tag{8}$$

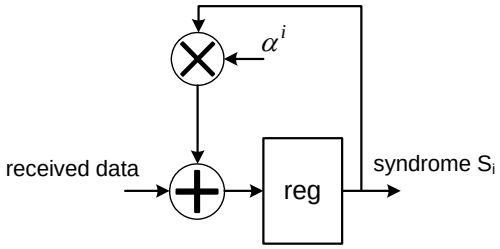


Fig.4 Partial Syndrome Calculator

Syndrome calculation can be done by an iterative process, such that the answer (2t syndrome symbols) is available as

soon as the last parity symbol has been read in. The circuit shown on Fig. 4 will generate the i^{th} syndrome, $2t$ of these will be needed for the full syndrome decoder. The syndromes depend only on the errors, not on the underlying encoded data.

A. Error Polynomial lambda – Berlecamp – Massey and Euclids algorithm

In Reed-Solomon decoding, two items of information are required to extract the original message: the error locations and the error magnitudes.

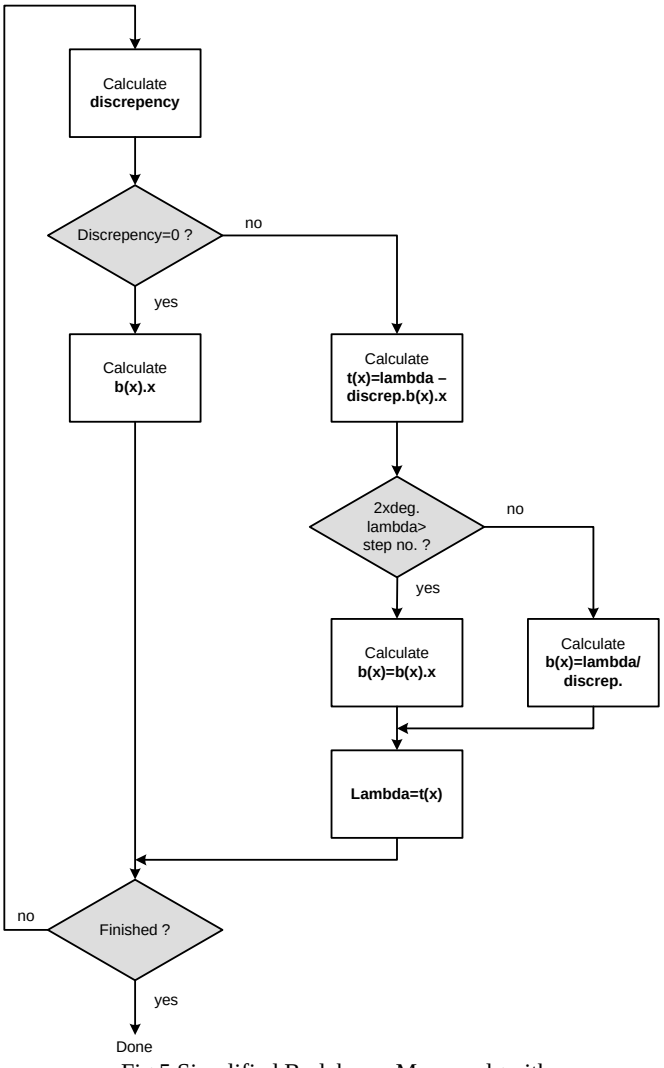


Fig.5 Simplified Berlekamp Massey algorithm

The Error Locator Polynomial is a polynomial whose roots directly define the error locations in the received word. This polynomial could be computed directly from the partial syndromes [7]. A number of methods exist to perform this computation, and the method that is most popular and most suited to this system is the Berlekamp-Massey Algorithm.

The Berlekamp-Massey algorithm is a shift-register synthesis algorithm [8] and the simplified block scheme of the algorithm is show on Fig.5. It takes as input the n-k partial syndromes and outputs the error locator polynomial $\sigma(x)$.

The algorithm aims to find an LFSR of minimal length such that the first (n-k) elements in the LFSR output sequence are the (n-k) syndromes. The taps of this shift register are the coefficients of the desired error locator polynomial, $\sigma(x)$ [8].

The algorithm iteratively solves the error locator polynomial by solving one equation after another and updating the error locator polynomial. If it turns out that it cannot solve the equation at some step, then it computes the error and weights it, increases the size of the error polynomial, and does another iteration. A maximum of $2t$ iterations are required. For n symbol errors, the algorithm gives a polynomial with n coefficients. At this point the decoder fails if there are more than t errors, and no corrections can be made. Doing so might actually introduce more errors than there were originally.

B. Error Polynomial Roots

After the Error Locator Polynomial has been computed, the roots of this polynomial have to be calculated in order to know the error locations. This is done by performing the Chien Search, which evaluates the Error Locator Polynomial at all elements of the GF(256) field. The algorithm checks if $\sigma(a^p)$ equals zero, p = 0, 1, 2 …., n, then a^p is a root of the polynomial, and α^{-p} is an error location, X_p.

Solving the key equation (Eq. 7) [7] determines the error evaluator or error magnitude polynomial, $\Omega(x)$:

$$S(x)\sigma(x) = \Omega(x) \bmod x^{n-k} \tag{9}$$

C. Error Magnitudes - Forney Algorithm

An efficient way of computing $\Omega(x)$ is to perform parallel computation of $\sigma(x)$ [9]. Solution to the key equation may be derived as follows:

$$\begin{aligned}\Omega(x) &= S(x)\sigma(x) \bmod x^{2t} = \\ &= \left(S_1 + S_2 x + \ldots + S_{2t} x^{2t-1}\right) \bullet \left(\sigma_1 + \sigma_2 x + \ldots + \sigma_t x^t\right) \bmod x^{2t} = \\ &= \Omega^{(0)} + \Omega^{(1)} x + \ldots + \Omega^{(t-1)} x^{t-1}\end{aligned} \tag{10}$$

$$\Omega^{(i)} = S_{i+1}\sigma_0 + S_i\sigma_1 + \ldots + S_1\sigma_i, \quad for\ i = 0,1,\ldots,t-1 \tag{11}$$

This leads to a direct computation of $\Omega(x)$, which requires fewer multiplications than the iterative algorithm [3].

The Forney Algorithm is used to compute for the error magnitudes, Y_i, corresponding to the respective error locations. Using the following equation:

$$Y_i = \frac{\Omega\left(X_i^{-1}\right)}{\sigma'\left(X_i^{-1}\right)}, \tag{12}$$

where X_i^{-1} indicates the root as computed from the Chien Search, and $\sigma'(x)$ the derivative of the error locator polynomial.

D. Error Correct

The error corrector block takes the received code and performs XOR-operation with the corresponding error

magnitudes computed at the respective error locations to attain the original message stream (Eq. 13).

$$c(X_i) = r(X_i) \oplus Y_i \tag{13}$$

IV. Simulation

We have presented the results of conducted simulation of Reed Solomon codes with following n and m parameters in according to DVB standard – RS (255, 239), RS (255, 235), RS (255, 223), RS (255, 205). The parameters setting of convolution code, rate $R=m/n$ selection in according to DVB standard – 1/2, 2/3, 3/4, 5/6, 7/8 and interleaver depth I settings – up to 20 [4].

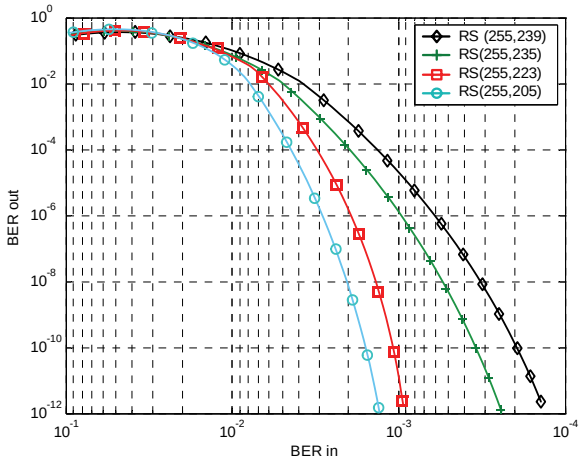


Fig. 6. Residual Bit – Error Rate value of various RS codes

Independently of whether the encoding and decoding is in frequency or time domain, the efficiency of the RS code is the same. The residual bit-error rate (BER) of various RS codes based on GF (2^8) is shown in Figure 6. Although the RS codes are symbol-oriented codes, the analysis of the efficiency takes bit errors into account. The efficiency of the code increases with an increase in the number of test symbols. At an input bit-error rate of 2.10^{-3} the residual bit error rate of the RS (255, 205) code is approx 1.10^{-10} – the coding gain is thus more than 10 to the power of 7 – whereas in the case of the RS (255, 239) code at the same input bit-error rate the output bit-error rate is 9.10^{-4} – the coding gain is only slightly greater than 0.5. For all DVB transmission standards a modified (shortened) RS (255, 239) code is used which makes it possible residual bit-error rate of approx 1.10^{-11} at an input bit--error rate of 2.10^{-4} while correcting up to 8 symbol errors per block.

The residual bit-error rate of convolutional codes of rate R is a function of E_b/N_0 (energy transmitted per bit divided by the noise-power density of the white Gaussian noise) and the parameter K describes the length of the code. The performance of the error correction increases with increased K. For the DVB standard a convolutional code of rate R=1/2 with a constraint length K=7 is used. With E_b/N_0 = 3.2 dB it is possible to achieve a bit-error rate of less than 2.10^{-4} at the output of the decoder, this ratio corresponds to the maximum of the bit-error rate at the input of the RS decoder, so finally a bit-error rate at the output of the RS decoder of less than 1.10^{-11} is obtained.

V. Conclusion

This paper has provided an overall view of the Reed Solomon codes implemented in digital television broadcast standards and we have conducted simulation with various RS codes. From the simulation is evidence that the efficiency of the codes increases with an increase in the number of test symbols.

Reliability of information is critical in any communication systems. Use of error-control codes reduce interference effects, and FECs in general, eliminate the need for retransmission of data streams. The theoretical performance of Reed-Solomon codes in bursty noise channels makes it a very good choice for FEC.

Acknowledgement

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

References

[1] T. Iliev, I. Lokshina, D. Radev, and G. Hristov, "Analysis and Evaluation of Reed–Solomon Codes in Digital Video Broadcasting Systems", – In: Proceedings of Seventh Annual Wireless Telecommunications Symposium WTS 2008, Pomona, CA, USA, (on CD & IEEE Xplore publishing), 2008.

[2] J.-H. Jeng and T.-K. Truong, "On Decoding of Both Errors and Erasures of a Reed–Solomon Code Using an Inverse-Free Berlekamp–Massey Algorithm", IEEE Transaction on Communication, vol. 47, 1488–1494, 1999.

[3] T.C. Lin, T. K. Truong and P. D. Chen, "A Fast Algorithm for the Syndrome Calculation in Algebraic Decoding of Reed–Solomon Codes," IEEE Transaction on Communication, vol. 55, 2240–2244, 2007.

[4] U. Riemers "*Digital Video Broadcasting. The International Standard for Digital Television*". Springer-Verlag Berlin Heildeberger New York, 2001.

[5] R. Roth and G. Ruckenstein, "Efficient decoding of reed-solomon codes beyond half the minimum distance," IEEE Transaction on Information Theory, vol. 46, 246–257, 2000.

[6] W. E. Ryan and P. Conoval, "A Method of Analysis for Interleaved Reed-Solomon Coding with Erasure Decoding on Burst Error Channels", IEEE Transaction on Communication, vol. 41, 430–434, 1993.

[7] T. K. Truong, J. H. Jeng, and I. S. Reed, "Fast algorithm for computing the roots of error locator polynomials up to degree 11 in Reed-Solomon decoders," IEEE Transaction on Communication, vol. 49, no. 5, 779–783, 2001.

[8] S. B. Wicker, "*Error Control Systems for Digital Communication and Storage*". Englewood Cliffs, NJ, 1995.

[9] S. B. Wicker and V. Bhargava, "*Reed Solomon Codes and their Applications"*, NY, IEEE Press (1994).

[10] EN 301192 v1.4.1: Digital Video Broadcasting (DVB); DVB Specification for Data Broadcasting, ETSI, Nov., 2004.

Comparative Study on Asymmetric Key Cryptography Algorithms for Heterogeneous Network

Maria Nenova, Georgi Iliev, Kiril Kassev

***Abstract* – A great amount of cryptographic devices and standards apply public key cryptography, which is based on the problem of factorization of prime big digits (RSA) and retrieval of discrete logarithm proposed by El Gamal. Implementation of the approach of eliptic curves gives equivalent protection in comparison with earlier developed protocols, and it has less amount of bits.**

***Keywords* – RSA, Elliptic curves, security of information, asymmetric cryptographic algorithms.**

I. INTRODUCTION

The ability of a cryptosystem to save the needed information is defined as strength of the system. This strength is provided by the implemented key and its number of bits. Public-key cryptography is one of the most widely used technology for secure transmission of information via internet and communication systems. The concept was proposed first by Diffie and Hellman [1]. This type of cryptography is called asymmetric, and the main asymmetric algorithm is RSA – developed by Rivest, Shamir, and Adleman [2]. In asymmetric encryption algorithms are used different keys for encryption and decryption, and the decryption key cannot be derived from the encryption key, and this property is their main advantage. They are also called public key methods, and are important because they can be used for transmitting encryption keys or other data securely. The types of asymmetric encryption algorithms in heterogeneous networks are as follows: RSA, Knapsack, Digital Signature Algorithm, El Gamal, ECDSA, XTR. [3], [4], [5].

The RSA encryption algorithm is based on the multiplying of two large secret prime numbers and this is an easy forward function. In the inverse way the finding factor operation is much more difficult. The problem for an attacker is the computing of factorizing **n**. The operation performed using the cryptosystem is arranged that the operations wished to be tractable require the multiplication. The operations which should be made difficult are the finding of the plaintext from the chifer text using only the public key. This action requires performing the inverse operation — solving the factoring problem.

[1]Maria V. Nenova is with the Faculty of Telecommunications, Technical University of Sofia, 1756 Sofia, Bulgaria, E-mail: mvn@tu-sofia.bg

[2]Georgi L. Iliev is with Faculty of Telecommunications, Technical University of Sofia, 1756 Sofia, Bulgaria, E-mail: gli@tu-sofia.bg

[3]Kiril M. Kassev is with Faculty of Telecommunications, Technical University of Sofia, 1756 Sofia, Bulgaria, E-mail: kmk@tu-sofia.bg

During the last decades in order to achieve higher level of security, the amount of bits increases, which leads to big computational load. The approach implementing elliptic curves *E* ensure equivalent security in difference with earlier developed protocols, with les amounts of bits. Cryptosystems based on elliptic curves *E* are proposed in 1986 by Victor Miller and Neal Koblitz [3].

TABLE I
A comparison of key sizes needed to achieve equivalent level of security with different methods.

ECC Key Size	RSA Key Size	Key-Size Ratio	AES Key Size
163	1,024	1:6	n/a
256	3,072	1:12	128
384	7,680	1:20	192
512	15,360	1:30	256
Key sizes in bits.		*Source: Certicom, NIST*	

Those cryptographic algorithms are based on the problem of finding a discrete logarithm for points of an elliptic curve over a finite field, and exactly they are used in modern cryptograph technology. The problem of factorization of large integers and finding discrete logarithms for elements of finite groups is the main task for attacks against the algorithm.

It was designed for devices with limited compute power and memory, such as smartcards and PDAs. Elliptical curve cryptography can operate more quickly than RSA. For elliptic curve keys is not necessary to be prime, which is an ease for key generation.

Over the past few years elliptic curve cryptography has been gaining popularity and being standardized around the world by agencies as ANSI (X9.62, X9.63), IEEE group P1363 and ISO/IEC, and RSA labs and Certicom. It is proposed only with one byte to be specified the type of the elliptic curve. The U.S. National Security Agency has endorsed ECC technology by including it in its Suite B set of recommended algorithms and allows their use for protecting information classified up to top secret with 384-bit keys.

Shorter key size of ECC is an advantage. It is a reason for application of ECC in:

– Wireless communications

– Smart cards. For example in smart cards the upper limit for number of bits for the implemented key is 1024bits, that is why ECC are good alternative.

– Web applications, for example SSL.

In general where security is needed but lacks the power, for storage and computational power for the cryptosystems.

For the reasons listed above ECC is important for wireless sensor networks[8], [9].

II. MATHEMATICAL EXPLANATION

One important property of a set of solutions of an elliptic curve is that it forms a group which enables us to do cryptography. The public key is a point in the curve and the private key is a random number. The public key is obtained by multiplying the private key with the generator point G in the curve.

Elliptic curves are the set of solutions of an equation of the form $y^2 = x^3 + ax + b$. Where the coefficeints a and b are elements of the field and $4a^3 + 27b^2 \neq 0$. For different values of a, and b can be received different elliptic curves. . If the number of points on elliptic curve is marked with E, then the upper, and the lower limit for it is calculated by Hasse's theorem, and is:

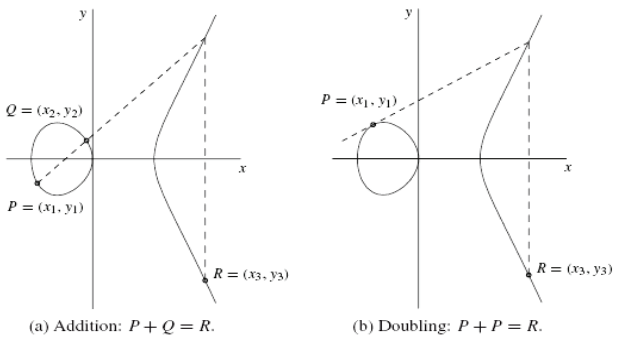
$$p + 1 - 2\sqrt{p} \leq E \leq p + 1 + 2\sqrt{p} \quad (1)$$

The elliptic curve cryptography usually is defined over two types of finite fields.

The NIST recommended [6] a certain set of elliptic curves for government use. This set of curves can be divided into two classes:

- Prime field F_p
- Binary field F_2^m

The elements of finite fields are integer numbers between 0 and p – 1, where p is representing the number of bits. All the operations such as addition, subtraction, division, multiplication are made over the integers. The prime number p is chosen such that there is finite large number of points on the elliptic curve to make the cryptosystem secure. SEC (Standards for Efficient Cryptography) [6], [7] specifies curves with p varying between 112-521 bits. In the experiments and results received in this paper is presented elliptic curve with 256 bits. The curves over F(p) are of the form [rsa lab] $y2 = x3 - 3x + b$ with b random, while the curves over F_2^m are either of the form $y2 + xy = x3 + x2 + b$ with b random or Koblitz curves. A Koblitz curve has the form $y2 + xy = x3 + ax2 + 1$ with $a = 0$ or 1.



(a) Addition: $P + Q = R$. (b) Doubling: $P + P = R$.

Fig.1. Addition and doubling of elliptic curve points

The main operations over elliptic curves are addition and doubling. Point multiplication is achieved by two basic elliptic curve operations.

- Point addition
- Point doubling

For example: if we assume m = 23 then mP can be expressed as R=23.P = 2(2(2(2P) + P) + P) + P.

If knowing two points on the curve can be received a unique third point which is the intersection of the curve with the line through the two points. If the line is tangent to the curve at a point, then that point is counted twice; if the line is parallel to the y-axis, is defined the third point as the point at infinity. Exactly one of these conditions then holds for any pair of points on an elliptic curve.

The main feature of elliptic curve cryptography is that if you have a point on the curve, all multiples of this point are also placed on the curve.

ECC are defined by the elements (p,a,b,G,n,h) defining the elliptic curve, that is called the domain parameters - a and b define the curve (by the equation); p defines the finite field F_p in the prime case - computations end by taking the remainder on division by p; G is the base point and defines the cyclic subgroup; order n represents the number of discrete points on the curve - the smallest non-negative number such that nG=O; h is called the cofactor and preferably is equal to 1.

Can be concluded that elliptic curve model is a hierarchical model of three layers:

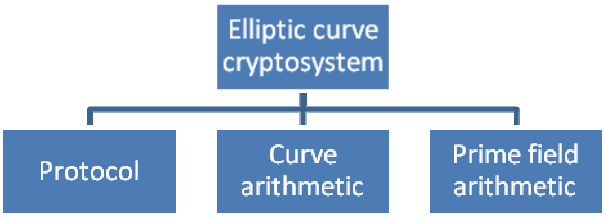


Fig.2 ECC structure

The first two layers were discussed above. On the top of the structure of ECC-system are cryptographic protocols. The protocol is a distributed algorithm for retrieving of a solution to a given cryptography task by the participants. Unfortunately between the trusted parties could be the "enemy". The aim of the protocol is predominantly to defend against unauthorized actions from the other participants. The protocols are realized for concrete aim. It is one of the fastest growing areas of theoretical cryptology, and is a field for further research.

For example a curve of the type shown bellow is used in the Microsoft Windows Media Digital Rights Management Version2.

$$y^2 = x^3 + 3176890812\ \ 5132550347\ \ 6317476413\ \ 8276932727\ \ 46955927\ x$$

$$+7905289660\ \ 7878758718\ \ 1205720257\ \ 1853543210\ \ 0651934$$

The process of the necessary amount of bits for the keys during next years is as follows:

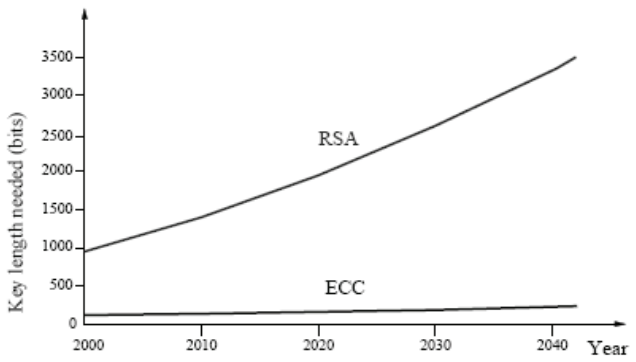


Fig 3. Compare between key length for RSA and Elliptic curves in the next few decades

The evolution in key length is according to the Moor's law. He proposed and improved that one each 18 months the number of bits increases [8], [9].

III. EXPERIMENTAL RESULTS

In this section the simulation results for the time need in key generation and the factorization are presented.

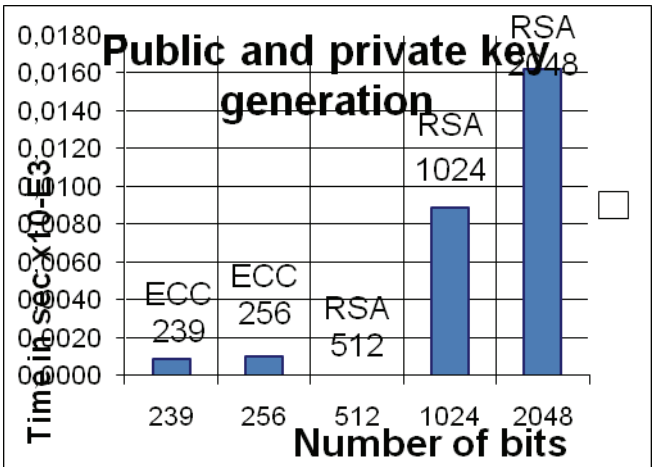


Fig5. Comparison between the time for key generation for RSA and Elliptic curves

In Fig. 5 is shown comparison between the RSA algorithms and elliptic curves implemented in cryptography and is seen the time necessary for the process of encryption. The times for the process of ECC encryption is in times smaller than the RSA even with keys of sizes of 2048 bits, which are not implemented at the moment for higher level of security (because attacks against 2048 bit keys are reported and improved to be braked).

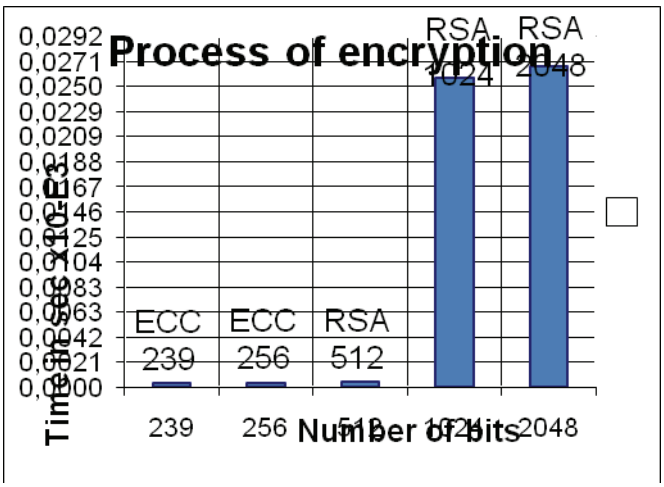


Fig6. Comparison between the times for encryption for RSA and Elliptic curves

In Fig. 6 is shown throughout the simulations comparison between times needed for generation of the keys for RSA cryptosystems and ECC system using different bit lengths. The times for the ECC239 and the ECC256 are similar to the time for RSA algorithm implementation for 512bits, which is not in use, because it is not secure.

In both graphics it can be observed that the times for generation, and encryption for the ECC crypto systems are considerably (in scales) smaller than for RSA even in the case of RSA512, that is the reason for implementation of such type of cryptography in heterogeneous networks.

In the next studied example an elliptic curve with 256 bits length is presented with all generated parameters for defining the curve as follows (the same curve is used in the previous examples):

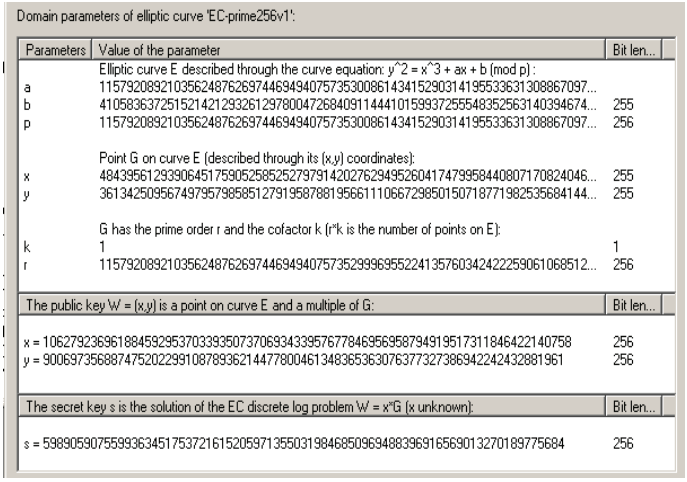


Fig. 4 Parameters of the elliptic curve

Details for the factorization of the

Input number =
77919633442827426182367047307721734498829067732826781658833934812666367

The respective next composite factor will be factorized into two factors:

1. Factorized number =
77919633442827426182367047307721734498829067732826781658833934812666367

Bit length = 239

Method: Brute Force. Time: 0.032 seconds.

First factor = 251

Bit length = 8, prime number.

Second factor =
31043678662481046287795636377578380278417955272042542493559120908871117

Bit length = 231, composite number.

2. Factorized number =
31043678662481046287795636377578380278417955272042542493559120908871117

Bit length = 231

Method: Lenstra. Time: 0.078 seconds.

First factor = 143501

Bit length = 18, prime number.

Second factor =
216330747956328153028868345012079220900327909018352084609578476 17

Bit length = 214, composite number.

3. Factorized number =
216330747956328153028868345012079220900327909018352084609578476 17

Bit length = 214

Method: Pollard. Time: 0.406 seconds.

First factor = 3293634479

Bit length = 32, prime number.

Second factor =
65681468097215699881209661399126742898153850975140974 23

Bit length = 183, prime number.

Found 4 factors in 0.516 seconds.

After generation of the public and private keys involving elliptic curve cryptology logic an example of factorization according to different methods is implemented. In the results a brute force attack is presented which is not the appropriate way for breaking the system private and public keys. A limit during the experiments is applied only for the first received factor, to be less than 1000.

IV. Conclusion

The security of a cryptosystem based on elliptic curves according to our experiments depends on the task how difficult it is to determine m given mP and P. This is referred to as the elliptic curve logarithm problem. The fastest known technique for taking the elliptic curve logarithm is known as the Pollard method and in one of our experiments we use this technique. It has been seen that a considerably smaller key size can be used for ECC compared with RSA, and DSA with curves is ten times faster than the classical RSA.

The most secure codes currently in use rely on public-key cryptography, whose security is based on the fact that computers today cannot factor very large numbers within a reasonable time period.

ACKNOWLEDGMENT

This work is supported by the Technical University of Sofia Science Fund – Grant No. 102ni191-7/2010 „Investigation of adaptive methods for QoS in communication networks”.

References

[1] Diffie, W., and Hellman, M. New directions inn cryptography, Trans. Inform. Theory IT-22, Nov. 1976, 644-654.
[2] R. L. Rivest. "Cryptology" in Handbook of Theoretical Computer Science, vol. A: Algorithms and Complexity, Elsevier and MIT Press (1990), 717-756.
[3] V.S. Miller, Use of elliptic curves in cryptography, Advances in Cryptology - Crypto '85, Springer-Verlag (1986), 417-426.
[4] T. El Gamal. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Transactions on Information Theory, v. IT-31, n. 4 469-472, 1985.
[5] http://www.nsa.gov.shtml
[6] FIPS 186-2
[7] According to http://www.secg.org/download/aid-385/sec1_final.pdf
[8] Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., and W. Polk, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile", RFC 5280, May 2008.
[9] Arjen K. Lenstra, Eric R. Verheul, “Selecting Cryptographyc Key Sizes” (1994). Journal of Cryptology of International Association for Cryptologyc Research

Forecasting the Number of IPTV Subscribers in Serbia

Valentina Radojičić[1], Goran Marković[2] and Marijana Vukašinović[3]

***Abstract* – In this paper we consider the development of IPTV service market in Serbia, which is highly attractive, both for telecom as well as cable TV operators. Based on the real statistical data, we performed the forecast of IPTV demand from the view of the main national telecom operator until the end of the year 2013. We used the main Bass diffusion model for this purpose, because the marketing promotion did not exist for this service.**

***Keywords* – Bass model, forecast, IPTV, operator, subscribers**

I. INTRODUCTION

Internet Protocol Television (IPTV) is advanced telecommunication service that delivers video and live television streams over a terrestrial broadband infrastructure, which is not tunneled through the Internet, but delivered via a service providers' private/closed network using an Internet Protocol (IP). The service is delivered to a Set Top Box (STB), which is then connected to a TV, allowing the user to consume media services in a manner very similar, but not limited to current cable/satellite TV offerings. IPTV is different to Internet TV, which is essentially video streaming via the public internet to a PC. Although it is possible to get Internet TV content to a TV, typically this is not supported by Quality of Service (QoS) and can not be controlled other than via the PC.

IPTV is a relatively new concept, which has been launched by the invasive progress of high-speed broadband technology. Western European telcos and internet service providers (ISPs) were the first to deploy IPTV services using ADSL broadband services in the early to mid 2000s. Most existing IPTV service delivery platforms deliver content in standard definition format which requires download speeds of between 2 – 4 Mbps per channel being streamed. However, as the world moves toward High Definition TV, the capacity requirements increase significantly to 8 – 10 Mbps per channel being streamed. IPTV technology has rapidly evolved since the initial deployments in the early 2000's to achieve significantly greater levels of quality and functionality.

Nowadays, IPTV is highly attractive service to both telecom operators as well as ISPs (Internet Service Providers), because it represents an entirely new revenue stream to bundle with their traditional voice and data services. It is expected that the IPTV will become the killer broadband application during the next few years. The worldwide growth of IPTV subscribers during the last years was very dynamic. The new forecast from Multimedia Research Group (MRG) indicates that the number of global IPTV subscribers will grow from 28 million in 2009 to 83 million in 2013, a compound annual growth rate of 31%, as illustrates Fig.1 [1]

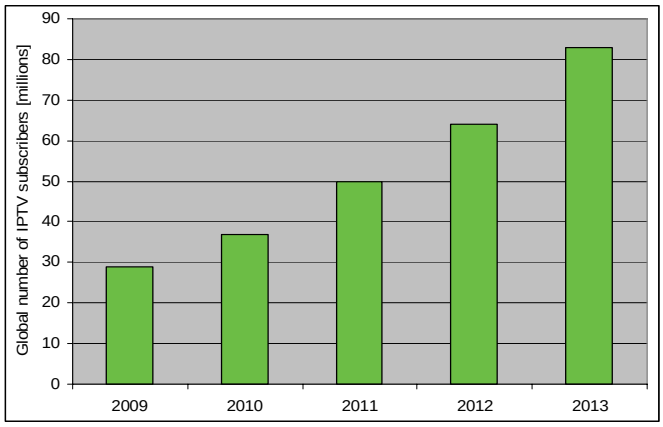


Fig.1 Global number of IPTV subscribers [source MRG, 2009]

IPTV adoption in Western Europe will continue to enjoy dynamic growth up to the 2014 — 19% per year — resulting in 20.3 million IPTV households in 2014 [2]. However, during that period, the dynamics of the IPTV market will change: The main growth factors that drove the IPTV market during its launch phase — home broadband adoption and the analog TV switch-off process — will gradually fade out, and short-term growth will mainly come from the conversion of secondary TVs to digital TV. The IPTV forecast for next five years is both conservative and optimistic, based on very detailed semi-annual analysis of individual Service Providers and on a country-by-country basis.

In this paper, we focused on forecasting the number of IPTV subscribers at Serbian telecommunication market for the main national telecom operator. For this purpose we used the main Bass diffusion model, because the marketing campaign did not exist [3]. The paper is organized in following manner. The second Section describes the basic analytical expression of the Bass diffusion model. The third Section gives the forecasted results for the IPTV users in Serbia. In the final section, the concluding remarks are given.

II. THE BASS DIFFUSION MODEL

The mathematical structure of the Bass model is derived from a hazard function corresponding to the conditional probability that an adoption will occur at time *t* given that it

[1]Valentina Dz. Radojičić is with the University of Belgrade - Faculty of Traffic and Transport Engineering, V. Stepe 305, 11000 Belgrade, Serbia, E-mail: valentin@sf.bg.ac.rs

[2]Goran Z. Marković is with the University of Belgrade - Faculty of Traffic and Transport Engineering, V. Stepe 305, 11000 Belgrade, Serbia, E-mail: g.markovic@sf.bg.ac.rs

[3]Marijana M. Vukašinović is with Telekom Serbia, Takovska 2, 11 000 Belgrade, marijanavuk@telekom.rs

has not occurred yet [4,5,6]. If *f(t)* is the density function of time to adoption and *F(t)* is the cumulative fraction of adopters at time *t*, the basic hazard function underlying the Bass model is given by following equation:

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

Parameter *q* reflects the influence of those consumers who have already adopted the product/service (i.e. word-of-mouth communication from previous adopters), while *p* captures the influence that is independent from the number of adopters (i.e. external communication). The solution of equation (1) is given by following equation:

$$F(t) = \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} \quad (2)$$

If *m* denotes the market potential for the new product/service, then the cumulative number of adoptions, *Y(t)*, at time *t* is given by:

$$Y(t) = m \cdot F(t), \quad (3)$$

while the current (non-cumulative) product or service sale at time *t*, *S(t)* is given by:

$$S(t) = mf(t) = pm + (q-p)mF(t) - qm(F(t))^2 =$$
$$= pm + (q-p)Y(t) - q/mY^2(t) = a + bY(t-1) + c[Y(t-1)]^2 \quad (4)$$

Where *t*=2,3,..., *Y*(*t*-1) is the cumulative sale at time *t*-1, *a*, *b*, *c* are the coefficients that have to be calculated based on the regression analysis. Hence, the value for *m* can be obtained as:

$$m = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \quad (5)$$

For *m*>0, the Bass model parameters of innovation (*p*) and imitation (*q*), could be obtained by following equations:

$$p = \frac{a}{m} \quad (6)$$

$$q = -mc \quad (7)$$

Based on thus estimated parameters *p* and *q*, the new service/product adoption for future period can be calculated.

III. Forecasting the Number of IPTV Subscribers

Based on the data of actual and forecasted total IPTV subscribers in the world, it was shown in our research that the estimated Bass parameters *p* and *q* are not directly applicable to the Serbian telecommunication market. Namely, if we use the same values for *p* and *q* (calculated based on the relevant historical IPTV statistical data), we obtained the great disproportion between the actual and forecasted results for the IPTV users in Serbia. Therefore, first of all, we try to estimate the valid values of parameters *p* and *q*, based on the real data from the Serbian IPTV market during the previous year.

Table I gives the data of the actual IPTV service sales during the year 2009. The total market potential for IPTV service in Serbia is currently estimated to be *m* = 500.000 subscribers.

In order to estimate the values for *p* and *q*, we performed the linear regression analysis to calculate the unknown parameters *a*, *b*, and *c*, from the equation (4).The values of (one time period) delayed cumulative sales *Y(t-1)*, as well as their square values, Y^2(t-1) could be simply obtained from the last column of Table I. The results of the regression analysis are given by Table II.

TABLE I
NUMBER OF IPTV SUBSCRIBERS DURING THE YEAR 2009

Month (2009)	Monthly # of IPTV subscribers *S(t)*	Cumulative # of IPTV subscribers *Y(t)*
January	1.975	1.975
February	1.594	3.569
March	1.425	4.994
April	1.152	6.146
May	1.930	8.076
June	3.408	11.484
July	3.421	14.905
August	2.640	17.545
September	1.679	19.224
October	1.509	20.733
November	1.704	22.437
December	1.579	**24.016**

TABLE II
THE RESULTS OF REGRESSION ANALYSIS

$S(t) = a + bY_{t-1} + cY_{t-1}^2$	
a	1302
b	0,5
c	-0,00000096

Based on the obtained values for *a*, *b* and *c*, we can now calculate the Bass parameters *p* and *q* using the equations (6) and (7), respectively:

$p = a/m = 1302/500000 = 0{,}002604$

$q = -cm = 0{,}00000096 * 500000 = 0.48$

Thus, the forecasted values of non-cumulative sales at time *t*, obtained by the Bass diffusion model could be calculated using the following expression:

$$p*(m-Y_{(t-1)}) + q*(Y_{(t-1)}/m)(m-Y_{(t-1)})$$

For the first time (month), the forecasted non-cumulative IPTV sale is 1302 subscribers. For the second month, the forecasted number of IPTV users equals to 0.002604*(500.000-1.302)+0,48*(1.302/500.000)*(500.000 - 1.302) = 1922. The calculated forecasted values for other months during the year 2009 are given in Table III.

TABLE III
FORECASTED NUMBER OF IPTV SUBSCRIBERS IN 2009
(p=0,002604 and q=0,4)

Month in year (2009)	Actual non-cumulative IPTV sale	Forecasted non-cumulative IPTV sale	Actual cumulative IPTV sale	Forecasted cumulative IPTV sale
January	1.975	1.302	1.975	1.302
February	1.594	1.922	3.569	3.224
March	1.425	2.831	4.994	6.055
April	1.152	3.549	6.146	9.604
May	1.930	4.309	8.076	13.913
June	3.408	4.994	11.484	18.908
July	3.421	5.660	14.905	24.568
August	2.640	6.279	17.545	30.847
September	1.679	6.865	19.224	37.712
October	1.509	7.411	20.733	45.124
November	1.704	7.922	22.437	53.046
December	1.579	8.396	24.016	61.442

Fig. 2 illustrates the comparative results for actual and forecasted cumulative number of IPTV users in a case of parameters values p=0, 002604 and q=0,4 for the year 2009.

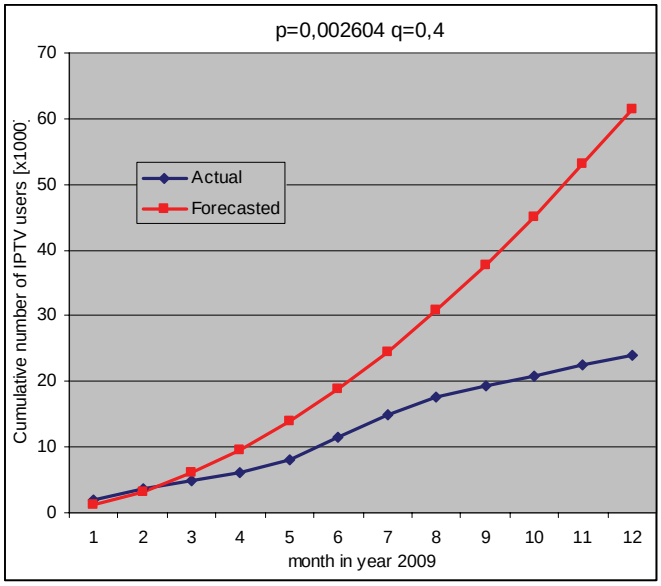


Fig.2 Cumulative number of IPTV users in year 2009 (p=0,002604 and q=0,4)

It could be seen from the Fig.2 that there are huge disagreements between the actual and forecasted cumulative number of IPTV subscribers. Therefore, the values for diffusion parameters have to be updated to achieve better accuracy between these two set of data. By keeping the same value of the parameter p and by iterative reducing the value of parameter q, it could be obtained better agreement between the actual and forecasted results. In the case when the values of parameters are q=0,08 and p=0,002604, the obtained forecasted results are given in Table IV.

TABLE IV
FORECASTED NUMBER OF IPTV SUBSCRIBERS IN 2009
(p=0,002604 and q=0,08)

Month in year (2009)	Actual non-cumulative IPTV sale	Forecasted non-cumulative IPTV sale	Actual cumulative IPTV sale	Forecasted cumulative IPTV sale
January	1.975	1.302	1.975	1.302
February	1.594	1.402	3.569	2.704
March	1.425	1.510	4.994	4.215
April	1.152	1.625	6.146	5.840
May	1.930	1.749	8.076	7.589
June	3.408	1.880	11.484	9.469
July	3.421	2.020	14.905	11.489
August	2.640	2.170	17.545	13.659
September	1.679	2.329	19.224	15.989
October	1.509	2.499	20.733	18.487
November	1.704	2.678	22.437	21.165
December	1.579	2.868	24.016	24.034

Fig. 3 illustrates the comparative results for actual and forecasted cumulative number of IPTV users in a case of parameters values p=0,002604 and q=0,08 for the year 2009.

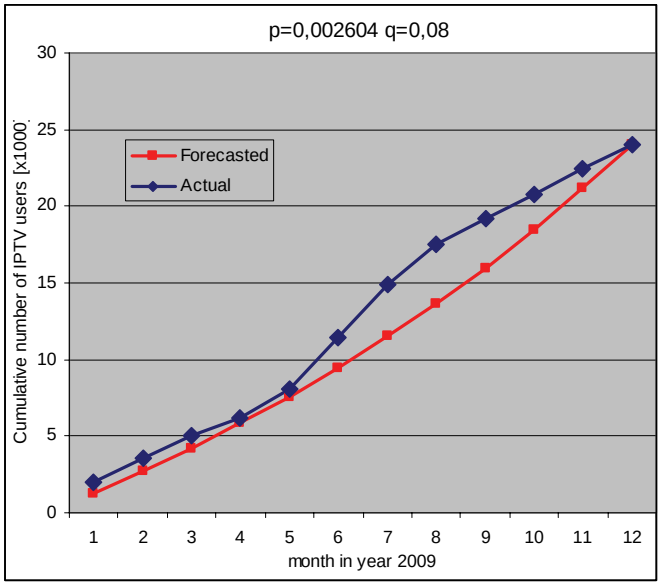


Fig.3 Cumulative number of IPTV users in year 2009 (p=0,002604 and q=0,08)

Based on the results given in Fig.3, it can be seen that in a case of parameters values p=0,002604 and q=0,08, there are acceptable disagreements between the actual and forecasted results. Therefore, we can used these diffusion parameters values to forecast the IPTV demand in short-term future period.

The forecasted cumulative and current (monthly) results for IPTV service demand, from the begining of the year 2009 up to the end of the year 2013 (60 months) are given by Fig. 4 and Fig.5, respectively.

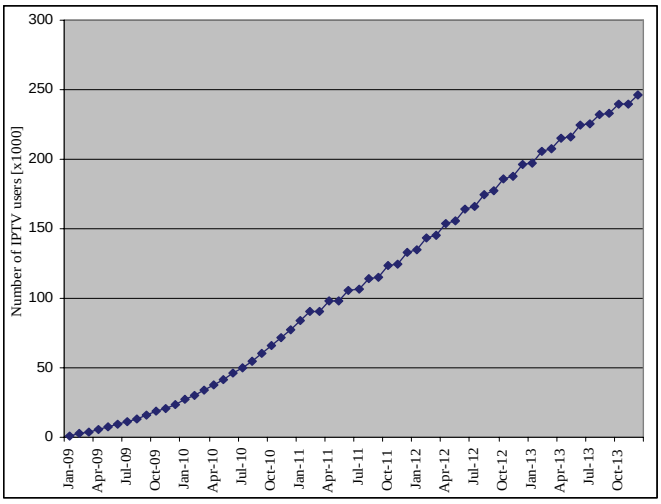


Fig. 4. Forecasted cumulative number of IPTV subscribers

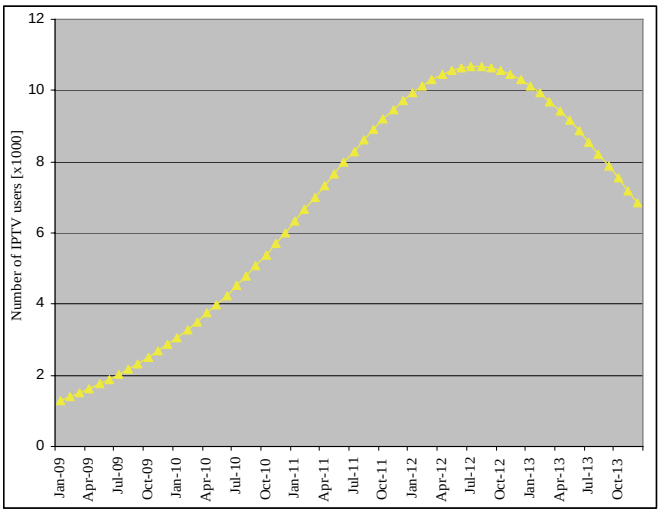


Fig. 5. Forecasted monthly number of new IPTV subscribers

Finally, in the Fig.6 we compared our forecasted cumulative number with the firstly planned number of the IPTV users after it being actually launched by the telecom operator (in the year 2008).

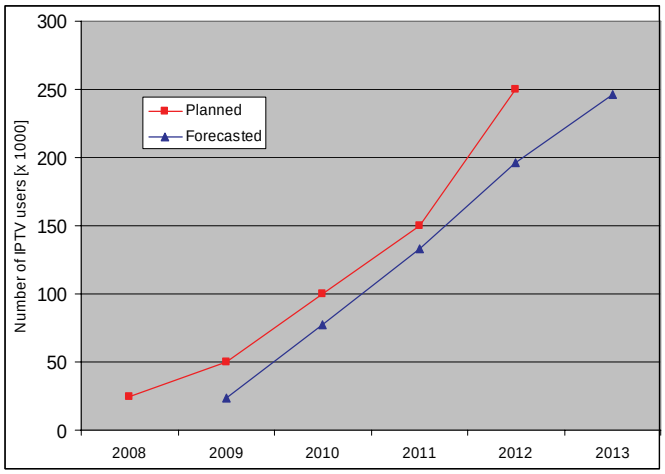


Fig. 6. Comparision of the forecasted and planned number of IPTV the user

The real number of users at the end of the year 2009 was nearly the forecasted number in that moment, which is less than the planned number of users, as could be seen from the Fig 6. It means that the plann was not fulfilled beginning from the service was launched.

IV. Conclusion

The IPTV service might become the killer broadband application in the near future. Both, telecom operators as well as cable operators are highly interested to offer the IPTV service to their customers, mainly due to increase their profits and also to achieve the greater satisfaction of the end users. There were many research studies concerned with the forecasting the number of IPTV users in different part of the world. However, a particular and precise forecast has to be done for each country as well as for each operator in the considered country. In this paper we presented the results of the forecasted number of IPTV subscribers from the point of view of the main Serbian telecommunication operator.

By analyzing the obtained results, we can conclude that the preliminary expectation plans about the adoption rate of the IPTV service were not achieved. The actual number of subscribers was only 50% than expected at the end of previous year. The main reasons could be found in facts that the time of the service launched was delayed, also due to relatively high service implementation costs to the individual subscribers. Such situation indicates that the potential market was not considered adequately. We took into account that circumstances in our forecasting procedure and try to obtain more precise data about the future demands for IPTV service.

Acknowledgement

This paper resulted from the researching project "*Next Generation Optical Networks – researching the possibilities for improving the Serbian optical transport network"* (project designation TR-11013) that is supported by the Serbian Ministry of Science and Technological Development.

References

[1] IPTV Global Forecast – 2009 to 2013: Semiannual IPTV Global Forecast Report - Edition 1, May 2009, http://www.mrgco.com

[2] European IPTV Forecast, 2009 To 2014, (http://www.forrester.com/rb/Research/european_iptv_forecast_2C_2009_to_2014/q/id/55325/t/2)

[3] V. Radojičić, *Forecasting in telecommunications*,Belgrade, Faculty of Traffic and Transp. Engin. 2003 (origin in Serbian).

[4] V. Radojičić, A. Kostić-Ljubisavljević, G. Marković: "*Forecasting the adoption of new telecommunication services*", Tehnika, no..5, Vol. 57, pp. 9-16, 2007 (origin in Serbian).

[5] V. Radojičić, G. Marković, » New Technology Forecasting using the Bass Model«,*Proceedings of the TELSIKS 2009 Conference*, Vol.1, pp. 277-280, 7-9 oct. 2009, Niš, Serbia

[6] V. Radojičić, G. Marković, „Model of the different brand diffusion at the telecommunication market", *Proceed. of the XVI TELFOR Conference* (*CD*) nov. 2008; Belgrade (origin in Serbian), www.telfor.org.yu

Express Mail Service Quality Improvement Using TETRA Communication System

Momčilo Dobrodolac[1], Dejan Marković[2], Bojan Stanivuković[3], Mladenka Blagojević[4]

Abstract **– Express and courier mail service is elite postal service. One of the basic conditions for achieving high quality of service is adequate communication between dispatchers and couriers. Council dilemma is how to transfer, as soon as possible, all the required information from the dispatcher to courier. The problem is that the courier is somewhere in the city and the communication with him\her is via a mobile terminal. The basic lack of current communication is that used system is part of the existing public system with the disadvantages that characterize it, related to security, availability, and others. Therefore, the purpose of this paper is to propose the implementation of TETRA communication system in order to improve the quality of express mail service.**

Keywords **– express mail service, quality, improvement, TETRA, communication system.**

I. INTRODUCTION

Companies involved in express mail industry provide fast, reliable, on demand, integrated, door-to-door transportation service of letters, reports, documents and other goods. Each shipment is tracked and its course of travel is fully controlled. This service is a business class in the field of postal traffic.

Express industry makes transportation of items easier and faster. Beside to the macroeconomic contribution, express industry has an important role by creating a direct impact on the capacity and competitiveness of other sectors of the economy, the productivity of companies that use its services, the reliability of the manufacturing process, reducing the cost of supplies, etc. It also promotes regional economic development and has impact on increasing investment and employment. [1] Modern business can not be imagined without the express industry.

Today, leading companies in this field are multinational, with worldwide coverage territory. These companies have their own fleet of aircrafts and vehicles for land transportation. Airline fleets of these companies are larger than many national airlines.

[1] Momčilo Dobrodolac is with The Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe Street 305, 11000 Belgrade, Serbia, E-mail: m.dobrodolac@gmail.com

[2] Dejan Marković is with The Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe Street 305, 11000 Belgrade, Serbia, E-mail: mdejan@sf.bg.ac.rs

[3] Bojan Stanivuković is with The Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe Street 305, 11000 Belgrade, Serbia, E-mail: stanivukovic@yahoo.com

[4] Mladenka Blagojević is with The Faculty of Transport and Traffic Engineering University of Belgrade, Vojvode Stepe Street 305, 11000 Belgrade, Serbia, E-mail: m.blagojevic@sf.bg.ac.rs

Serbian Post, beside its rich history, has introduced express and courier services in year 2002, by department called Post Express. As basic characteristics of these services, the Serbian Post highlights security in transportation, speed, high quality and reliability. Within a relatively short time, the national postal operator, has taken a significant share of the express mail market in Serbia. [2]

The user can request the service by phone, by calling the Call Center or by coming to Post Office. Given that one of the most important elements that provide additional quality to express courier service is exactly the coming to client's home or business address, this type of service request is very common. Speed of responding to the call is a parameter which the users give a great importance in assessing the overall quality of service.

The basic element for successful technological process and the high quality of service lies in adequate communication system between all the entities participating in this process, especially between dispatchers and couriers. Therefore, the subject of this paper is precisely the impact of different communication systems on the technological process. The new system for communication between dispatchers and couriers is proposed and it could contribute to increasing the productivity of enterprises and greater customer satisfaction.

II. THE ROLE OF COMMUNICATION SYSTEM IN TECHNOLOGICAL PROCESS

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

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

information about the sender, recipient, type of required services, some specific services, weight and shipping volume, the way of payment and others.

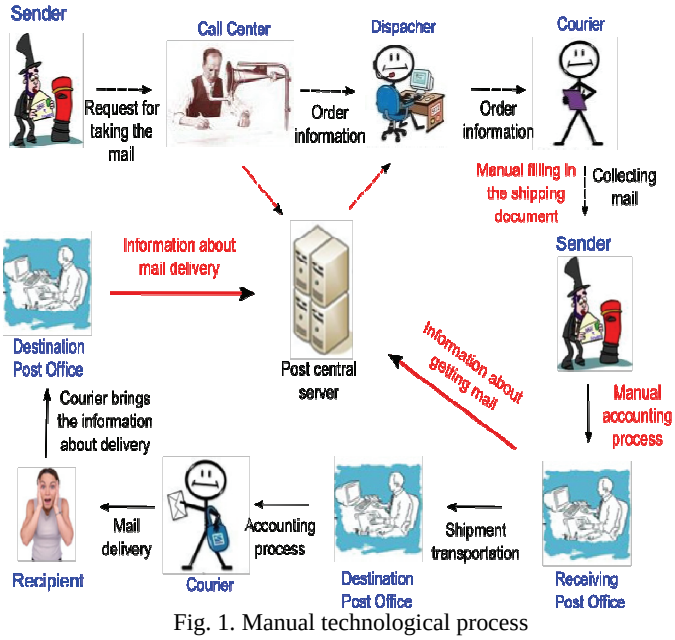


Fig. 1. Manual technological process of express mail transportation

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

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

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

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

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

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

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

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

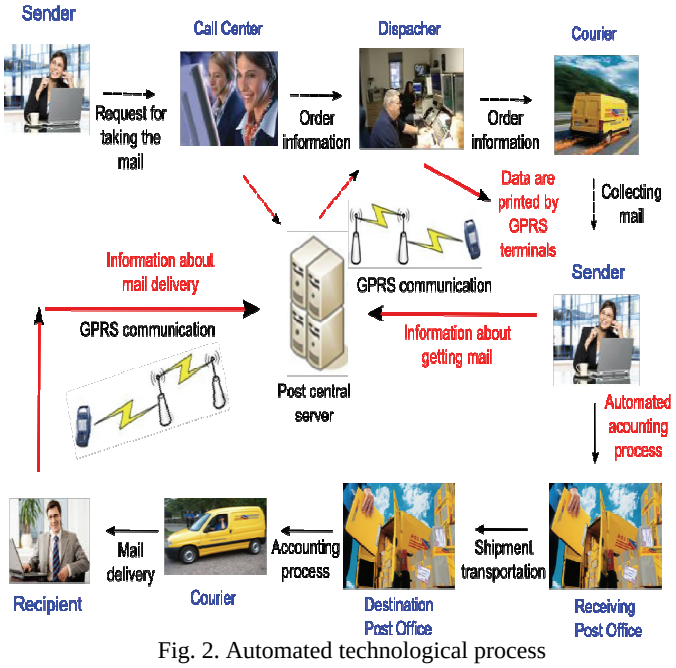


Fig. 2. Automated technological process of express mail transportation

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

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

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

Comparing manual and automated production process, it can be concluded that there are many advantages in using the GPRS terminal at several points of production cycle. These benefits greatly contribute to reducing costs, increasing speed of providing services, better utilization of labor, increasing the quality of services and better productivity. However, beside all these advantages, the automated system is implemented in a very small number of cases. The most common reason for this is poor availability of GPRS terminals because the system is designed to be part of a public communication network. As a potential solution to this problem, in this paper the new communication system for closed user groups is introduced.

III. TETRA COMMUNICATION SYSTEM

Terrestrial Trunked Radio (TETRA) is a digital trunked mobile radio standard developed by the European Telecommunications Standards Institute (ETSI). The purpose of the TETRA standard was to meet the needs of traditional Professional Mobile Radio (PMR) user organizations such as those listed below:

- Public Safety: police, customs,
- National safety: military, border guards,
- Emergency services: fire, rescue, first aid, ambulance,
- Authorities: prison administrations, other authorities,
- Transportation,
- Utilities,
- Government,
- Military,
- Public Access Mobile Radio (PAMR),
- Commercial & Industry,
- Fleet Management,
- Closed User Groups.

Because the TETRA standard has been specifically developed to meet the needs of a wide variety of traditional PMR user organizations it has a scaleable architecture allowing economic network deployments ranging from single site local area coverage to multiple site wide area national coverage. Besides meeting the needs of traditional PMR user organizations, the TETRA standard has also been developed to meet the needs of Public Access Mobile Radio (PAMR) operators.

The common mode of operation for TETRA is a group calling mode in which a single button push will connect the user to the users in a selected call group and/or a dispatcher. It is also possible for the terminal to act as a one-to-one walkie talkie but without the normal range limitation since the call still uses the network. If enabled by the Subscriber Management, TETRA terminals can act as mobile phones, with a full-duplex direct connection to other TETRA Users or the PSTN. Emergency buttons, provided on the terminals, enable the users to transmit emergency signals, to the dispatcher, overriding any other activity taking place at the same time.

TETRA uses Time Division Multiple Access (TDMA) with four user channels on one radio carrier and 25 kHz spacing between carriers. Both point-to-point and point-to-multipoint transfer can be used. Digital data transmission is also included in the standard though at a low data rate.

TETRA Mobile Stations (MS) can communicate Direct Mode or using Trunked infrastructure (Switching and Management Infrastructure or SwMI) made of TETRA Base Stations (TBS). As well as allowing direct communications in situations where network coverage is not available, Direct Mode or DMO also includes the possibility of using one (or a chain) of TETRA terminals as relays for a signal. This functionality is called DMO gateway (from DMO to TMO) or DMO Repeater (DMO to DMO). In Emergency situations this feature allows direct communications underground or in areas of bad coverage.

In addition to voice and dispatch services, the TETRA system supports several types of data communication. Status messages and Short Data Services (SDS) are provided over the system's main control channel, while Packet Data or Circuit switched data communication uses specifically assigned traffic channels.

TETRA provides Authentication Mechanisms of Terminals towards Infrastructure and vice versa. For protection against eavesdropping, air interface encryption and end-to-end encryption is available.

The main advantages of TETRA over other technologies, such as GSM, are:

- High spectral efficiency - 4 channels in 25 kHz and no guard bands, compared to GSM with 8 channels in 200 kHz and guard bands,
- The much lower frequency used gives longer range, which in turn permits very high levels of geographic coverage with a smaller number of transmitters, thus cutting infrastructure costs,
- Very fast call set-up - a one to many group call is generally set-up within 0.5 seconds (typical less than 250 msec for a single node call) compared with the many seconds (typically 7 to 10s) that are required for a GSM network.,

- The system contains several mechanisms, designed into the protocols and radio parameters, to ensure communication success even during overload situations (e.g. during major public events or disaster situations), thus calls will always get through unlike in cellular systems. The system also supports a range of emergency calling modes,
- TETRA infrastructure is usually separate from (but connected to) that of the public (mobile) phone networks, resulting in (normally) no call charges for the system owners, substantially more diverse and resilient communications and it is easy to customize and integrate with data applications (vehicle location, GIS databases, dispatch systems etc).
- Communication between different types of networks is becoming increasingly important. TETRA has the ability to connect with a variety of external networks. For example, the TETRA network can be connected with the public and private telephone networks, different types of data network and a large management and signaling networks. All of these networks can be accessed directly from a mobile terminal. The sheme of various possibilities of tetra system is shown in Fig. 3.

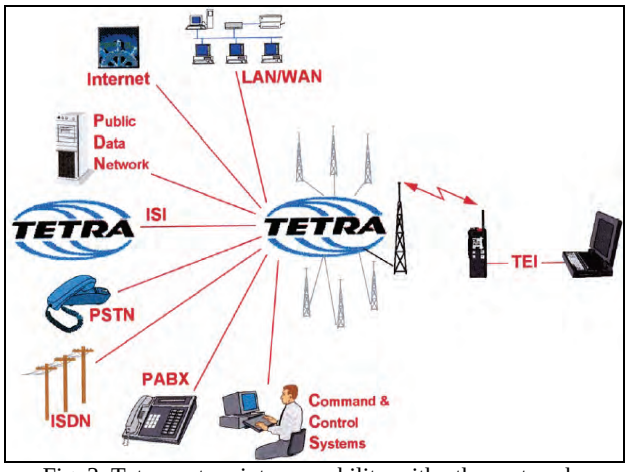


Fig. 3. Tetra system interoperability with other networks

- Unlike most cellular technologies, TETRA networks typically provide a number of fall-back modes such as the ability for a base station to process local calls. So called Mission Critical networks can be built with TETRA where all aspects are fail-safe/multiple-redundant.
- In the absence of a network mobiles/portables can use 'direct mode' whereby they share channels directly (walkie-talkie mode).
- Gateway mode - where a single mobile with connection to the network can act as a relay for other nearby mobiles that are out of range of the infrastructure.
- TETRA also provides a point-to-point function that traditional analogue emergency services radio systems did not provide. This enables users to have a one-to-one trunked 'radio' link between sets without the need for the direct involvement of a control room operator/dispatcher.
- unlike the cellular technologies, which connect one subscriber to one other subscriber (one-to-one) then TETRA is built to do one-to-one, one-to-many and many-to-many. These operational modes are directly relevant to the public safety and professional users.
- Equipment is available from many suppliers around the world, thus providing the benefits of competition

Main disadvantages of TETRA system are:

- Handsets are more expensive than cellular (about 750 EUR in 2003, about 600 EUR in 2006). This is due to the more difficult technology, smaller economies of scale, and different business model (e.g.: need for security, high powers and robustness),
- Data transfer is efficient and long range (many km), but slow by modern standards at 7.2 kbit/s per timeslot (3.5 kbit/slot net packet data throughput), although up to 4 timeslots can be combined into a single data channel to achieve higher rates whilst still fitting into a single 25 kHz bandwidth channel. Latest version of standard supports 115.2 kbit/s in 25 kHz or up to 691.2 kbit/s in an expanded 150 kHz channel,
- Requires a linear amplifier to meet the stringent RF specifications that allow it to exist alongside other radio services.

Having in mind all the advantages and disadvantages of TETRA communication system, the conclusion can be reached that implementation of this communication system would lead to improvement of technological process of express mail service.

IV. Conclusion

In this paper, the importance of communication system between dispatcher and courier is presented. It is clear that huge optimization of technological process of express mail service could be done be using adequate communications. Saving could be achieved in materials, labor and working time. Quality of service could be improved and productivity of express company increased. As one of the possibility for adequately communication, TETRA system is introduced with its advantages and disadvantages.

References

[1] Dejan Marković, Momčilo Dobrodolac, „The Impact of Express Industy on Global Economy“, 24. Symposium on Novel Technologies in Postal and Telecommunication Traffic – PosTel 2006, Conference Proceedings, pp. 196 – 207, Belgrade, Serbia, 2006.

[2] Momčilo Dobrodolac, “The Research of Models for Market Analyzing of Postal and Courier Systems”, Master Thesis, The Faculty of Transport and Traffic Engineering University of Belgrade, Serbia, 2008.

[3] J. Dunlop, D. Girma, J. Irvine, *Digital Mobile communications and the TETRA system*, Wiley, New York, USA, 2000.

[4] http://www.bakom.ch/

[5] http://www.tetramou.com/

Sequential Symbol Synchronizers based on Pulse Comparation by Positive Transitions at Quarter Rate

Antonio D. Reis[1,2], Jose F. Rocha[1], Atilio S. Gameiro[1] and Jose P. Carvalho[2]

***Abstract* – This work presents the synchronizer based on pulse comparation, between variable and fixed pulses.**

This synchronizer has two variants, one operating by both transitions at the bit rate and other operating by positive transitions at quarter rate. Each variant has two versions namely the manual and the automatic.

The objective is to study the four synchronizers and evaluate their output jitter UIRMS (Unit Interval Root Mean Square) versus input SNR (Signal Noise Ratio).

***Keywords* – Synchronism in Digital Communications**

I. INTRODUCTION

This work studies the sequential symbol synchronizer, with a phase comparator based on a pulse comparation, between a variable pulse Pv and a fixed reference pulse Pf.

The synchronizer has four types supported in two variants one operating by both transitions at the bit rate and other operating by positive transitions at quarter rate [1, 2, 3, 4, 5].

The variant at the rate has two versions namely the manual (b-m) and automatic (b-a). The variant at quarter rate has two versions namely the manual (p-m/4) and automatic (p-a/4).

The difference between them is in the phase comparator since the other blocks are equal [6, 7, 8, 9, 10, 11, 12].

The error pulse Pe (Pv - Pf) controls the VCO (Voltage Controlled Oscillator) to synchronize with the input data. The VCO output is the clock, with good quality, that samples appropriately the input data and retimes its bit duration.
Fig.1 shows the blocks of the symbol synchronizer.

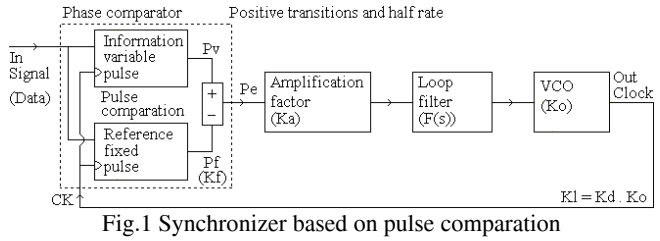


Fig.1 Synchronizer based on pulse comparation

Kf is the phase comparator gain, F(s) is the loop filter, Ko is the VCO gain and Ka is the loop gain factor that controls the root locus and then the loop characteristics.

In priori and actual-art state was developed various synchronizers, now is necessary to know their performance.

The motivation of this work is to create new synchronizers and to evaluate their performance with noise. This contribution increases the knowledge about synchronizers.

Following, we present the variant both transitions at rate with their manual (b-m) and automatic (b-a) versions. Next, we present the variant positive transitions at quarter rate with their manual (p-m/4) and automatic (p-a/4) versions.

After, we present the design and tests. Then, we present the results. Finally, we present the conclusions.

II. SYNCHRONIZERS OPERATING AT THE RATE

The synchronizer with its phase comparator operates, here, by both transitions at the data transmission rate.

This variant has the manual (b-m) and automatic (b-a) versions, the difference is in phase comparator. The variable pulse Pv, produced by the first flip flop with exor, is equal in the two versions, but the fixed pulse Pf is different [1, 2].

A. Both transitions, at the rate and manual

The manual version has a phase comparator, where the fixed pulse Pf is produced by an exor with a delay $\Delta t=T/2$, that needs a previous manual adjustment (Fig.2)

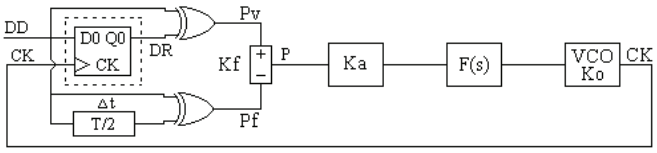


Fig.2 Synchronizer both at the rate and manual (b-m)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

Fig.3 shows the waveforms of the synchronizer operating at the rate and manual version.

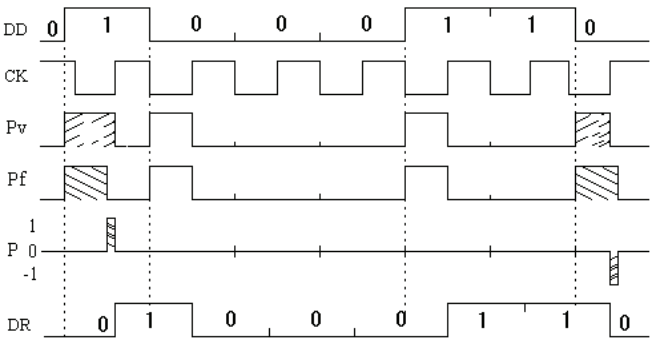


Fig.3 Waveforms of the synchronizer at the rate and manual

The error pulse Pe diminishes and disappear at the equilibrium point.

[1]Antonio D. Reis, Jose F. Rocha and Atilio S. Gameiro are with Dep. Electrónica, Universidade Aveiro / Instituto Telecomunicações, 3800 Aveiro, Portugal, E-mail: amg@det.ua.pt; frocha@det.ua.pt.

[2]Antonio D. Reis and Jose P. Carvalho are with the Dep. Fisica, Universidade da Beira Interior Covilhã / Unidade Detecção Remota, 6200 Covilhã, Portugal, E-mail: adreis@ubi.pt; pacheco@ubi.pt.

B. Both transitions, at the rate and automatic

The automatic version has a phase comparator where the fixed pulse Pf is produced automatically by the second flip flop with exor, without previous adjustment (Fig.4).

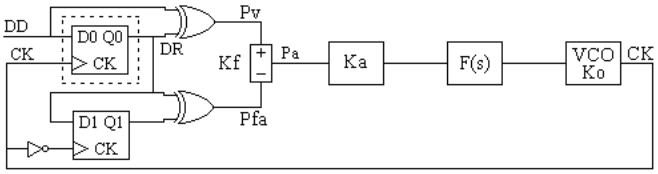


Fig.4 Synchronizer both at the rate and automatic (b-a)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

Fig.5 shows the waveforms of the synchronizer operating at the rate and automatic version.

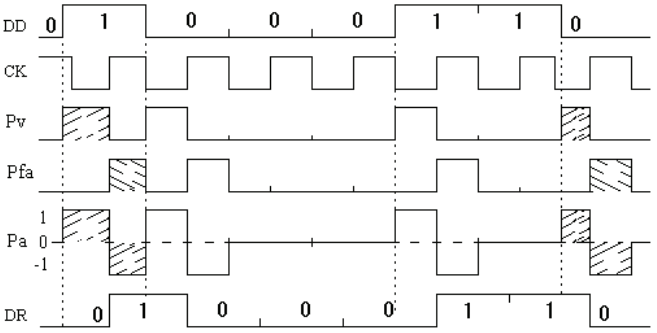


Fig.5 Waveforms of the synchronizer at the rate and automatic

The error pulse Pe don't disappear, but the variable area Pv is equal to the fixed one Pf at the equilibrium point.

III. SYNCHRONIZERS OPERATING AT QUARTER RATE

The synchronizer with its phase comparator operates, here, by positive transitions at quarter data transmission rate.

This variant has the manual (p-m/4) and the automatic (p-a/4) versions, the difference is only in the phase comparator. The variable pulse Pvp, based in the four first flip flops with multiplexer, is equal in the two versions, but the fixed pulse Pfp is produced from a different way [3, 4].

A. Positive transitions, at quarter rate and manual

The manual version has a phase comparator, where the fixed pulse Pf is produced by an exor with a delay Δt=T/2, that needs a previous manual adjustment (Fig.6).

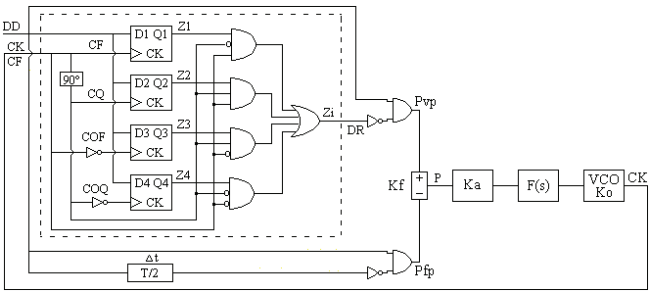


Fig.6 Synchronizer positive at quarter rate and manual (p-m/4)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

Fig.7 shows the waveforms of the synchronizer operating at quarter rate and manual version.

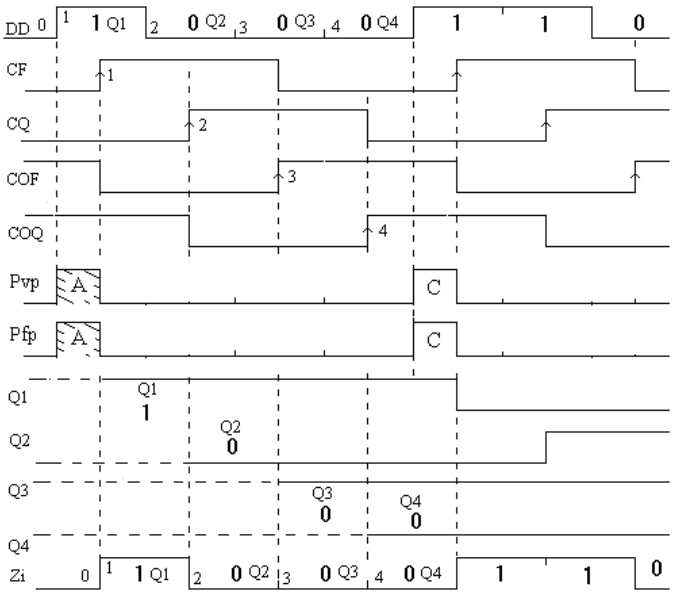


Fig.7 Waveforms of the synchronizer at quarter rate and manual

The error pulse Pe diminishes and disappear at the equilibrium point

B. Positive transitions, quarter rate and automatic

The automatic version has a phase comparator, where the fixed pulse Pf is produced automatically by the seconds flip flops and multiplexer with exor, without previous adjustment (Fig.8).

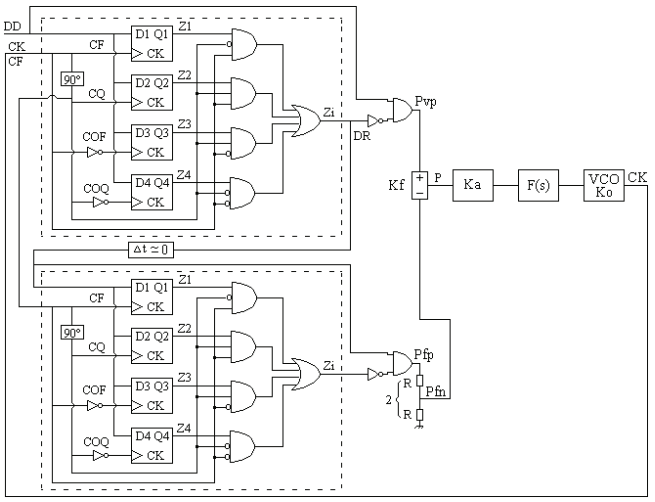


Fig.8 Synchronizer positive at quarter rate and automatic (p-a/4)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

Fig.9 shows the waveforms of the synchronizer at quarter rate and automatic version.

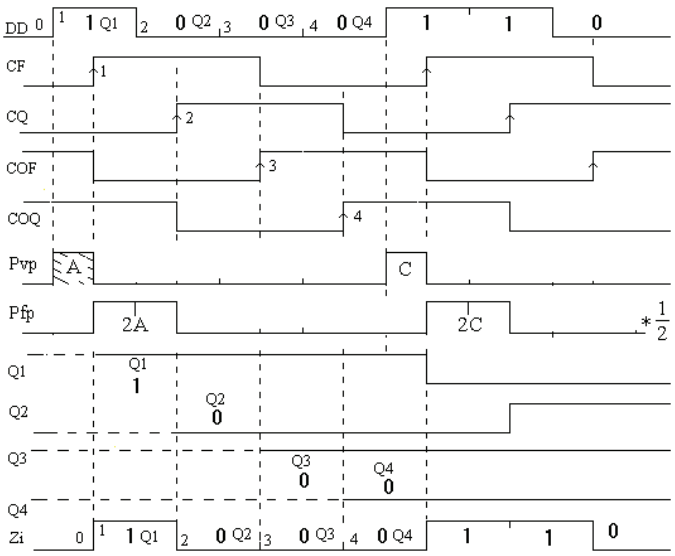


Fig.9 Waveforms of the synchronizer at quarter rate and automatic

The error pulse Pe don't disappear but the positive area is equal to the negative at the equilibrium point.

IV. DESIGN, TESTS AND RESULTS

We will present the design, the tests and the results of the referred synchronizers [5].

A. Design

To get guaranteed results, it is necessary to dimension all the synchronizers with equal conditions. Then it is necessary to design all the loops with identical linearized transfer functions.

The general loop gain is Kl=Kd.Ko=Ka.Kf.Ko where Kf is the phase comparator gain, Ko is the VCO gain and Ka is the control amplification factor that permits the desired characteristics.

For analysis facilities, we use a normalized transmission rate tx=1baud, what implies also normalized values for the others dependent parameters. So, the normalized clock frequency is fCK=1Hz.

We choose a normalized external noise bandwidth Bn = 5Hz and a normalized loop noise bandwidth Bl = 0.02Hz. Later, we can disnormalize these values to the appropriated transmission rate tx.

Now, we will apply a signal with noise ratio SNR given by the signal amplitude Aef, noise spectral density No and external noise bandwidth Bn, so the SNR = $A^2_{ef}/(No.Bn)$. But, No can be related with the noise variance σn and inverse sampling Δτ=1/Samp, then $No=2\sigma n^2.\Delta\tau$, so $SNR=A^2_{ef}/(2\sigma n^2.\Delta\tau.Bn) = 0.5^2/(2\sigma n^2*10^{-3}*5)= 25/\sigma n^2$.

After, we observe the output jitter UI as function of the input signal with noise SNR. The dimension of the loops is

- 1st order loop:

The loop filter F(s)=1 with cutoff frequency 0.5Hz (Bp=0.5 Hz is 25 times bigger than Bl=0.02Hz) eliminates only the high frequency, but maintain the loop characteristics.

The transfer function is

$$H(s)=\frac{G(s)}{1+G(s)}=\frac{KdKoF(s)}{s+KdKoF(s)}=\frac{KdKo}{s+KdKo} \quad (1)$$

the loop noise bandwidth is

$$Bl=\frac{KdKo}{4}=Ka\frac{KfKo}{4}=0.02Hz \quad (2)$$

Then, for the analog synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Km=1, A=1/2, B=1/2; Ko=2π)

$$(Ka.Km.A.B.Ko)/4 = 0.02 \rightarrow Ka=0.08*2/\pi \quad (3)$$

For the hybrid synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Km=1, A=1/2, B=0.45; Ko=2π)

$$(Ka.Km.A.B.Ko)/4 = 0.02 \rightarrow Ka=0.08*2.2/\pi \quad (4)$$

For the combinational synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Kf=1/π; Ko=2π)

$$(Ka*1/\pi*2\pi)/4 = 0.02 \rightarrow Ka=0.04 \quad (5)$$

For the sequential synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Kf=1/2π; Ko=2π)

$$(Ka*1/2\pi*2\pi)/4 =0.02 \rightarrow Ka=0.08 \quad (6)$$

The jitter depends on the RMS signal Aef, on the power spectral density No and on the loop noise bandwidth Bl.
For analog PLL the jitter is
$\sigma\phi^2=Bl.No/Aef^2=Bl.2.\sigma n^2.\Delta\tau=0.02*10^{-3}*2\sigma n^2/0.5^2=16*10^{-5}.\sigma n^2$
For the others PLLs the jitter formula is more complicated.

- 2nd order loop:

The second order loop is not shown here, but the results are identical to the ones obtained above for the first order loop.

B. Tests

The following figure (Fig.10) shows the setup that was used to test the various synchronizers.

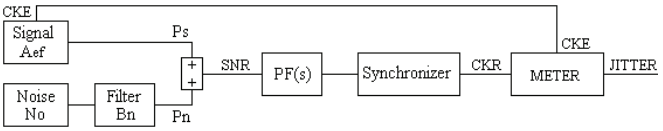


Fig.10 Block diagram of the test setup

The receiver recovered clock with jitter is compared with the emitter original clock without jitter, the difference is the jitter of the received clock.

C. Jitter measurer (Meter)

The jitter measurer (Meter) consists of a RS flip flop, which detects the random variable phase of the recovered clock (CKR), relatively to the fixed phase of the emitter clock (CKE). This relative random phase variation is the recovered clock jitter (Fig.11).

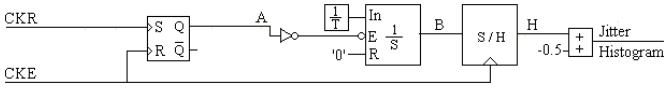


Fig.11 The jitter measurer (Meter)

The other blocks convert this random phase variation into a random amplitude variation, which is the jitter histogram. Then, the jitter histogram is sampled and processed by an appropriate program, providing the RMS jitter and the peak to peak jitter.

D. Results

We will present the results (graphics of output jitter UIRMS - input SNR) for the four symbol synchronizers.

Fig.12 shows the jitter-SNR curves of the four synchronizers namely both transitions at rate manual (b-m), both transitions at rate automatic (b-a), positive transitions at quarter rate manual (p-m/4) and positive transitions at quarter rate automatic (p-a/4).

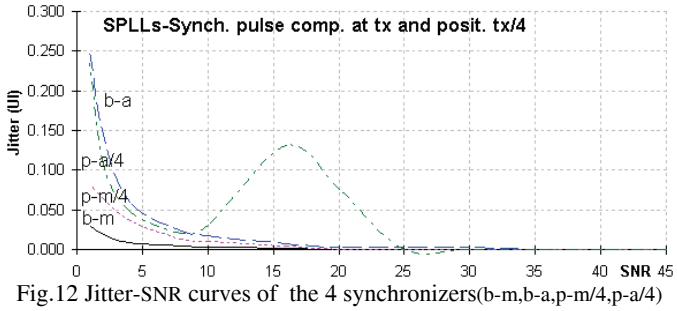


Fig.12 Jitter-SNR curves of the 4 synchronizers(b-m,b-a,p-m/4,p-a/4)

We see that, in general, the output jitter UIRMS decreases gradually with the input SNR increasing. However, the positive quarter rate automatic (p-a/4) has some irregularities.

For high SNR, the four synchronizer jitter curves tend to be similar. However, for low SNR, the manual versions (b-m, p-m/4) are significantly better than the automatic versions (b-a, p-a/4), the both transitions at rate manual (b-m) is slightly the best. Also, for an intermediate SNR (SNR ≅ 16), the positive transitions quarter rate automatic (p-a/4) has a very significant jitter perturbation, due to some losses of synchronism.

V. CONCLUSIONS

We studied four synchronizers with one variant operating by both transitions at the rate that has two versions namely the manual (b-m) and automatic (b-a) and other variant operating by positive transitions at quarter rate that has two versions namely the manual (p-m/4) and automatic (p-a/4). Then, we tested their jitter - noise curves.

We observed that, in general, the output jitter curves decreases gradually with the input SNR increasing. However, the positive quarter automatic (p-a/4) has some irregularities.

We verified that, for high SNR, the four synchronizers jitter curves tend to be similar, this is comprehensible since all the synchronizers are digital and have similar noise margin. However, for low SNR, the manual versions (b-m, p-m/4) are significantly better than the automatic versions (b-a, p-a/4), this is comprehensible since the automatic versions have more digital states, then the error state propagation is aggravated. The version both transitions at rate manual (b-m) is slightly the best because has less digital states. Also, for an intermediate SNR (SNR ≅ 16) the positive transitions at quarter rate automatic (p-a/4) has a very significant jitter perturbation due to some losses of synchronism.

In the future, we are planning to extend the present study to other types of synchronizers.

ACKNOWLEDGMENTS

The authors are grateful to the program FCT (Foundation for sCience and Technology) / POCI2010.

REFERENCES

[1] J. C. Imbeaux, “performance of the delay-line multiplier circuit for clock and carrier synchronization”, IEEE Jou. on Selected Areas in Communications p.82 Jan. 1983.

[2] Werner Rosenkranz, “Phase Locked Loops with limiter phase detectors in the presence of noise”, IEEE Trans. on Communications com-30 Nº10 pp.2297-2304. Oct 1982.

[3] H. H. Witte, “A Simple Clock Extraction Circuit Using a Self Sustaining Monostable Multivibrat. Output Signal”, Electronics Letters, Vol.19, Is.21, pp.897-898, Oct 1983.

[4] Charles R. Hogge, “A Self Correcting Clock Recovery Circuit”, IEEE Tran. Electron Devices p.2704 Dec 1985.

[5] A. D. Reis, J. F. Rocha, A. S. Gameiro, J. P. Carvalho “A New Technique to Measure the Jitter”, Proc. III Conf. on Telecommunications pp.64-67 FFoz-PT 23-24 Apr 2001.

[6] Marvin K. Simon, William C. Lindsey, “Tracking Performance of Symbol Synchronizers for Manchester Coded Data”, IEEE Transactions on Communications Vol. com-2.5 Nº4, pp.393-408, April 1977.

[7] J. Carruthers, D. Falconer, H. Sandler, L. Strawczynski, “Bit Synchronization in the Presence of Co-Channel Interference”, Proc. Conf. on Electrical and Computer Engineering pp.4.1.1-4.1.7, Ottawa-CA 3-6 Sep. 1990.

[8] Johannes Huber, W. Liu “Data-Aided Synchronization of Coherent CPM-Receivers” IEEE Transactions on Communications Vol.40 Nº1, pp.178-189, Jan. 1992.

[9] Antonio D’Amico, A. D’Andrea, Reggianni, “Efficient Non-Data-Aided Carrier and Clock Recovery for Satellite DVB at Very Low SNR”, IEEE Jou. on Sattelite Areas in Comm. Vol.19 Nº12 pp.2320-2330, Dec. 2001.

[10] Rostislav Dobkin, Ran Ginosar, Christos P. Sotiriou “Data Synchronization Issues in GALS SoCs”, Proc. 10th International Symposium on Asynchronous Circuits and Systems, pp.CD-Ed., Crete-Greece 19-23 Apr. 2004.

[11] N. Noels, H. Steendam, M. Moeneclaey, “Effectiveness Study of Code-Aided and Non-Code-Aided ML-Based Feedback Phase Synchronizers”, Proc. IEEE Int Conf. on Comm.(ICC’06) pp.2946-2951, Ist.-TK, 11-15 Jun 2006.

[12] A. D. Reis, J. F. Rocha, A. S. Gameiro, J. P. Carvalho “Effects of the Prefilter Type on Digital Symbol Synchronizers”, Proc. VII Symposium on Enabling Optical Network and Sensors (SEONs 2009) pp.35-36, Lisboa (Amadora)-PT 26-26 June 2009.

Prefilter Bandwidth Effects in Sequential Symbol Synchronizers based on Pulse Comparation by Positive Transitions at Quarter Rate

Antonio D. Reis[1,2], Jose F. Rocha[1], Atilio S. Gameiro[1] and Jose P. Carvalho[2]

***Abstract* – This work studies the effects of the prefilter bandwidth in the sequential symbol synchronizers based on pulse comparation at bit rate and at half bit rate. We consider three prefilter bandwidths namely B1=∞, B2=2.tx and B3=1.tx, where tx is the bit rate. The synchronizer has two variants one operating by both transitions at bit rate and other operating by positive transitions at quarter rate. Each variant has two versions namely the manual and the automatic. The objective is to study the prefilter bandwidth with four synchronizers and to evaluate their output jitter UIRMS (Unit Interval Root Mean Square) versus input SNR (Signal Noise Ratio).**

***Keywords* – Prefilter, Synchronism, Communication Systems**

I. INTRODUCTION

This work studies the prefilter bandwidth effects on the jitter-SNR behavior of four sequential symbol synchronizers.

The prefilter, applied before the synchronizer, switches their bandwidth between three values namely first B1=∞, after B2=2.tx and next B3=1.tx, where tx is the bit rate [1, 2, 3, 4]

The synchronizer has four types supported in two variants, one operating by both transitions at the rate with versions manual (b-m) and automatic (b-a) and other operating by positive transitions at quarter rate with versions manual (p-m/4) and automatic (p-a/4). The difference between the four synchronizers is only in the phase comparator, since the other blocks are equal. The synchronizer VCO (Voltage Controlled Oscillator) is the clock whose performance determines, in good part, the system quality [5, 6, 7, 8].

Fig.1 shows the prefilter followed of the synchronizer.

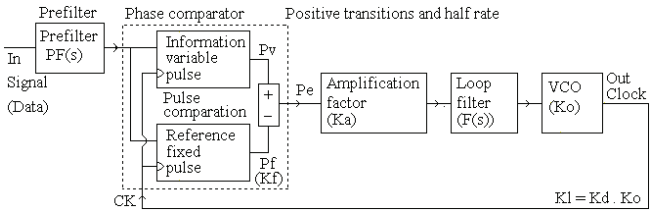


Fig.1 Prefilter with the synchronizers based on pulse comparation

PF(s) is the prefilter (low pass filter). The synchronizer has various blocks, namely Kf is the phase detector gain, F(s) is the loop filter, Ko is the VCO gain and Ka is the loop gain factor that controls the root locus and loop characteristics.

[1]Antonio D. Reis, Jose F. Rocha and Atilio S. Gameiro are with Dep. Electrónica, Universidade Aveiro / Instituto Telecomunicações, 3800 Aveiro, Portugal, E-mail: amg@det.ua.pt; frocha@det.ua.pt.

[2]Antonio D. Reis and Jose P. Carvalho are with the Dep. Fisica, Universidade da Beira Interior Covilhã / Unidade Detecção Remota, 6200 Covilhã, Portugal, E-mail: adreis@ubi.pt; pacheco@ubi.pt.

In priori and actual-art state was developed various synchronizers, but is necessary to know their performance.

The motivation of this work is to create new synchronizers and evaluate their performance with noise. This contribution increases the knowledge about synchronizers.

Following, we present the prefilter with their three different bandwidths (B1=∞, B2=2.tx, B3=1.tx) [9, 10, 11, 12].

After, we present the variant by both transitions at rate with their manual (b-m) and automatic (b-a) versions. Next, we present the variant by positive transitions at quarter rate with their manual (p-m/4) and automatic (p-a/4) versions.

After, we present the design and tests. Then, we present the results. Finally, we present the conclusions.

II. PREFILTER BANDWIDTH EFFECTS

The prefilter, applied before the synchronizer, filters the noise but disturbs slightly the signal. The prefilter bandwidth B switches between three values (B1=∞, B2=2.tx, B3=1.tx).

Fig.2 shows the prefilter with their three bandwidths.

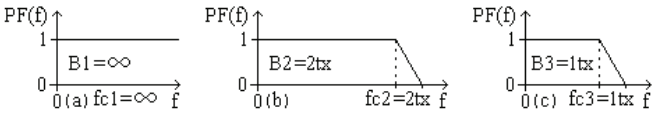


Fig.2 Three prefilter bandwidths: a) B1=∞; b) B2=2.tx; c) B3=1.tx

a) First, as shown in Fig.2a, the prefilter has a bandwidth equal to infinite (B1= ∞).

b) Second, as shown in Fig.2b, the prefilter has a bandwidth equal to times the bit rate (B2 = 2.tx).

c) Third, as shown in Fig.2c, the prefilter has a bandwidth equal to the bit rate (B3 = 1.tx).

We will evaluate the three bandwidth effects (B1, B2, B3) on the jitter-SNR curves of the four symbol synchronizers.

III. SYNCHRONIZERS OPERATING AT THE RATE

The synchronizer with its phase comparator operates, here, by both transitions at the data transmission rate.

This variant has the manual (b-m) and automatic (b-a) versions, the difference is in phase comparator. The variable pulse Pv, produced by the first flip flop with exor, is equal in the two versions, but the fixed pulse Pf is different [1, 2].

A. Both transitions, at the rate and manual

The manual version has a phase comparator, where the fixed pulse Pf is produced by an exor with a delay Δt=T/2, that needs a previous manual adjustment (Fig.3)

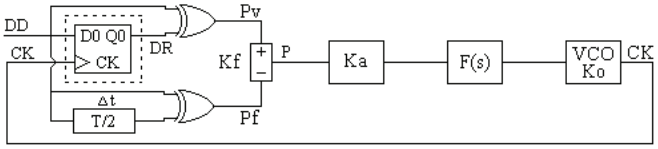


Fig.3 Synchronizer both at the rate and manual (b-m)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

B. Both transitions, at the rate and automatic

The automatic version has a phase comparator where the fixed pulse Pf is produced automatically by the second flip flop with exor, without previous adjustment (Fig.4).

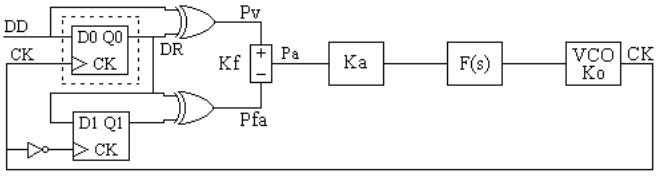


Fig.4 Synchronizer both at the rate and automatic (b-a)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

IV. SYNCHRONIZERS OPERATING AT QUARTER RATE

The synchronizer with its phase comparator operates, here, by positive transitions at quarter data transmission rate.

This variant has the manual (p-m/4) and the automatic (p-a/4) versions, the difference is only in the phase comparator. The variable pulse Pvp, based in the four first flip flops with multiplexer, is equal in the two versions, but the fixed pulse Pfp is produced from a different way [3, 4].

A. Positive transitions, at quarter rate and manual

The manual version has a phase comparator, where the fixed pulse Pf is produced by an exor with a delay Δt=T/2, that needs a previous manual adjustment (Fig.5).

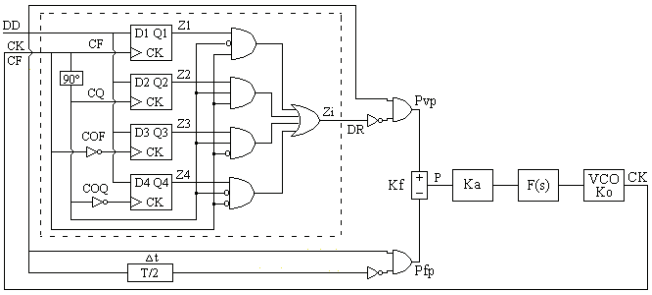


Fig.5 Synchronizer positive at quarter rate and manual (p-m/4)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

B. Positive transitions, quarter rate and automatic

The automatic version has a phase detector, where the fixed pulse Pf is produced automatically by the seconds flip flops and multiplexer with exor, without previous adjustment (Fig.6).

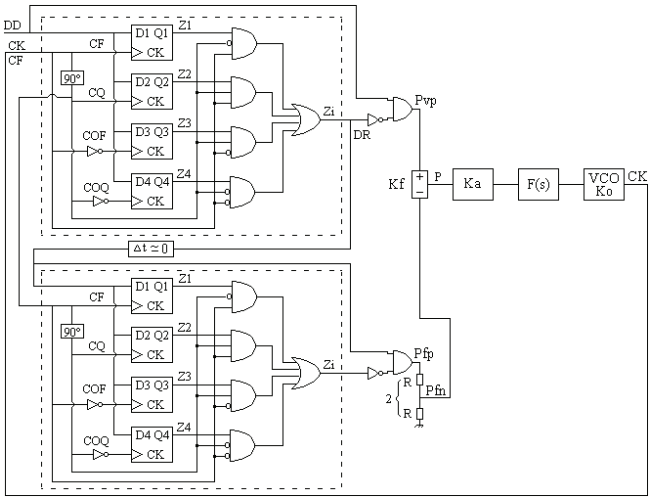


Fig.6 Synchronizer positive at quarter rate and automatic (p-a/4)

The variable pulse Pv minus the fixed pulse Pf (Pv-Pf) determines the error phase that controls the VCO.

V. DESIGN, TESTS AND RESULTS

We will present the design, the tests and the results of the referred synchronizers [5].

A. Design

To get guaranteed results, it is necessary to dimension all the synchronizers with equal conditions. Then it is necessary to design all the loops with identical linearized transfer functions.

The general loop gain is Kl=Kd.Ko=Ka.Kf.Ko where Kf is the phase comparator gain, Ko is the VCO gain and Ka is the control amplification factor that permits the desired characteristics.

For analysis facilities, we use a normalized transmission rate tx=1baud, what implies also normalized values for the others dependent parameters. So, the normalized clock frequency is fCK=1Hz.

We choose a normalized external noise bandwidth Bn = 5Hz and a normalized loop noise bandwidth Bl = 0.02Hz. Later, we can disnormalize these values to the appropriated transmission rate tx.

Now, we will apply a signal with noise ratio SNR given by the signal amplitude Aef, noise spectral density No and external noise bandwidth Bn, so the $SNR = A^2_{ef}/(No.Bn)$. But, No can be related with the noise variance σn and inverse sampling Δτ=1/Samp, then $No=2\sigma n^2.\Delta\tau$, so $SNR=A^2_{ef}/(2\sigma n^2.\Delta\tau.Bn) = 0.5^2/(2\sigma n^2*10^{-3}*5)= 25/\sigma n^2$.

After, we observe the output jitter UI as function of the input signal with noise SNR. The dimension of the loops is

- 1st order loop:

The loop filter F(s)=1 with cutoff frequency 0.5Hz (Bp=0.5 Hz is 25 times bigger than Bl=0.02Hz) eliminates only the high frequency, but maintain the loop characteristics.

The transfer function is

$$H(s)=\frac{G(s)}{1+G(s)}=\frac{KdKoF(s)}{s+KdKoF(s)}=\frac{KdKo}{s+KdKo} \quad (1)$$

the loop noise bandwidth is

$$Bl=\frac{KdKo}{4}=Ka\frac{KfKo}{4}=0.02Hz \quad (2)$$

Then, for the analog synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Km=1, A=1/2, B=1/2; Ko=2π)
$(Ka.Km.A.B.Ko)/4 = 0.02 \rightarrow Ka=0.08*2/\pi$ (3)

For the hybrid synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Km=1, A=1/2, B=0.45; Ko=2π)
$(Ka.Km.A.B.Ko)/4 = 0.02 \rightarrow Ka=0.08*2.2/\pi$ (4)

For the combinational synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Kf=1/π; Ko=2π)
$(Ka*1/\pi*2\pi)/4 = 0.02 \rightarrow Ka=0.04$ (5)
For the sequential synchronizers, the loop bandwidth is
$Bl=0.02=(Ka.Kf.Ko)/4$ with (Kf=1/2π; Ko=2π)
$(Ka*1/2\pi*2\pi)/4 =0.02 \rightarrow Ka=0.08$ (6)

The jitter depends on the RMS signal Aef, on the power spectral density No and on the loop noise bandwidth Bl.
For analog PLL the jitter is
$\sigma\phi^2=Bl.No/Aef^2=Bl.2.\sigma n^2.\Delta\tau=0.02*10^{-3}*2\sigma n^2/0.5^2=16*10^{-5}.\sigma n^2$
For the others PLLs the jitter formula is more complicated.

- 2nd order loop:

The second order loop is not shown here, but the results are identical to the ones obtained above for the first order loop.

B. Tests

The following figure (Fig.7) shows the setup that was used to test the various synchronizers.

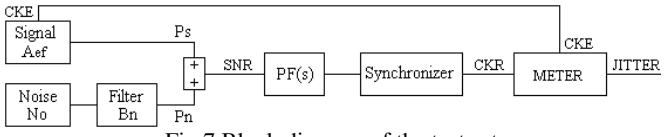


Fig.7 Block diagram of the test setup

The receiver recovered clock with jitter is compared with the emitter original clock without jitter, the difference is the jitter of the received clock.

C. Jitter measurer (Meter)

The jitter measurer (Meter) consists of a RS flip flop, which detects the random variable phase of the recovered clock (CKR), relatively to the fixed phase of the emitter clock (CKE). This relative random phase variation is the recovered clock jitter (Fig.8).

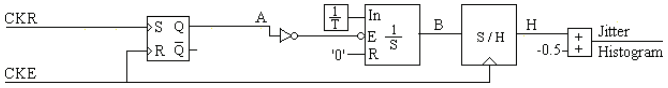


Fig.8 The jitter measurer (Meter)

The other blocks convert this random phase variation into a random amplitude variation, which is the jitter histogram.
Then, the jitter histogram is sampled and processed by an appropriate program, providing the RMS jitter and the peak to peak jitter.

D. Results

We will present the results (jitter - noise graphics) for the prefilter with the four synchronizers.
Fig.9 shows the jitter-SNR curves of the prefilter bandwidth B1=∞ with the four synchronizers namely both transitions at rate manual (b-m), both transitions at rate automatic (b-a), positive transitions at quarter rate manual (p-m/4) and positive transitions at quarter rate automatic (p-a/4).

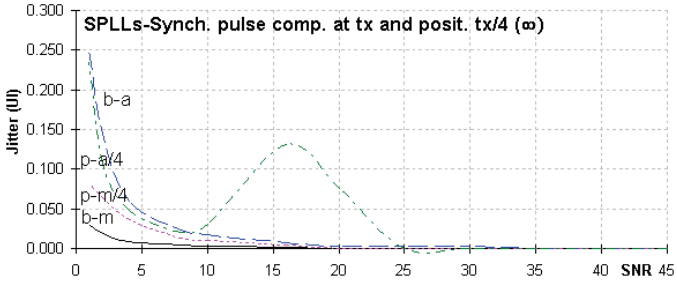


Fig.9 Jitter-SNR curves of B1+ 4 synchronizers(b-m,b-a,p-m/4,p-a/4)

We see that, in general, the output jitter UIRMS decreases gradually with the input SNR increasing. However, the positive at quarter rate (p-a/4) has some irregularities.
For prefilter B1=∞, for high SNR, the four synchronizer jitter curves tend to be similar. However, for low SNR, the manual versions (b-m, p-m/4) are significantly better than the automatic versions (b-a, p-a/4), the both transitions at rate manual (b-m) is slightly the best. Also, for an intermediate SNR (SNR ≅ 16), the positive transitions at quarter rate automatic (p-a/4) has a very significant jitter perturbation, due to some losses of synchronism.
Fig.10 shows the jitter-SNR curves of the prefilter bandwidth B2=2.tx with the four synchronizers namely both transitions at rate manual (b-m), both transitions at rate automatic (b-a), positive transitions at quarter rate manual (p-m/4) and positive transitions at quarter rate automatic (p-a/4).

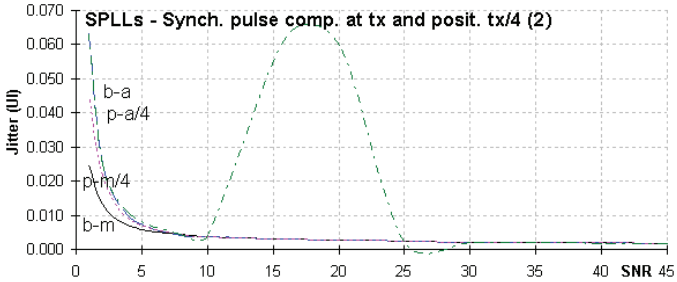


Fig.10 Jitter-SNR curves of B2+ 4 synchronizers(b-m,b-a,p-m/4,p-a/4)

For prefilter B2=2.tx, we verify that, it becomes the jitter-

SNR curves more similar between themselves. For high SNR, it harms slightly the jitter-SNR curves. However, for low SNR, it benefits significantly the jitter-SNR curves. Also, for an intermediate SNR (SNR ≅ 20), the positive transitions at quarter rate automatic (p-a/4) has a great jitter perturbation, due to some losses of synchronism.

Fig.11 shows the jitter-SNR curves of the prefilter bandwidth B3=1.tx with the four synchronizers namely both transitions at rate manual (b-m), both transitions at rate automatic (b-a), positive transitions at quarter rate manual (p-m/4) and positive transitions at quarter rate automatic (p-a/4).

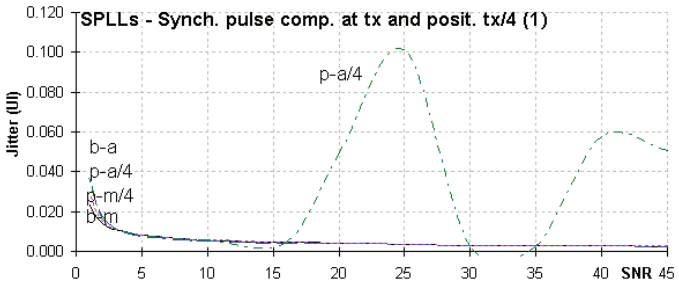


Fig.11 Jitter-SNR curves of B3+ 4 synchronizers(b-m,b-a,p-m/4,p-a/4)

For prefilter B3=1.tx, we verify that, it becomes the jitter-SNR curves still more similar between themselves. For high SNR, it harms more the jitter-SNR curves. However, for low SNR, it benefits less the jitter-SNR curves. Also, for an intermediate SNR (SNR ≅ 25), the positive transitions at quarter rate automatic (p-a/4) has a great jitter perturbation, due to some losses of synchronism.

VI. CONCLUSIONS

We studied three prefilter bandwidths (B1=∞, B2=2.tx, B3=1.tx) with four synchronizers, one variant operates by both transitions at the rate with two versions namely the manual (b-m) and automatic (b-a) and other variant operates by positive transitions at quarter rate with two versions namely the manual (p-m/4) and automatic (p-a/4). Then, we tested their jitter - SNR curves.

We observed that, in general, the output jitter curves decreases gradually with the input SNR increasing. However, the positive transitions at quarter rate (p-a/4) has some undesired irregularities.

For prefilter B1=∞, we verified that, for high SNR, the four synchronizers jitter curves tend to be similar, this is comprehensible since all the synchronizers are digital and have similar noise margin. However, for low SNR, the manual versions (b-m, p-m/4) are significantly better than the automatic versions (b-a, p-a/4), this is comprehensible since the automatic versions have more digital states, then the error state propagation is aggravated. The version both transitions at rate manual (b-m) is slightly the best, because has less digital states. Also, for an intermediate SNR (SNR ≅ 16) the positive transitions at quarter rate automatic (p-a/4) has a very significant jitter perturbation due to some losses of synchronism.

For prefilter B2=2.tx, we verify that, it becomes the jitter-SNR curves more similar between themselves. For high SNR, it harms slightly the jitter-SNR curves. However, for low SNR, it benefits significantly the jitter-SNR curves. Also, for an intermediate SNR (SNR ≅20), the synchronizer (p-a/4) has a great jitter perturbation, due to some losses of synchronism.

For prefilter B3=1.tx, we verify that, it becomes the jitter-SNR curves still more similar between themselves. For high SNR, it harms more the jitter-SNR curves. However, for low SNR, it benefits less the jitter-SNR curves. Also, for an intermediate SNR (SNR ≅ 25), the synchronizer (p-a/4) has a great jitter perturbation, due to some losses of synchronism.

In the future, we are planning to extend the present study to other types of synchronizers.

ACKNOWLEDGMENTS

The authors are grateful to the program FCT (Foundation for sCience and Technology) / POCI2010.

REFERENCES

[1] J. C. Imbeaux, “performance of the delay-line multiplier circuit for clock and carrier synchronization”, IEEE Jou. on Selected Areas in Communications p.82 Jan. 1983.

[2] Werner Rosenkranz, “Phase Locked Loops with limiter phase detectors in the presence of noise”, IEEE Trans. on Communications com-30 Nº10 pp.2297-2304. Oct 1982.

[3] H. H. Witte, “A Simple Clock Extraction Circuit Using a Self Sustaining Monostable Multivibrat. Output Signal”, Electronics Letters, Vol.19, Is.21, pp.897-898, Oct 1983.

[4] Charles R. Hogge, “A Self Correcting Clock Recovery Circuit”, IEEE Tran. Electron Devices p.2704 Dec 1985.

[5] A. D. Reis, J. F. Rocha, A. S. Gameiro, J. P. Carvalho “A New Technique to Measure the Jitter”, Proc. III Conf. on Telecommunications pp.64-67 FFoz-PT 23-24 Apr 2001.

[6] Marvin K. Simon, William C. Lindsey, “Tracking Performance of Symbol Synchronizers for Manchester Coded Data”, IEEE Transactions on Communications Vol. com-2.5 Nº4, pp.393-408, April 1977.

[7] J. Carruthers, D. Falconer, H. Sandler, L. Strawczynski, “Bit Synchronization in the Presence of Co-Channel Interference”, Proc. Conf. on Electrical and Computer Engineering pp.4.1.1-4.1.7, Ottawa-CA 3-6 Sep. 1990.

[8] Johannes Huber, W. Liu “Data-Aided Synchronization of Coherent CPM-Receivers” IEEE Transactions on Communications Vol.40 Nº1, pp.178-189, Jan. 1992.

[9] Antonio D’Amico, A. D’Andrea, Reggianni, “Efficient Non-Data-Aided Carrier and Clock Recovery for Satellite DVB at Very Low SNR”, IEEE Jou. on Sattelite Areas in Comm. Vol.19 Nº12 pp.2320-2330, Dec. 2001.

[10] Rostislav Dobkin, Ran Ginosar, Christos P. Sotiriou “Data Synchronization Issues in GALS SoCs”, Proc. 10th International Symposium on Asynchronous Circuits and Systems, pp.CD-Ed., Crete-Greece 19-23 Apr. 2004.

[11] N. Noels, H. Steendam, M. Moeneclaey, “Effectiveness Study of Code-Aided and Non-Code-Aided ML-Based Feedback Phase Synchronizers”, Proc. IEEE Int Conf. on Comm.(ICC’06) pp.2946-2951, Ist.-TK, 11-15 Jun 2006.

[12] A. D. Reis, J. F. Rocha, A. S. Gameiro, J. P. Carvalho “Effects of the Prefilter Type on Digital Symbol Synchronizers”, Proc. VII Symposium on Enabling Optical Network and Sensors (SEONs 2009) pp.35-36, Lisboa(Amadora)-PT 26-26 June 2009.

Throughput Improvement in Gigabit DSL Communications

Vladimir Poulkov[1], Miglen Ovcharov[2] and Pavlina Koleva[3]

Abstract – **In this paper we propose a scheme for throughput improvement in Gigabit Digital Subscriber Line (GDSL) systems by Radio Frequency Interference (RFI) mitigation, using an Identification and Cancellation algorithm. The research shows that the GDSL link performance depends strongly on the parameters of the Multiple Input Multiple Output (MIMO) DSL system as well as on the RFI suppression algorithm efficiency.**

Keywords – **Gigabit DSL, GDSL, DMT, RFI, MIMO**

I. Introduction

The rapid growth of communication technologies, digital signal processing (DSP) and computational power of microprocessors have made possible today's Gigabit DSL communications, offering throughput of several Gigabits, using more than one twisted pairs in the telephone cables.

The multi-pair MIMO DSL vectored communication technology was originally proposed by Cioffi and Ginis [1, 2] in 2002. The method adopts common DSLAM equipment for all service providers in the Central Office, synchronised block transmission and Dynamic Spectrum Management algorithms. The theoretical studies prove that DSL loops can achieve data rates as high as 10 Gbps at 500 m using four twisted pair telephone cable, assuming DSLAM at both sides.

The Block Diagram of a GDSL communication system is presented on Figure 1.

Along with the methods typical for vectored DSL technology, GDSL benefits from Common Mode (CM) transmission due to its much lower attenuation. On the other side, in CM, the crosstalk disturbers are much stronger than in Differential Mode (DM). This phenomenon is utilized for transfer of additional information by applying appropriate precoding methods.

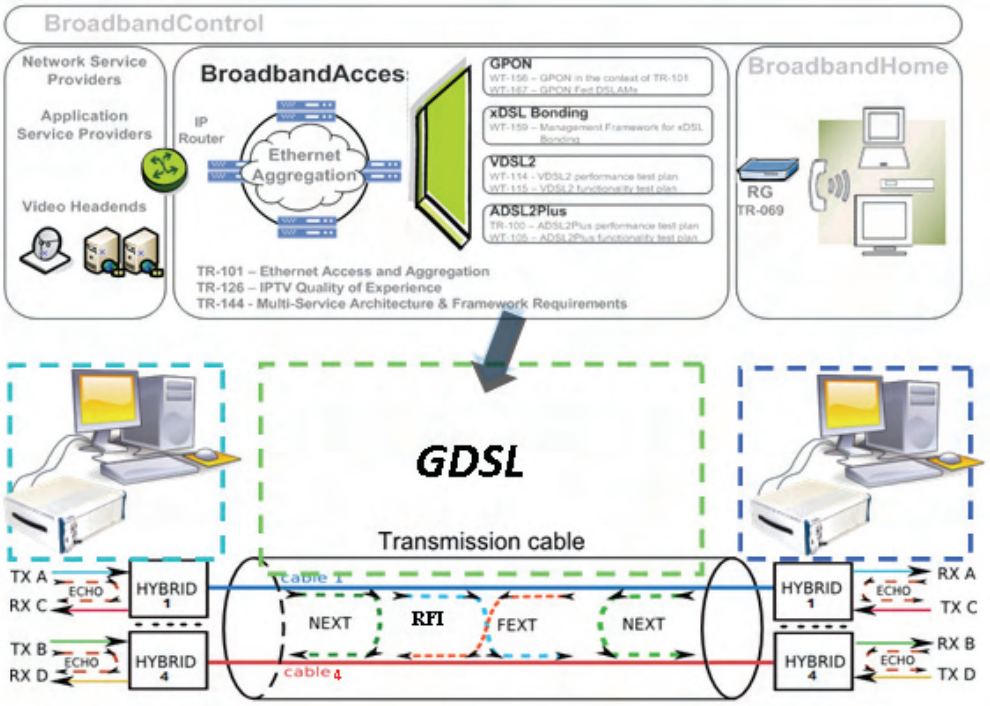


Fig. 1. Block diagram of MIMO DMT GDSL communication system

The transfer function of a MIMO DMT GDSL communication system can be described by the matrix equation [3]:

$$Y = HX + N \quad (1)$$

Where: **Y** is the output signal vector which components are the output signals of each twisted pair; **X** is the input signal vector which components are the inputs signals of each

[1]Vladimir Poulkov is with the Faculty of Telecommunications, Technical University of Sofia, Sofia 1000, 8 Kl. Ohridski Blvd., Bulgaria, e-mail: vkp@tu-sofia.bg

[2]Miglen Ovtcharov is with the Faculty of Telecommunications, Technical University of Sofia, Sofia 1000, 8 Kl. Ohridski Blvd, Bulgaria, e-mail: miglen.ovcharov@jci.com

[3]Pavlina Koleva is with the Faculty of Telecommunications, Technical University of Sofia, Sofia 1000, 8 Kl. Ohridski Blvd, Bulgaria, e-mail: pkoleva@tu-sofia.bg

twisted pair; **H** is the transfer matrix of the MIMO DSL channel; **N** is the additive noise vector at the input of MIMO DSL channel, including AWGN, RFI, impulse noises. The crosstalk function between different wires, is included in the channel matrix **H**. Equation (1) is valid for each tone of the DMT symbol, therefore the tone index will be not be included in the analysis below.

In order to perform successful demodulation at the receiver, the channel transform matrix **H** of MIMO DSL channel must be known. Because of the physical nature of the cable cannel, it's assumed that the channel parameters are constant in time or vary slowly as a function of outside temperature. Therefore the MIMO cable channel parameters can be identified at the start-up initialization procedure.

GDSL DMT technology uses mainly the Common Mode transmission as shown on Figure 2 [2].

The matrix impedances $\mathbf{Z_L}$ and $\mathbf{Z_S}$ can have two main

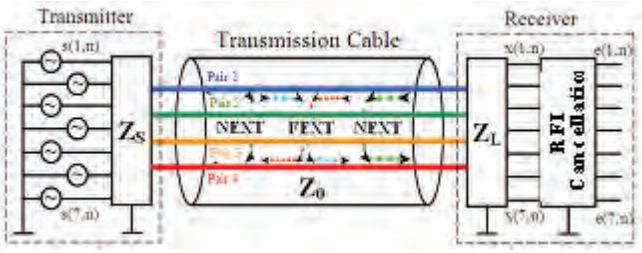


Fig. 2. MIMO GDSL Common Mode System Model [2]

configurations: either both matrix impedances $\mathbf{Z_L}$ and $\mathbf{Z_S}$ to match the characteristic impedance matrix $\mathbf{Z_0}$ or the diagonal elements of $\mathbf{Z_L}$ and $\mathbf{Z_S}$ to match the diagonal elements of $\mathbf{Z_0}$.

The channel matrix **H** contains a frequency dependent transform function of the pair or wire at its main diagonal $T_{i,i}$ and a frequency dependent crosstalk transform function between pairs or wires. The channel matrix for Common Mode and Differential Mode in case of four twisted pairs can be written as:

$$H_{cm} = \begin{bmatrix} T_1 & FEXT_{12} & \Lambda & FEXT_{16} & FEXT_{17} \\ FEXT_{21} & T_2 & \Lambda & FEXT_{26} & FEXT_{27} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{M} \\ FEXT_{61} & FEXT_{62} & \Lambda & T_6 & FEXT_{67} \\ FEXT_{71} & FEXT_{72} & \Lambda & FEXT_{76} & T_{77} \end{bmatrix} \quad (2)$$

Using Singular Value Decomposition (SVD) method, the cannel matrix **H** can be presented as [3]:

$$H(f) = U(f)S(f)V^*(f) \quad (3)$$

$$S(f) = diag\ (\lambda_1, \lambda_2, ..., \lambda_L) \quad (4)$$

Where: **U** and **V** are unitary matrixes and $\mathbf{UU^{-1}} = \mathbf{U^{-1}U} = \mathbf{VV^{-1}} = \mathbf{V^{-1}V} = \mathbf{I}$; $\mathbf{U^{-1}} = \mathbf{U^*}$; $\mathbf{V^{-1}} = \mathbf{V^*}$; L – number of twisted pairs; f – linear frequency and $f = n\Delta f$ for $n = 1,2,..., N_{sc}$; Δf – sub-tones bandwidth; λ_i – singular values of channel matrix per each tone. From equations (1-4) the input signal of the MIMO DMT receiver can be written as:

$$Y = U\ S\ V^*X + N \quad (5)$$

Multiplying (5) by $\mathbf{U^*}$ at left side, the received signal at the modem inputs is:

$$\tilde{Y} = S\ \tilde{X} + \tilde{N} \quad (6)$$

$$\tilde{y}_l = \lambda_l \tilde{x}_l + \tilde{n}_l, \qquad l = 1,2,..., L \quad (7)$$

$$\tilde{Y} = U^*Y; \quad \tilde{X} = V^*X; \quad \tilde{N} = U^*N \quad (8)$$

Equation (6) describes the MIMO DMT DSL system as a set of parallel scalar independent SISO DMT DSL communication systems. Consequently, the complexity of MIMO DMT GDSL system will be proportional to the complexity of L independent DMT DSL communication systems plus a little bit overhead due to the supervisor control algorithm and calculations required for matrixes: $\tilde{Y}$, $\tilde{X}$, $\tilde{N}$.

The Block Diagram in frequency domain of a single DMT DSL communication channel is presented on Figure 3 [2].

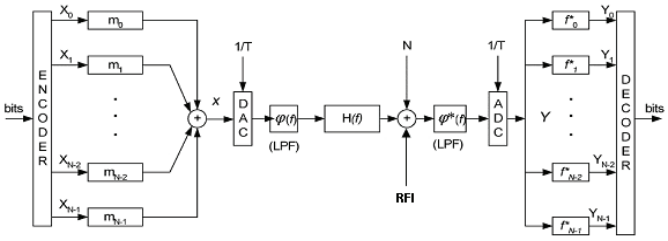


Fig. 3. Single scalar DMT DSL communication channel [2]

II. RFI CANCELLATION

The frequency identification and cancelation method is implemented as two phases algorithm. In the beginning of the first phase, the complex RFI frequency is estimated by finding maximum in oversampled signal spectrum [4]:

$$\omega_n = \arg(\omega) \in L_\infty \max(P_R(\omega)) \quad (9)$$

After this, the amplitude and phase estimation is done. The received sampled signal $r(k)$ can be expressed as:

$$r(k) = s(k) + n(k) + i_n(k) \quad (10)$$

The interference signal is defined as [4]:

$$i_n(k) = a_n \cos(\omega_n k / T) + jb_n \sin(\omega_n k / T) \quad (11)$$

On the basis of (11) a matrix form to represent the sampled interference can be used. In particular, it can be rewritten in matrix form as [4]:

$$I_n = MX \quad (12)$$

Where matrix $\boldsymbol{M}$ is defined as:

$$M = \begin{bmatrix} \cos(\omega_n k_1 / T) & j\sin(\omega_n k_1 / T) \\ \cos(\omega_n k_2 / T) & j\sin(\omega_n k_2 / T) \\ ..... & ..... \\ \cos(\omega_n k_N / T) & j\sin(\omega_n k_N / T) \end{bmatrix} \quad (13)$$

Vector $\boldsymbol{X}$ gathers the coefficients a_n and b_n:

$$X = \begin{bmatrix} a_n & b_n \end{bmatrix}^T \quad (14)$$

Applying Maximum Likelihood (ML) algorithm, the solution is given by [4]:

$$X = \left(M^T M\right)^{-1} M^T R \quad (15)$$

Where the input signal vector $\boldsymbol{R}$ is defined as:

$$R = \begin{bmatrix} r(k_1) & r(k_2) & .... & r(k_N) \end{bmatrix}^T \quad (16)$$

Thus the information about the complex amplitude of the RFI tone is derived.

The second phase of the algorithm uses the estimation results from the first phase, as initial conditions of NLS optimization procedure, applied to each scalar channel [4]:

$$f(\omega, A, \varphi) = \sum_{i=1}^{N} \left| r(t) - \sum_{m=1}^{M} A_m e^{j(\omega_m t + \varphi_m)} \right|^2 \quad (17)$$

Here ω_m is the frequency of the *m*-th interfering tone and A_m and φ_m respectively are the amplitude and phase of the *m*-th interfering sinusoidal tone and *M* is the total number of interfering tones. The sum from 1 to *N* is referred to the *N* samples representing the sampled monocycle.

III. Simulation Model

To evaluate the performance of GDSL link, MATLAB simulations relative to the complex baseband presentation are conducted, assuming a typical MIMO DMT DSL receiver. The Channel Encoder is implemented as Convolutional Encoder with code rate: $R_C = 1/2$.The Interleaver – Deinterleaver is modeled as a permutation table random access algorithm. The Optimal Linear Precoder for complex signals is realized using the methods described in [2]. For the simulations, the number of used cable pairs is: $N = [1,2,3,4]$ and the corresponding number of MIMO inputs/outputs is: $K = 2N - 1$.The Digital Modulator is implemented as 8192-point Inverse Fast Fourier Transform (IFFT) process. The DMT symbol consists of up to 4096 tones. Each DMT data tone can use a different Grey encoded QAM modulation format which depends on the SNR of the data tone bandwidth. After the IFFT process, the prefix and suffix guard intervals are added.

The MIMO digital wire-line subscriber channel model is realized as given in [6].

$$y = H\,x + n + z \quad (18)$$

Where: **y** is a *[K×1]* output column vector whose components are the outputs of the individual transmission lines and **x** is a *[K×1]* input column vector. In DM, the MIMO channel *[K×K]* transfer matrix **H** (typically constant or varying slowly with temperature) models the DSL cable test loop for common mode excitation, using the "ABCD" parameters block matrix [6, 7], **n** is a *[K×1]* Complex Additive White Gaussian Noise (AWGN) column vector and **z** is a *[K×1]* RFI column vector.

For the experiments, a 24-gauge cable segment (AWG 24) 300 m long, is chosen. The background noise with Power Spectrum Density (PSD) at -140 dBm/Hz is modeled as complex AWGN - *Nc(0,1)*.

By the time being, there is no globally accepted model for Common Mode transmission in twisted pair cables. In the proposed simulation model, the method for CM cable transfer function determination is realized by substituting the DM primary cable parameters [10, 11]:

$$R_C(f) = 0{,}55\,R(f), \quad L_C(f) = 4{,}4\,L(f) \quad (19)$$

$$G_C(f) = 2\,G(f), \quad C_C(f) = 0{,}95\,C(f) \quad (20)$$

Where: *R(f), L(f), G(f)* and *C(f)* are the frequency dependent primary DSL cable parameters in DM; $R_C(f)$, $L_C(f)$, $G_C(f)$ and $C_C(f)$ are frequency dependent primary DSL cable parameters in CM. Then, the CM MIMO DSL channel transform function can be written in a form [10]:

$$H_{CM} = H_{CM}^{D}\left(I + H_{CM\ NEXT} + H_{CM\ FEXT}\right) \quad (21)$$

$$H_{CM\,NEXT_{m,n}}(f,L) = K_{NEXT} f^{3/2}\left(1 - C_{CM\,NEXT_{m,n}}(f,L)^4\right) \quad (22)$$

$$H_{CM\,FEXT_{m,n}}(f,L) = K_{FEXT} f^{2}\left(1 - C_{CM\,FEXT_{m,n}}(f,L)^2\right) \quad (23)$$

Where: $\mathbf{H^D_{CM}}$ is a *[2K-1×2K-1]* diagonal matrix that contains the end to end complex line attenuations of *2K-1* channels in CM, **I** is a *[2K-1×2K-1]* identity matrix, $\mathbf{C_{CM\ NEXT}}$ and $\mathbf{C_{CM\ FEXT}}$ are *[2K-1×2K-1]* "zero-diagonal" matrixes that reflect the NEXT and FEXT coupling in CM, *L* - the loop length and *f* - frequency. K_{NEXT} and K_{FEXT} are power normalization constants.

In the DMT Demodulator the guard prefix and suffix intervals are removed and 8192-point FFT is applied. Further, Frequency Domain Per-Tone Adaptive Channel Equalization and DMT demodulation are performed. Finally, OLP, 64-QAM demodulation and Error Correction Decoding are implemented.

IV. Simulation Results

Using the above general simulation model of a MIMO Gigabit DSL system, different experiments have been performed, estimating the Bit Error Ratio (BER) and Throughput as a function of the Signal to Noise Ratio (SIR) and RFI presence. The RFI is modeled as a complex single tone, the frequency of which is located in the middle between two adjacent DMT tones. In respect of the number of used twisted pairs, four DSL systems are considered: Single Input Single Output (SISO) VDSL2 (1-pair), MIMO GDSL: 2, 3 and 4-pair. The DSL channels are subject to FEXT, NEXT and background AWGN with flat variable PSD.

In Figure 4 the complex MIMO DSL channel 24 AWG, 300 m long, without RFI suppression is considered. The background AWGN with Power Spectrum Density (PSD) at -140 dBm/Hz is considered. Unsurprisingly, 4-pair MIMO GDSL system achieves the best performance.

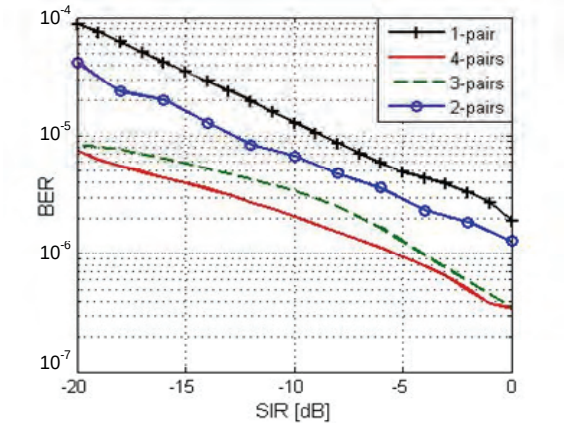


Fig. 4. BER as a function of SIR for GDSL

Figure 5 presents the relation between BER, Throughput and SNR for 4x4 MIMO GDSL DMT single user

communication system: (a) – without RFI cancellation; (b) – with RFI cancellation.

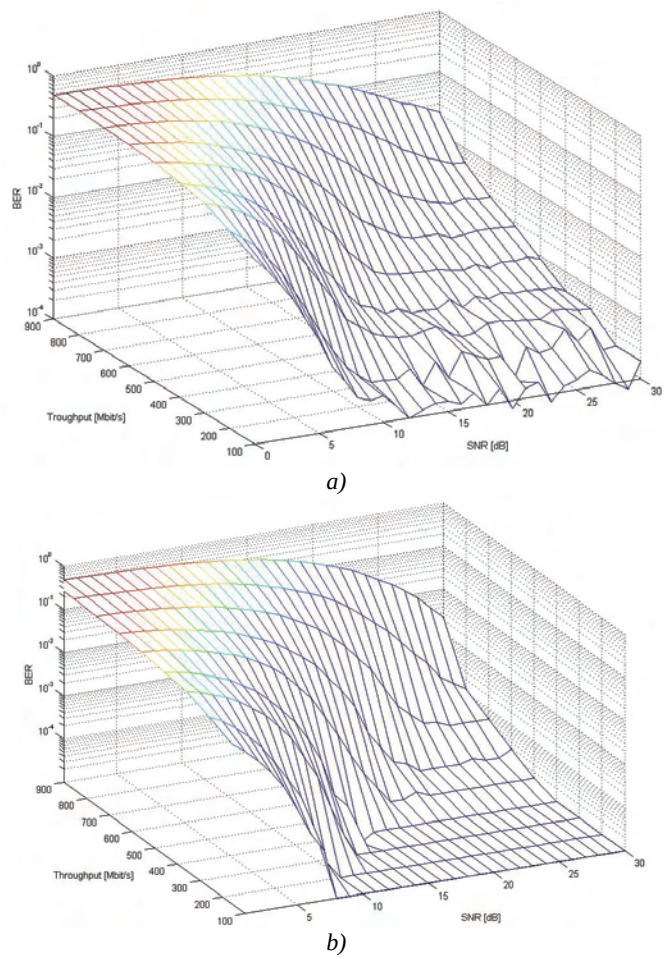


Fig. 5. Relation between BER, Throughput and SNR for 4x4 MIMO GDSL DMT single user communication system for: SIR = -10dB, Cable 24 AWG, L = 300 m.

The GDSL system throughput as a function of SNR having:

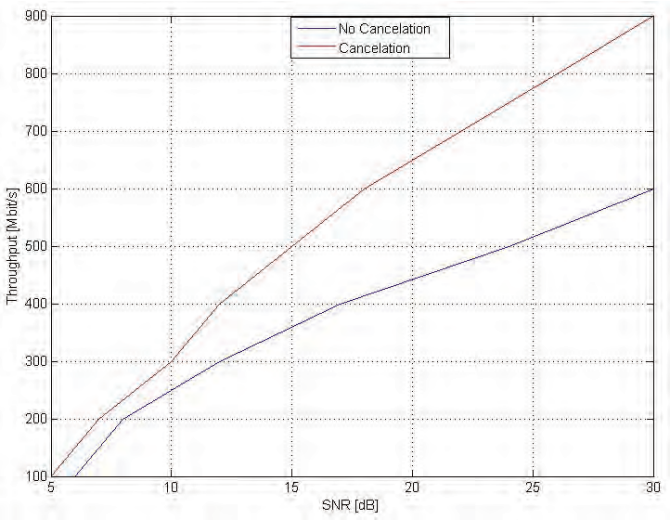


Fig. 6. Throughput as a function of SNR for 4x4 MIMO GDSL DMT single user communication system for: SIR = -10 dB, BER = 10^{-3}, Cable 24 AWG, L = 300 m.

RFI level fixed to SIR = -10 dB and BER fixed to BER = 10^{-3} is plotted in Figure 6. From the figure can be concluded that applying such a method of RFI cancellation leads to significant improvement in the throughput performance, up to a level where the system transmission speed is limited by the background AWGN.

V. Conclusions

In this paper a method for throughput improvement in MIMO Gigabit DSL DMT communication systems, using RFI filtering via frequency identification and cancellation, is proposed. Analyzed are the relations between BER and Throughput as a function of SNR for two cases: with and without RFI cancellation. The experimental results bring forward the significant improvement of GDSL link performance when applying such a RFI cancellation.

References

[1] G.Ginis and J.M. Cioffi, "Vectored transmission for digital subscriber line systems", *IEEE J. on Sel. Areas Commun.*, vol. 20, no. 5, pp. 1085–1104, June 2002.

[2] Bin Lee, John M. Cioffi, Sumanth Jagannathan, Mehdi Mohseni, "Gigabit DSL", IEEE TRANS ON COMMS, VOL. 55, NO. 9, SEPTEMBER 2007, pp. 1689-1692.

[3] ANSI Dynamic Spec Rep Draft, T1E1.4/2003-RB, May 2004.

[4] E. Baccareli, M. Baggi, L. Tagilione, "A Novel Approach to In-Band Interference Mitigation in Ultra Wide Band Radio Systems", IEEE Conference on Ultra Wide Band Systems and Technologies, 2002.

[5] S. Tezuka, P. L'Ecuyer, R. Couture, "On the lattice structure of the add-with-carry and subtract-with-borrow random number generators", *TOMACS*, vol. 3 , No 4, pp. 315 – 331, Oct. 1993.

[6] D. Zanatta Filho at all, "The Capacity of Binders for MIMO Digital Subscriber Lines", *VI International Telecommunications Symposium (ITS2006)*, Sep 3-6, 2006, Fortaleza-CE, Brazil.

[7] Clayton R. P., "*Analysis of Multi-conductor Transmission Lines*", Wiley and Sons, 1994.

[8] Milos Jacovljevic at all, "Common Mode Characterization and Channel Model Verification for Shielded Twisted Pair (STP) Cable", expects for publication in the *ICC2008 Proc. of IEEE.*

[9] Le Nir V., Moonen M., ``Study and optimisation of the common mode exploitation for xDSL application", Internal Report 07-161, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2007.

[10] A. H. Kamkar-Parsi, M. Bouchard, G. Bessens, and T. H. Yeap, "A wideband crosstalk canceller for xdsl using common-mode information", *IEEE Transactions on Communications*, vol. 53, no. 2, pp. 238–242, Feb. 2005.

Performance Analysis of SC Diversity System with Different Correlation Models in α-μ fading Environment

Jelena Anastasov[1], Mihajlo Stefanović[2], Stefan Panić[3], Ivana Petrović[4] and Edis Mekić[5]

***Abstract* – In this paper, we present the performance analysis of selection combining (SC) diversity system operating over the channel subjected to α-μ fading and cochannel interference. Assuming the short distance between the received antennas, we consider the existed correlation between received signal envelopes. Constant correlation and exponential correlation models are analysed for proposed channel and system model and brief analytical and numerical results are given for both cases. Outage probability, as an important system's performance measure, is the main focus of the paper.**

***Keywords* – fading environment, outage probability, selection diversity**

I. INTRODUCTION

Many urban mobile communication systems are subjected to fading, caused by multipath propagation due to reflections, refractions and scattering by buildings and other large obstucles [1]. As a result, the receiver sees the superposition of multiple copies of the transmitted signal, each traversing a different path. Several statistical models have been used in the literature to describe the fading envelope of the received signal [2]-[4]. For example, the Rayleigh and Rician distributions are used to characterize the envelope of faded signals over small geographical areas or short term fades while the log-normal distribution is used when much wider geographical areas are involved. Nakagami-*m* distribution is frequently used fading model distribution because of its wide range of applicability for many physical propagation channels than other distributions. Further, more general fading distribution is the α-μ distribution [5]. It includes all other distributions: the Gamma (and its discrete versions Erlang and central Chi-squared), Nakagami-*m* (and its discrete version Chi), exponential, Weibull, one-sided Gaussian and Rayleigh.

There are many proposed techniques to reduce bad fading influences. Space diversity techniques are extensively used to ameliorate signal propagation in fading channels and to improve the system performance gain. Maximal-ratio combining (MRC), equal-gain combining (EGC), selection combining (SC), switch and stay combining (SSC), or a combination of MRC and SC called generalized selection combining (GSC), are several principal types of diversity combining techniques that can be generally performed [6]. SC space-diversity technique is with lower complexity nature opposed to MRC and EGC combining techniques which require the amount of the channel-state information of transmitted signal.

In general, the SC combiner selects the branch with the highest signal-to-noise ratio (SNR), actually the branch with the strongest signal assuming equal noise power among the branches [3]. Also, it has been proposed the selection of the branch with the highest signal plus noise [7]. In environments where the level of the cochannel interference is sufficiently high as compared to the thermal noise (interference-limited systems), SC picks the highest signal-to-interference ratio (SIR; SIR-based selection diversity) [8].

The fading among the channels is correlated due to insufficient antenna spacing, which is a real scenario in practical diversity systems, resulting in a degradation of the diversity gain [1]. Therefore, it is important to understand how the correlation between received signals affects the system performance. Several correlation models have been proposed in the literature for the performance analysis of various wireless systems, corresponding to specific modulation, detection, and diversity combining schemes. Though useful in mathematics and some situations in engineering, the assumption of constant or exponential correlation generally matches the practical environment in mobile communications. When there is an antenna array with a totally symmetrical triangular configuration, we have a constant correlation. On the contrary, the assumption of exponential correlation is somewhat close to the situation in a linear array, but it requires equispaced diversity antennas.

II. CHANNEL AND SYSTEM MODEL

The desired signal received by the *i*-th antenna can be written as [9]:

[1]Jelena Anastasov is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18 000 Niš, Serbia, E-mail: anastasovjelena@gmail.com

[2]Mihajlo Stefanović is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18 000 Niš, Serbia, E-mail: mihajlo.stefanovic@elfak.ni.ac.rs

[3]Stefan Panić is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18 000 Niš, Serbia, E-mail: stefanpnc@yahoo.com

[4]Ivana Petrović is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18 000 Niš, Serbia

[5]Edis Mekić is with the Faculty of Electronic Engineering, A. Medvedeva 14, 18 000 Niš, Serbia

$$D_i(t) = R_i(t) e^{j\phi_i(t)} e^{j[2\pi f_c t + \Phi(t)]}, i = 1,2 \quad (1)$$

where f_c is carrier frequency, $\Phi(t)$ desired information signal, $\phi_i(t)$ the random phase uniformly distributed and $R_i(t)$ the α-μ distributed random amplitude process given by [5]:

$$f_{R_i}(t) = \frac{\alpha_i \mu_i^{\mu_i} t^{\alpha_i \mu_i - 1}}{\hat{t}^{\alpha_i \mu_i} \Gamma(\mu_i)} \exp\left(-\mu_i \frac{t^{\alpha_i}}{\hat{t}_i^{\alpha_i}}\right), t \geq 0 \quad (2)$$

where $\Gamma(.)$ is the Gamma function [10, eq. (8.31)], with $\hat{t} = \sqrt[\alpha_i]{E(R_i^{\alpha_i})}$ being the α -root mean value, μ is the inverse of normalized variance of R^α i.e. $\mu = E^2(R^\alpha)/ V(R^\alpha)$, $\mu > 0$, and $E(\cdot)$ and $V(\cdot)$ are, respectively, the expectation and variance operators.

The resultant interfering signal received by the i-th antenna is:

$$C_i(t) = r_i(t) e^{j\theta_i(t)} e^{j[2\pi f_c t + \psi(t)]}, i = 1,2 \quad (3)$$

where $r_i(t)$ is also α-μ distributed random amplitude process, $\theta_i(t)$ is the random phase, and $\psi(t)$ is the information signal. This model refers to the case of a single cochannel interferer.

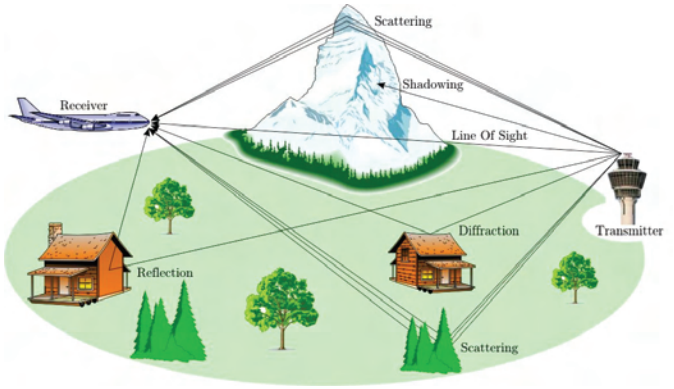


Fig. 1. Fading propagation environment

The performance of the SC can be carried out by considering the effect of only the strongest interferer, assuming that the remaining interferers are combined and considered as uncorrelated interference between antennas, as in [11]. Furthermore, $R_i(t)$, $r_i(t)$, $\phi_i(t)$, and $\theta_i(t)$ are assumed to be mutually independent and the effect of the interference on system performance is higher than that of the thermal noise (interference-limited environment) [9]. Now, due to a scenario with closely placed diversity antennas, both desired and interfering signal envelopes experience correlative α–μ fading.

There is a need to derive the joint statistics for two α-μ variables. We are relaying on some results which are already available in the literature for the constant correlation model of Nakagami-m distribution [12, eq. (10)] and exponential correlation model of Nakagami-m distribution [13, eq. (3)]. We also consider the proposed relations between α-μ and Nakagami-m variables [5, eqs. (19),(31)].

So, by setting $\Sigma_{i,j} \equiv 1$ for $i = j$ and $\Sigma_{i,j} \equiv \rho$ for $i \neq j$ in correlation matrix [14], where ρ denotes the power correlation coefficient defined as $cov(R_i^2, R_j^2)/(var(R_i^2)var(R_j^2))^{1/2}$ and after the standard statistical procedure of transformation of variates and some mathematical manipulations, joint probability density functions (joint pdfs) for both desired and interfering signal envelopes can be expressed as:

$$p_{R_1,R_2}(R_1,R_2) = \frac{\left(1-\sqrt{\rho_d}\right)^{\mu_d}}{\Gamma(\mu_d)} \sum_{k_1,k_2=0}^{\infty} \frac{\Gamma(\mu_d + k_1 + k_2) \rho_d^{\frac{k_1+k_2}{2}}}{\left(1-\sqrt{\rho_d}\right)^{2\mu_d+k_1+k_2}} \times \mu_d^{2\mu_d+k_1+k_2} \left(\frac{1}{1+\sqrt{\rho_d}}\right)^{\mu_d+k_1+k_2} \times \prod_{i=1}^{2} \frac{\alpha_i R_i^{\alpha_i(\mu_d+k_i)-1}}{\Gamma(\mu_d+k_i) k_i! \hat{R}_i^{\alpha_i(\mu_d+k_i)}} \exp\left(-\frac{\mu_d R_i^{\alpha_i}}{\hat{R}_i^{\alpha_i}(1-\sqrt{\rho_d})}\right) \quad (4)$$

and

$$p_{r_1,r_2}(r_1,r_2) = \frac{\left(1-\sqrt{\rho_c}\right)^{\mu_c}}{\Gamma(\mu_c)} \sum_{l_1,l_2=0}^{\infty} \frac{\Gamma(\mu_c + l_1 + l_2) \rho_c^{\frac{l_1+l_2}{2}}}{\left(1-\sqrt{\rho_c}\right)^{2\mu_c+l_1+l_2}} \times \mu_c^{2\mu_c+l_1+l_2} \left(\frac{1}{1+\sqrt{\rho_c}}\right)^{\mu_c+l_1+l_2} \times \prod_{i=1}^{2} \frac{\alpha_i r_i^{\alpha_i(\mu_c+l_i)-1}}{\Gamma(\mu_c+l_i) l_i! \hat{r}_i^{\alpha_i(\mu_c+l_i)}} \exp\left(-\frac{\mu_c r_i^{\alpha_i}}{\hat{r}_i^{\alpha_i}(1-\sqrt{\rho_c})}\right) \quad (5)$$

where ρ_d and ρ_c are the correlation coefficients, and μ_d and μ_c are the fading severity parameters for the desired and interference signal, respectively. The average desired signal and interference powers at the i-th branch are denoted by $\hat{R}_i$ and $\hat{r}_i$, respectively.

Otherwise, by setting $\Sigma_{i,j} \equiv \rho^{|i-j|}$ in correlation matrix [14], for both desired signal and interference, the exponential correlation model can be obtained. So, the joint pdfs of desired and interfering signal envelopes for the exponential correlation system model can be expressed as:

$$p_{R_1,R_2}(R_1,R_2) = \sum_{k=0}^{\infty} \alpha_1 \alpha_2 \frac{\rho_d^{2k}}{2^{2(\mu_d+k)} \Gamma(\mu_d) \Gamma(\mu_d+k) k!} \times \frac{1}{\left(1-\rho_d^2\right)^{\mu_d+2k}} \times \exp\left(-\frac{R_1^{\alpha_1} + R_2^{\alpha_2}}{2(1-\rho_d^2)}\right) \times R_1^{\frac{\alpha_1}{2}(\mu_d+1)-1} R_2^{\frac{\alpha_2}{2}(\mu_d+1)-1} \quad (6)$$

and

$$p_{r_1,r_2}(r_1,r_2)=\sum_{l=0}^{\infty}\alpha_1\alpha_2\frac{\rho_c^{2l}}{2^{2(\mu_c+l)}\Gamma(\mu_c)\Gamma(\mu_c+l)l!}\times\frac{1}{\left(1-\rho_c^2\right)^{\mu_c+2l}}\times\exp\left(-\frac{r_1^{\alpha_1}+r_2^{\alpha_2}}{2(1-\rho_c^2)}\right)\times r_1^{\frac{\alpha_1}{2}(\mu_c+1)-1}\,r_2^{\frac{\alpha_2}{2}(\mu_c+1)-1} \tag{7}$$

Instantaneous values of SIR at the two input branches can be defined as $\lambda_1=R_1^2/r_1^2$, $\lambda_2=R_2^2/r_2^2$. The selection combiner (Fig. 2.) chooses and outputs the branch with the largest SIR:

$$\lambda=\lambda_{out}=\max(\lambda_1,\lambda_2) \tag{8}$$

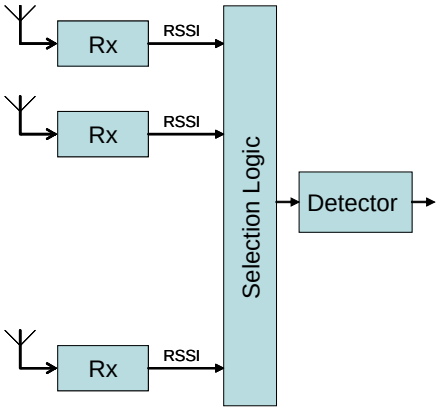


Fig. 2. SC receiver

Let $S_1=\hat{R}_1^2/\hat{r}_1^2$, $S_2=\hat{R}_2^2/\hat{r}_2^2$ be the average SIR's at the two input branches of the selection combiner. The joint pdf of instantaneous values of SIRs at the two input branches of SC, λ_1, λ_2 can be given by [15]:

$$p_{\lambda_1,\lambda_2}(t_1,t_2)=\frac{1}{4\sqrt{t_1t_2}}\int_0^{\infty}\int_0^{\infty}p_{R_1,R_2}\left(r_1\sqrt{t_1},r_2\sqrt{t_2}\right)p_{r_1,r_2}(r_1,r_2)\times r_1r_2\,dr_1dr_2 \tag{9}$$

For this case, the joint cumulative distribution function (joint cdf) of λ_i, $i=1,2$ can be written as [15]:

$$F_{\lambda_1,\lambda_2}(t_1,t_2)=\int_0^{t_1}\int_0^{t_2}p_{\lambda_1,\lambda_2}(x_1,x_2)dx_1dx_2 \tag{10}$$

Substituting the derived version of expression (9) in (10) and after two successive integrations, the joint cdf becomes:

$$F_\lambda(t)=\sum_{k_1,k_2=0}^{\infty}\sum_{l_1,l_2=0}^{\infty}G\times\prod_{i=1}^{2}\mathrm{B}\left(\frac{\mu_d t^{\frac{\alpha_i}{2}}}{\mu_d t^{\frac{\alpha_i}{2}}+\mu_c\frac{\left(1-\sqrt{\rho_d}\right)}{\left(1-\sqrt{\rho_c}\right)}S_i^{\frac{\alpha_i}{2}}},\mu_d+k_i,\mu_c+l_i\right) \tag{11}$$

with

$$G=\frac{\left(1-\sqrt{\rho_d}\right)^{\mu_d}\left(1-\sqrt{\rho_c}\right)^{\mu_c}}{\Gamma(\mu_d)\Gamma(\mu_c)}\Gamma(\mu_d+k_1+k_2)\Gamma(\mu_c+l_1+l_2)\rho_d^{\frac{k_1+k_2}{2}}\rho_c^{\frac{l_1+l_2}{2}}\times\left(\frac{1}{1+\sqrt{\rho_d}}\right)^{\mu_d+k_1+k_2}\left(\frac{1}{1+\sqrt{\rho_c}}\right)^{\mu_c+l_1+l_2}\times\prod_{i=1}^{2}\frac{\Gamma(\mu_d+\mu_c+k_i+l_i)}{\Gamma(\mu_d+k_i)\Gamma(\mu_c+l_i)k_i!l_i!} \tag{12}$$

for the constant correlation system model and

$$F_\lambda(t)=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\left(\Gamma(\mu_d+\mu_c+k+l)\right)^2\rho_d^{2k}\rho_c^{2l}\left(1-\rho_d^2\right)^{\mu_d}\left(1-\rho_c^2\right)^{\mu_c}}{\Gamma(\mu_d)\Gamma(\mu_c)\Gamma(\mu_d+k)\Gamma(\mu_c+l)k!l!}\times\prod_{i=1}^{2}\mathrm{B}\left(\frac{t^{\frac{\alpha_i}{2}}}{\frac{\left(1-\rho_d^2\right)}{\left(1-\rho_c^2\right)}+t^{\frac{\alpha_i}{2}}},\mu_d+k,\mu_c+l\right) \tag{13}$$

for the exponential correlation system model, with $\mathrm{B}(z,a,b)$ being the incomplete Beta function [10, eq. 8.391].

III. Outage Probability

Outage probability is an important system's performance measure, used to control the cochannel interference level. The designers of wireless communications systems readjust the system's operating parameters considering outage probability, in order to achieve the quality of service (QoS) and grade-of-service (GoS) demands. In the interference-limited environment, outage probability P_{out} is defined as the probability which combined-SIR falls below a given outage threshold q, also known as a protection ratio. If q is the protection ratio, defined as the ratio of the desired signal power to the interference power at the output of the combiner, the outage probability can be expressed as:

$$P_{\text{out}}=\text{Probability}(\lambda<q)=\int_0^q p_\lambda(t)dt=F_\lambda(q) \tag{14}$$

The protection ratio q depends upon the used modulation technique as well as on the desired QoS.

Fig. 3 and Fig. 4 illustrate the influence of correlation coefficients ρ_d and ρ_c and different values of fading parameters μ_d and μ_c on P_{out}. It is shown that outage probability increases when correlation among branches increases, which means degradation in performance gain. Otherwise, when parameter μ_d and/or μ_c increase (fading severity decreases), the values of outage probability also decreases, which means better system performances.

Comparing the two correlation models, it is noticed that there is compromise between them. In some cases the constant correlation model gives better results, in the other the

exponential correlation model, which depends on elected correlation coefficients and fading parameters.

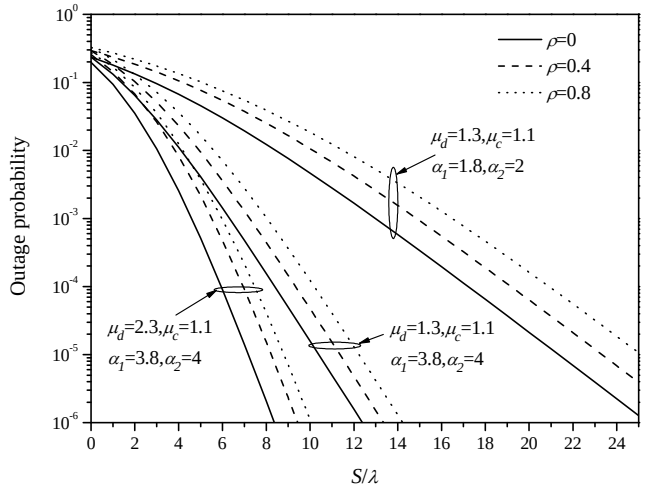


Fig. 3. Outage probability of constant correlation model versus S_1/q for different values of correlation coefficients and fading severity

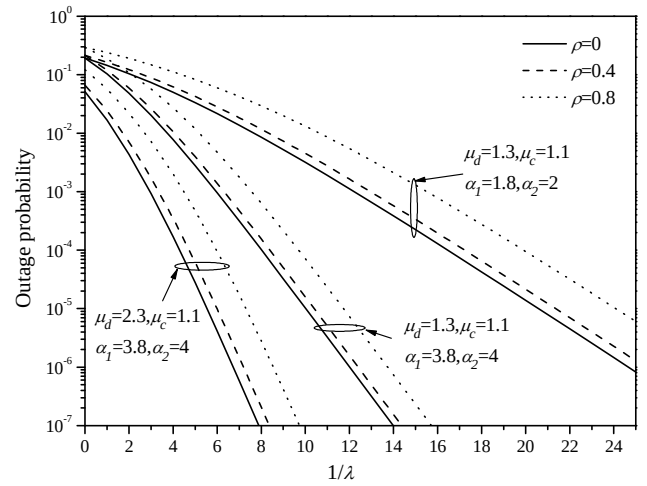


Fig. 4. Outage probability of exponential correlation model versus $1/q$ for different values of correlation coefficients and fading severity

IV. CONCLUSION

In this paper, the performance analysis of system with SC, based on SIR over correlated α–μ fading channels, was obtained. Constant correlation and exponential correlation models were observed for evaluating performances of proposed diversity system. The desired signal envelopes and interference were α–μ distributed. By considering α–μ distribution, we have included all important distributions which describe the statistics of radio signal through various environments. The analytical analysis for evaluating outage probability was presented. Using those results, the effects of various parameters such as the fading severity and level of correlation to the outage probability were observed.

ACKNOWLEDGEMENT

This paper was supported in part by the Ministry of Science of Serbia within the project "Development and realization of new generation software, hardware and services based on software radio for specific purpose applications" (TR-11030).

REFERENCES

[1] Simon MK, Alouini MS. *Digital Communication Over Fading Channels*. Wiley:New York, 2000.

[2] H. Suzuki, "A statistical model for urban radio propagation," *IEEE Trans. Commun*., vol. COM-25, pp. 673-680, July 1977

[3] W. C. Jakes, Mobile Communication Engineering. New York: Wiley,1974.

[4] S . Stein, "Fading channel issues in system engineering," *IEEEJ. Select. Areas Commun.*, vol. SAC-5, pp. 6849, Feb. 1987.

[5] M. D. Yacoub, "The α-μ distribution: a physical fading model for the stacy distribution," *IEEE Trans. Veh. Technol.*, vol. 56, no. 1, pp. 27–34, Jan. 2007.

[6] Brennan D. Linear diversity combining techniques. *Proc. IEEE*, Jun 1959; 47(6),pp. 1075–1102.

[7] E. A. Neasmith and N. C. Beaulieu, "New results in selection diversity," *IEEE Trans. Commun.*, vol. 46, pp. 695–704, May 1998.

[8] S. Okui, "Effects of CIR selection diversity with two correlated branches in the m-fading channel," *IEEE Trans. Commun.*, vol. 48, pp.1631–1633, Oct. 2000.

[9] G. K. Karagiannidis, "Performance analysis of SIR-based dual selection diversity over correlated Nakagami-*m* channels," *IEEE Trans. Veh. Technol.*, vol. 52, no. 5, pp. 1207-1216, Sep. 2003.

[10] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and products*, 5th, Ed. New York: Academic, 1994.

[11] M. Č. Stefanović, D. M. Milović, A. M. Mitić and M. M. Jakovljević, "Performance analysis of system with selection combining over correlated Weibull fading channels in the presence of cochannel interference," *AEÜ - International Journal of Electronics and Communications*, vol. 62, no. 9, pp. 695-700, October 2008.

[12] J. Reig, "Multivariate Nakagami-*m* distribution with constant correlation model," *Int. J. Electron. Commun. (AEUE)*, DOI: 10.1016/j.aeue.2007.10.009, 2007.

[13] Karagiannidis GK, Zogas DA, Kotsopoulos SA. On the multivariate Nakagami-m distribution with exponential correlation. *IEEE Trans Commun 2003*, COM-51, pp. 1240–1244.

[14] V. A. Aalo, "Performance of maximal-ratio diversity systems in a correlated Nakagami fading environment," *IEEE Trans. Commun.*, vol. 43, pp. 2360-2369, 1995.

[15] C. W. Helstrom, *Probability and Stohastic Processes for Engineers*, 2nd ed. New York: MakMillian, 1991.

Application of MPPT controller for hybrid alternative electrical power grid independent source

Vladimir Smiljakovic[1] and Zoran Zivanovic[2]

Abstract – **In this paper one example of the field proven application of the MPPT (maximum power point tracking) controller for hybrid photovoltaic solar and wind generated alternative electrical power is presented. This power source is located in uninhabited hard to reach mountain region and completely independent from electrical power grid or any other source of energy. Successful uninterrupted operation in harsh environment including winter period of year completely approved chosen approach.**

Keywords – **Maximum power point tracking (MPPT) controller, alternative electrical power sources, photovoltaic (PV) panel, wind generator**

I. Introduction

Alternative sources of electrical energy that use renewable energy (for example: Sun energy, wind energy, sea wave and tide energy) become very significant in recent years. Besides ecological aspects, economic aspects including investment and maintenance costs are motivating factor of rising influence in process of decision making concerning choice of acceptable and cost effective energy sources of future.

As an illustration, in Germany that is one of the most developed countries of the World, constant growth of installed power of PV sources of electrical energy of more than 20% year on year is registered, reaching by the end of 2009. more than 8,5 GW of total installed power [1]. Also building and exploitation of other renewable sources of electrical energy – e.g. wind generators concentrated in so called wind farms or in single stand alone installation is in great expansion nowadays.

One of the most interesting alternative sources of electrical energy is conversion of Sun energy in photovoltaic semiconductor cells in electrical energy.

Due to properties of PV semiconductor material electrical power generated by PV semiconductor array decreases with growth of temperature (presented at Fig 1.). At the other side, electrical power generated by PV panels-arrays increases by increase of insolation of Sun (presented at Fig 2).

So in regions with continental climate hybrid configuration consisting of: PV arrays and wind generator as mutually independent and complemented by nature of generating of electrical power sources regarding day/night and seasons of the year conditions is natural choice. These alternative energy sources configurations as necessary parts have accumulators of electrical energy and controller [2].

[1]Vladimir Smiljakovic is with the Institute of Microwave Technique and Electronics IMTEL communications, Bul. M.Pupina 165b, 11070 N. Belgrade, Serbia, E-mail: smiljac@insimtel.com

[2]Zoran Zivanovic is with the Institute of Microwave Technique and Electronics IMTEL communications, Bul.M.Pupina 165b, 11070 N.Belgrade, Serbia

II Maximum Power Tracking (MPPT) Controller for PV Arrays

Especially in situations when alternative electrical energy sources are not connected to power grid, i.e. when electrical power supply relies exclusively on local alternative energy source it is of the highest importance to use as much energy from available source of energy as possible. So obtaining maximum coeficient of energy utility is highly appreciated, as well as reliability in operation.

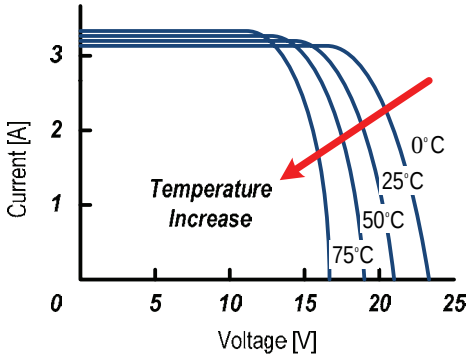


Fig 1. Dependence of PV array voltage (power) on temperature

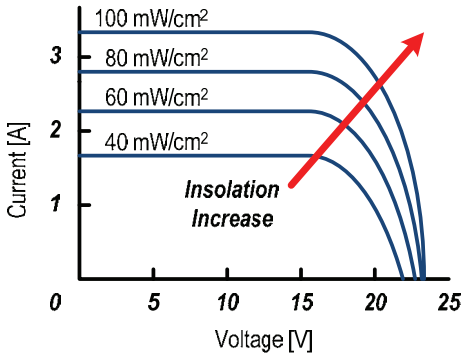


Fig 2. Dependence of PV array voltage (power) on Sun ray intensity per square meter (insolation)

Concerning obtaining as much electrical energy as possible from PV panel array two aspects are of greatest influence:

1) mechanical i.e. position of PV array upper surface to Sun rays direction – optimal position to receive maximum Sun energy is surface perpendicular to Sun light ray and

2) electrical – optimization of coefficient of utilization – conversion of PV array electrical energy to electrical energy given to electrical load (equipment and batteries)

There are two kinds of tracking concerning PV array panels [3]:

1) Panel position mechanical tracking of Sun position – panels are on mount that tracks Sun position across the sky for maximum sunlight energy reception. Increase of obtained electrical energy by this tracking is 15% in winter and up to 35% in summer (at Northern hemisphere of Earth) according to [4]. One should bear in mind that application of mechanical tracking system pose difficulties in maintenance due to existence of electromotor drive and periodic gears lubrication, icing and similar preventive maintenance actions.

2) Maximum power point tracker (MPPT) effects are just the opposite: 20 to 45% power gain in winter (i.e. in the most critical part of year when PV supply is concerned) and 10-15% in summer. This is pure electronic way of tracking and has nothing to do with moving panels. The electronic MPPT controller measure voltage output of the PV panel and voltage of the battery to find out what is the best (i.e. highest) power PV panel can take out to charge the battery. The MPPT controller takes this voltage from PV panel array and converts it to the voltage that gives maximum current in actual situation to charge the battery. According to results presented in literature, configurations with MPPT controllers have 92 - 97% efficiency in the energy conversion while classical approach can give as much as 35% worse efficiency. Actual gain vary significantly all the time depending on the changing weather conditions at the particular location (e.g. Sun light insolation, actual shadowing of PV panel by clouds, rain, fog, leafs, dust, snow etc), ambient temperature and battery state.

As the MPPT controller actually is DC/DC converter that poses optimum load to PV panel and thus makes the PV panel operate at its peak power delivery regime it is basic difference from classic controller solution that takes PV voltage value directly as battery charging voltage.

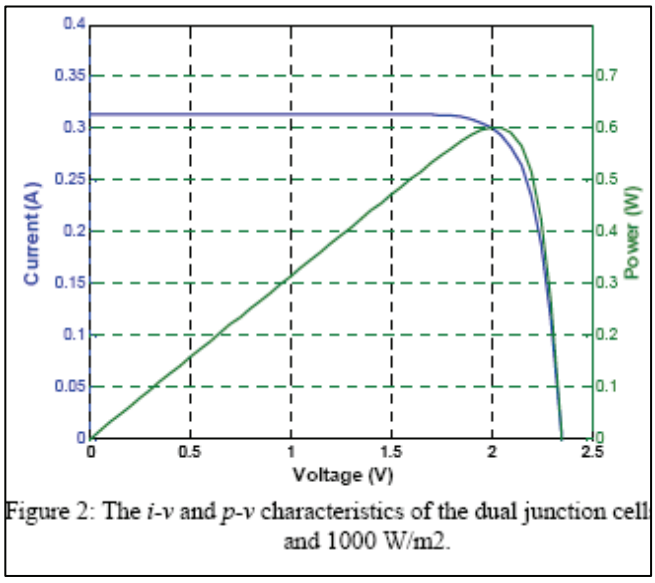


Figure 2: The *i-v* and *p-v* characteristics of the dual junction cell and 1000 W/m2.

Figure 3. Curent-voltage and power-voltage dependence of the typical semiconductor junction (PV cell) – 1000 W/m2 insolation atambient temperature 20degree ^0C

Since the ambient working conditions are continually changing (amount and distribution of radiant sunlight, temperature of the panel first of all) MPPT must constantly adapt its state to obtain maximum energy to battery (consumer) from PV panel. Working point is in the vicinity of zero slope point of p-v curve of PV cell (PV cell is basic element of every PV array that is formed by connecting PV cells serial and parallel configurations according to request and needs of end users). One can easily see so called "knee point" at the p-v curve characterizing PV cell at Figure 3. (i.e. actual maximum power of PV panel point). At that point from mathematical point of view is zero slope of characteristic, so MPPT controller in varying environmental conditions (i.e. varying working point) tends to be as close as possible to that point to obtain as much as possible power from PV source of energy to deliver to battery and load.

III Description of the Chosen Alternative Source MPPT Configuration

Due to the very hard access to the location and severe meteorological conditions at the location, the redundant configuration of the power supply system for the communication node is chosen. It consists of three completely independent electrical power generators: two PV solar systems with identical characteristics and one independent wind generator based electrical energy generating system. Block diagram of applied hybrid alternative energy supply with MPPT controllers for PV electrical energy power sources is presented at Figure 4.

The total requested autonomy (in the worst case) is 10 days: without any Sun energy available and without any wind stronger than 3m/s available. Complete detailed analysis and estimation based on calculations is presented in [4].

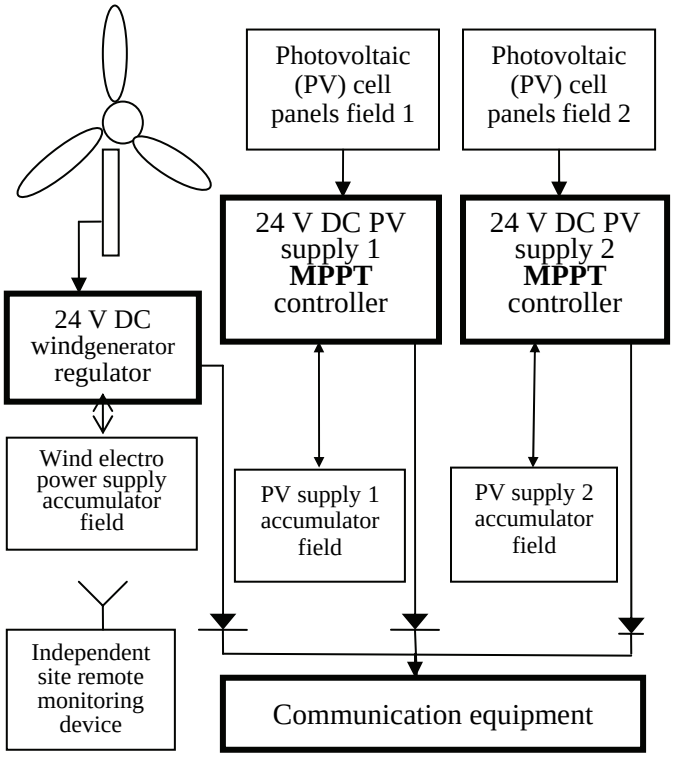


Fig. 4 Block diagram of applied hybrid alternative energy supply with MPPT controllers for PV electrical energy power sources

Due to MPPT controller application availability and autonomy of realized PV electrical power is significantly improved. At the same time MPPT application allowed use of PV panels with smaller dimensions at the field, thus enabling lighter supporting mechanical construction and easier transportation to site.

Reference [5] shows that a single PV solar system fulfills the daily request for electrical power supply, so part of the system based on the wind generator is added in purpose to upgrade the reliability of the electrical power generation 365 days a year. The accumulators are chosen to provide a 10 day total autonomy of the requested power supply for the communication node. The described hybrid alternative power system is for a microwave digital radio relay active repeater station power supply and have to secure uninterrupted functioning 24h/day all over the year in very heavy conditions that are at high mountain hard to access locations. Complete finished repeater station at work with alternative power source is presented at Figure 5. Two PV panel arrays are at the top of cabinet, facing South, and with panel angle to horizon equal to Latitude+22.5^0 (optimal position for the worst position of the site to Sun rays).

Fig 5 Photo of the realized alternative hybrid energy supply for TELEKOM Serbia microwave communication node at hard to reach Orlovac mountain location in south east part of Serbia (1670m Above Sea Level)

Wind generator is placed at the top of the tower to gain as much energy as possible from the local winds. Cabinet contains MPPT controllers, wind generator controller, wiring, fuses, batteries and independent wireless site remote monitoring device (for acquisition of all relevant data).

IV Results Obtained in the Field

The solution with MPPT controllers applied gives exelent results in the field exploitation in extremely harsh mountain environment [2],[4],[5]. Request of the maintenance free power grid independent electrical power source is fully satisfied. By continual monitoring of all vital parameters of digital radio relay node including complete parameters of power supply system – two independent PV supply systems and wind generator system as well, logging all monitored data and periodically in regular intervals sending data packets by independent wireless communication channel to remote monitoring center, not a single one situation of miss functioning or lack of energy for communication node is observed. All the time all the requested parameters of presented alternative power source were deep in the optimal region.

Results obtained during almost one year of exploitation without any problem in the field continually supplying uninterrupted electrical energy for communication node located in the extremely hard to reach mountain region at height more than 1600 m above sea level highly approve applied approach.

V Conclusion

The hybrid alternative electropower generator system based on PV panel electrical energy source and wind generator is build as uninterrupted electrical energy supply of 15 W DC power for communication node situated in very harsh and hard to reach mountain ambient.

For the PV based part of described electrical power generating system maximum power point tracking (MPPT) controller is applied, resulting in field proven high quality utilization of electrical energy obtained from PV array. Almost optimal utilization of PV array generated electrical energy due to MPPT controller usage enabled reliable and stable electrical power supply for communication equipment in all working conditions, including toughest.

Acknwledgement

The results presented in this paper are part of results obtained working at the project with the support of the Ministry of science of the Republic of Serbia

References

[1] Paul Buckley „Global PV Market Gains 20% Year on Year“, EE Times Europe Power Management News Letter, April 13, 2010.

[2] IMTEL Communications: Active microwave repeater station, Available at: http://www.insimtel.com

[3] www.solar-electric.com/charge_controls/mppt.htm

[4] Zoran Zivanovic, Vladimir Smiljakovic “Example of Hybrid Power Supply Combining Solar and Wind Power”, Proceedings of ETRAN 2008 Conference (in Serbian), part A, Palic, Serbia 2008.

[5] Vladimir Smiljakovic, Zoran Zivanovic “Hybrid Alternative Power Supply for Microwave Active Repeater Operating in Severe Conditions”, XLIV ICEST Conference Proceedings, Veliko Trnovo, Bulgaria, 2009.

This page intentionally left blank.

Wideband Receiver for Signal Detection in Frequency Range from 15 to 19 GHz

Sinisa P. Jovanovic, Predrag S. Manojlovic, Dragan D. Obradovic and Nemanja M. Mitrovic

***Abstract* – The paper presents the concept, realization and measured results for a receiver prototype designed for signal detection in a frequency range from 15 to 19 GHz. The characteristics of the presented unit are described in detail as well as the enhancements relative to the previously developed laboratory model. As the final result, a compact, dependable, repeatable and inexpensive device is obtained. This receiver is one of the essential parts of the device for detection of radar signals as well as others signal sources from the microwave frequency range.**

***Keywords* – Radar signal detection, receiver, microwave frequency range.**

I. Introduction

Together with the invention of the first radars emerged the need for devices which would, in real time, determine the frequency, intensity and direction of the unknown electromagnetic radiation. In the beginning, those devices, as well as radars, were heavy and complicated facilities primarily intended for use on a strategic level. However, at the same time as the development and increase in the use of radars in the preparation and executing of all sorts of battle operations, the importance of the use of the devices for their detection also increased, as well as the need for their improvement and widespread use so that in the conditions of modern warfare every vehicle or object that is the aim of radar observation could be equipped with them.

II. A Short Description of the Detection Device

Detecting and locating an unknown microwave signal in real time is not an easy task. It requires determining at least three parameters of that signal: the time of appearance, frequency and at least one spatial coordinate-azimuth. In order for the task to be realistically achievable, it is necessary to define certain tradeoffs right at the beginning. For example the observing frequency range needs to be limited to the most interesting frequencies based on experience and available data of the radars' characteristics that are planned to be observed. Also, certain limitations should be established regarding the resolution or accuracy with which the unknown frequency and azimuth are determined. The device discussed in this paper is projected for the detection of an unknown signal in the frequency range from 7 to 19 GHz, which is divided into three sub-ranges: from 7 to 11 GHz, from 11 to 15 GHz and from 15 to 19 GHz, marked on Fig. 1 and Fig. 2 as lower (L), middle (M) and upper (U) frequency ranges respectively. Thanks to that, the construction and the realization of the receiving antennas are significantly simplified by narrowing the frequency working range, of course at the price of tripling the total number of receiving antennas in the system. Another significant convenience is the possibility of using the common local oscillator for all three sub-ranges [1], which is very important considering the complexity and price of this subassembly, as well as the need for the synchronization and control of the whole system.

Fig. 1 shows the schematic of the entire device. Its antenna system is made up of 24 linearly polarized horn antennas, with 8 on every 3 sub ranges. In the azimuth plane, the antennas from the same sub range are spaced by 45°. All antennas have a 3 dB width of the main lob of 60° in both the E and H plane, which allows every antenna to cover a sector of the azimuth 45° in width. Depending on the relative position of the radiation source and radiation pattern and orientation of the receiving antennas, different levels of unknown signal will be received by the antennas of this system, which would be the base for determining the azimuth of the unknown signal by employing a special algorithm [2] in the signal processing block. The measuring accuracy, before everything else, depends on the equalized characteristics of both the horn antennas and the receiving blocks and their components. In order to enable the detection of both the horizontally and the vertically polarized signals, all the antennas are positioned so that the E-plane of their radiation pattern forms a 45° angle with the horizontal plane.

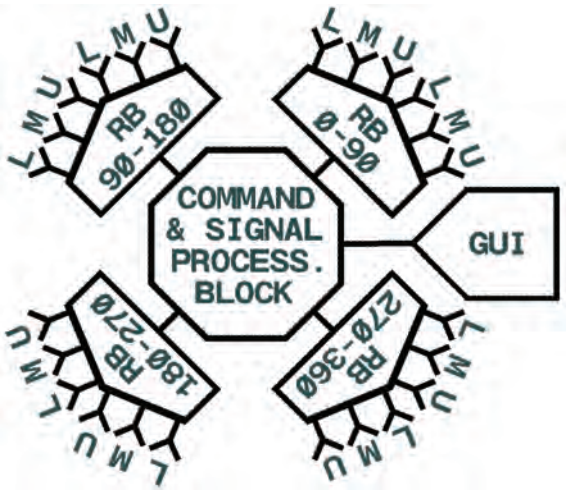


Fig. 1 The block scheme of the device for radar detection
L, M, U- receiving antennas for the Lower, Middle and Upper frequency sub-range respectively; RB xxx-xxx- receiving blocks for the related sectors of the azimuth; GUI – Graphic User Interface

The authors are with the IMTEL Communications, Blvd. M. Pupina 165B, 11070 N. Belgrade, Serbia, E-mail: siki@insimtel.com

Fig.2 shows a schematic of one of the four identical receiving blocks of the detection device, each intended to cover a 90° wide azimuth sector, for which a pair of antennas in each frequency sub range is needed. Each pair is connected to a corresponding receiver via the antenna switch. It is also possible to have a configuration of the receiver block without antennas switches, but it would require twice more of all other subassemblies. The local oscillator generates signals in frequency range from 11 to 15 GHz, with a 100 MHz step, which are divided into three samples and amplified up to the levels required for the normal operation of mixers in all three receivers.

The block for the control and acquisition directs the changes in frequency of the signal of the local oscillator and the work of the antenna switch. This block also reads the analogue values on the exit of the detector of all three receivers and communicates with the command and processing block.

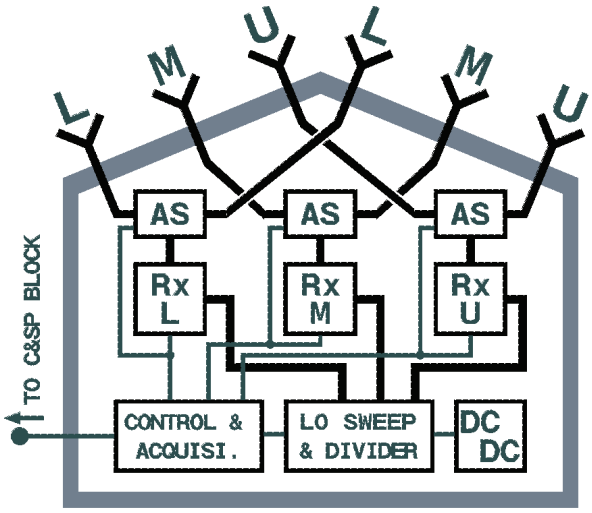


Fig. 2. The block scheme of one receiving block.
AS: antenna switches; DC/DC: power supply;
Rx(L,M,U): receivers for the Lower, Middle and Upper sub-range.

Unlike other, mentioned up to now, parts of the receiver block which are identical or common for all three sub ranges, receivers RxL, RxM and RxU are realized for each sub range separately. Although their configurations are mutually similar, they are different when it comes to working frequencies of the input band-pass RF filters as well as the IF band-pass filters. Since the length of this paper isn't sufficient for the description of all three types of receivers, in the remaining part of this paper only the receiver for the highest (upper) sub range will be presented and described.

III. Laboratory Model of the Receiver

On Fig.3 a block scheme of the laboratory model of the receiver is given, which in that phase of development was projected for a frequency sub range from 15 to 18GHz. The same picture shows an antenna switch and a wideband band-pass filter which are realized in separate housing. The low noise amplifier (LNA) shown on Fig. 4 is realized as a two-stage transistor amplifier using transistors NE3210S01 [3]. The amplifier has a gain of 14 dB with a noise figure of 3 dB.

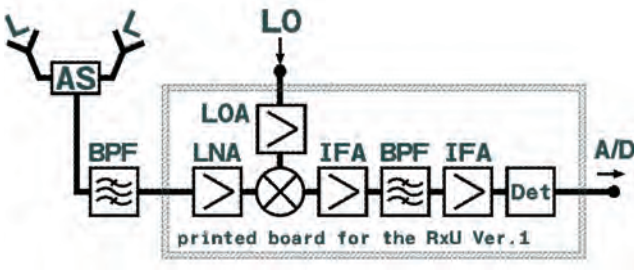


Fig. 3. A block- scheme of the laboratory model of the receiver

The LO signal is amplified with LOA (HMC441) to the level required by the mixer (HMC144) [4]. The obtained signal at intermediate frequency with two ERA1 amplifiers [5] both before and after passing through the narrow band-pass filter (fc=3 GHz, BW_{3dB}=200 MHz) made as a printed filter with slow-wave open-loop resonators [6]. The logarithmic detector [7] (Det) produces a DC signal reversely proportional to the signal level at its input, which is further processed in the control and acquisition block. The photo of the realized laboratory model of the receiver is presented on Fig. 4.

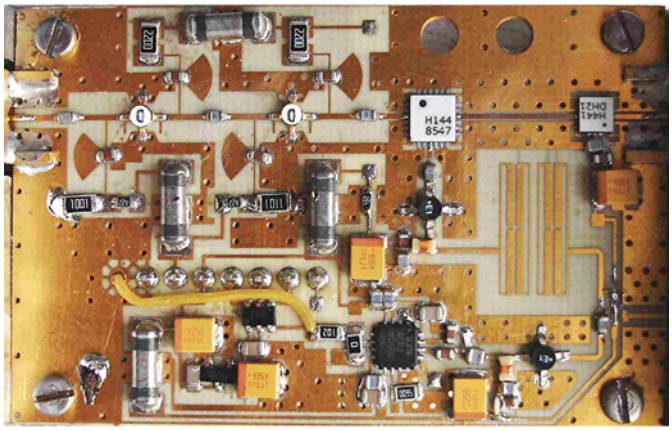


Fig. 4 Photo of the laboratory model of the receiver

The evaluation of the realized receiver usually means the excitation of its entrance with a series of signals of a known level and frequency and the reading and memorizing of the levels of the detected signal at its output. The results obtained in that way are shown on Fig.5.

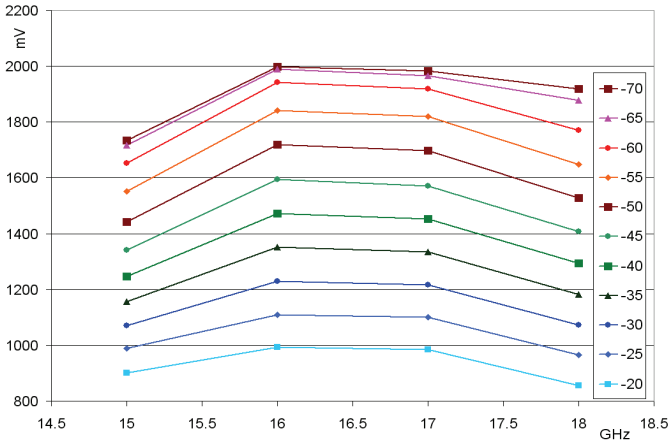


Fig.5: The measured results of the signal at output of the receiver depending on the frequency and level of the input signal

The frequency of input signals has changed from 15 to 18 GHz, with a step of 1 GHz, and level from -70 to -20 dBm, with a step of 5 dB. In the case of the ideal logarithm detector, the obtained curves would be horizontal, parallel and equidistant. The results shown on Fig. 5 reveal a decline of the sensitivity and the overall conversion gain in the centre of the working bandwidth. Comparing the results from Fig. 5 and the characteristics of the logarithmic detector AD8318 shows that this version of the receiver has quite a low overall conversion gain that vary from 0 to 3 dB over working bandwidth.

IV. The New Version of the Receiver

The experiences gathered during the making and evaluation of the laboratory model have led to some significant improvements in the next version. Primarily, the intermediate frequency has been changed from 3 to 4GHz, which has enabled the expansion of the receiver sub range from 15-18 GHz to 15-19 GHz, respectively the expansion of the working frequency range of the whole device for radar detection from 8-18 GHz to 7-19 GHz. The next improvement was the integration of the receiver with the input band pass filter as well as the antenna switch, as shown on Fig. 6.

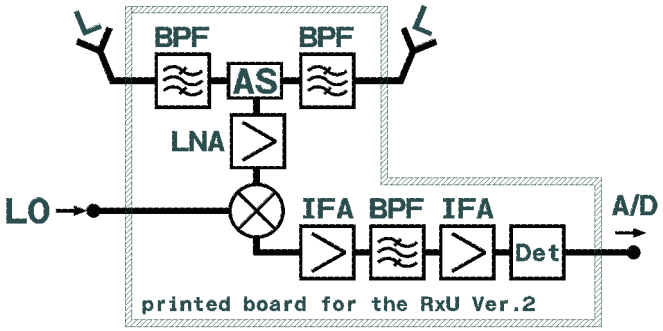


Fig. 6 A block-scheme of the prototype of the receiver

Instead of one filter behind the switch, like on Fig. 3, the new version has two filters, one behind each antenna entry of the receiver, but that hasn't changed the total entrance attenuation, or the noise factor of the receiver, nor has it changed the total size of the plate on which it is realized. The dimensions of the previous version of the receiver were 38 × 58 mm. The new version of the receiver, in spite of having an even bigger number of elements, is realized on only one board with dimensions 30 x 67 mm. All the components, except for the connector for the connection with the motherboard, are installed on one side of the board. The integration of the receiver with a band pass filer and antenna switch has significantly lowered the number of components needed for its connection with the surrounding subassemblies, which is shown in Table 1. This way, next to lowering the production costs, a significantly higher reliability of the entire device is achieved, because the number of connections that could become loose due to mechanical shocks to which the device is exposed in real working conditions is scientifically reduced.

In order for the new version to be suitable for serial production, it was necessary to design printed filters whose characteristics will be stable with variations of projected dimensions typical for available mass production technology. The printed filters which needed to be designed are: the input RF band pass filter with pass-band (1 dB) from 15 to 19 GHz (instead of from 15 to 18 GHz that was required for the previous receiver's version); as well as the IF band-pass filter having fc=4 GHz and relatively narrow (for printed filters) 3 dB-pass-band of about 150 MHz. The filters were designed and realized on substrate RO4350 (h=0.762 mm, εr=3.48, tanδ=0.0037), which is for above requirements more suitable than the substrate used for laboratory model of the receiver RO4003 (h=0.2 mm, εr=3.38, tanδ=0.003) [8]

TABLE I
REQUIRED NUMBER OF COMPONENTS VS RECEIVER'S VERSION

Component type	Receiver's version	
	v1 (lab. model)	v2 (prototype)
SMA connector	20	12
Flat connector	4	2
Semi-rigid cable	6	3
Flat cable	1	0
Mechanical case	3	1

The band pass filter is realized as a standard printed filter with parallel coupled resonators of the fourth order [9], whose dimensions are optimized with a program for electromagnetic simulation [10] to make it resistive to tolerances during production. The chosen version's characteristics stay within specified limits even for realized widths that deviate from the projected ones for ±50 μm, which makes a very big relative error considering that the minimal projected width of the lines or gaps is 120 μm.

Separate test versions of the filters are fabricated in the same time and under the same conditions as the receiver's board containing integrated filters to enable measurements of the filters' characteristics. Fig. 7 shows the measured results of the realized input band-pass test filter. Over the working bandwidth (m1 to m2), the maximum insertion loss is 1.82 dB with a variation lower than 0.5 dB (m1 and m2), while the return loss is better than -18 dB.

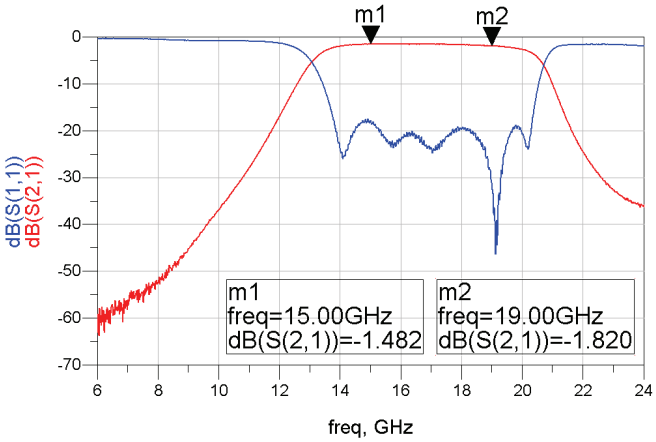


Fig.7 Measured characteristics of the input RF band-pass filter

The 3 dB bandwidth of the IF band-pass filter defines the measuring accuracy of the whole detection device. In the previous version that width amounted to 200 MHz, or 6.6% relative to the central frequency of 3 GHz. During the prototype development phase, one of the basic requirements was the

narrowing of its 3dB bandwidth to at least 150 MHz with the central frequency of 4 GHz, which gives a relative bandwidth of 3.75%. Like in the previous version, the filter is realized with four resonators with an open loop [5], but the change of the substrate enabled the required narrowing of the bandwidth.

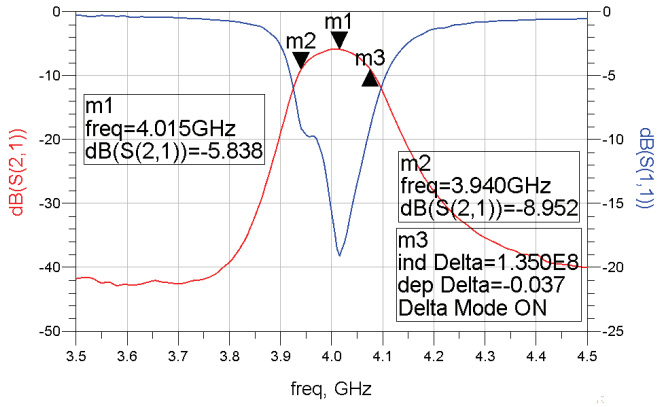


Fig.8 Measured characteristics of the IF band-pass filter

Fig. 8 shows the measured values of the frequency characteristics of this filter realized on a test board. The narrowing of the bandwidth is realized partially at the price of a somewhat bigger insertion loss at the central frequency of about 6 dB (m1). The insertion loss in the band-stop region is about 40 dB, while the 3 dB bandwidth is 135 MHz (m2 to m3).

V. Realization and Measured Results

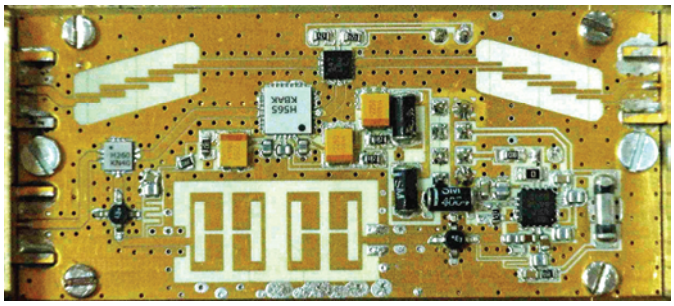


Fig.9 Photo of the prototype version of the receiver

TABLE II
MAJOR CHARACTERISTICS OF BOTH RECEIVERS' VERSIONS

Characteristics	Receiver's version	
	lab. model	prototype
Dimensions [mm]	36×56	30×67
Operating range [GHz]	15-18	15-19
Max conversion gain [dB]	3	16
Min conversion gain [dB]	0	15
Overall noise figure [dB]	12	8
Sensitivity [dBm]	-70[1]	-81
Max RF_{IN} (±1dB error) [dBm]	-5	-16
Dynamic Range [dB]	65	65
Useful dynamic range [dBm] (for max expected RF_{IN} =-30dBm)	40	51

[1]This low sensitivity is caused by low conversion gain so that the input RF signal range wasn't adjusted to the dynamic range of the logarithmic detector

Fig. 9 shows a realized prototype of the receiver, while Table II shows measured results of its major characteristics compared to the previous version.

The major improvement is increasing the conversion gain for 13dB and lowering the overall noise figure (that includes all losses between the horn antenna and the receiver) for 4dB. The higher overall conversion gain is achieved by employing an input amplifier with higher gain (HMC565) and a mixer with lower conversion loss (HMC260) [4], while the lower noise figure is mainly achieved by integration of the components and avoiding unnecessary losses in connecting cables as well as the realization of the input BPF with lower insertion loss.

VI. Conclusion

In this paper two versions of the receiver for the detection of an unknown signal are described in detail. The improvements in the prototype version, which have been achieved based on experiences with the laboratory model, are especially emphasized. The realized prototype version of the receiver has a conversion gain of 16 dB, with a noise figure of 8 dB and a sensitivity of -81 dBm. As a final result a reliable, repeatable, compact and inexpensive subassembly is obtained, which is one of the key elements of the device for the detection of radar radiation and other unknown sources of signals of microwave frequency.

Acknowledgement

The authors wish to thank to Miss Maja Jovanovic for her help in translating this paper into English.

This work is supported by the Ministry of Science and Technological Development of the Republic of Serbia.

References

[1] N. Mitrovic, D. Obradovic, S. Tasic: "High Frequency Low Cost Fast Channel Switching Synthesizer", Proceedings of 9th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service - TELSIKS, Nis, 7-9 Oct. 2009, pp.48-51.

[2] N. Mitrovic, D. Obradovic, P. Manojlovic: "One Method Source Radiations Azimuth Finding Using Horn Antennas", Proceedings of papers, 53th ETRAN Conference, Vrnjacka Banja, 15.-18. June 2009, paper AP1.2, str.1-4.

[3] www.cel.com/pdf/datasheets/ne3210s1.pdf

[4] www.hittite.com

[5] Hong, J.-S., and Lancaster, M.J.: 'Theory and experiment of novel microstrip slow-wave open-loop resonator filters', IEEE Trans. Microw. Theory Tech., 1997, 45, (12), pp. 2358–2365

[6] www.minicircuits.com/pdfs/ERA-1+.pdf

[7] www.analog.com

[8] www.rogerscorp.com

[9] Ian Hunter: "Theory and Design of Microwave Filters", University Press, Cambridge, 2001.

[10] IE3D User's Manual, Zeland Software Inc.

Cassegarin Antenna of 0.9m at 10.5GHZ

Zoran Mićić, Vladimir Smiljaković and Ivan Jovanović

***Abstract* – The paper presents analysis and simulation of a standard axially symmetric dual-reflectors Cassegrain antenna with minimal blockade, linear polarization (horizonatal and/or vertical), 0.9m diameter, operating from 10.2 to 10.7GHz range. Dimensions of the primary radiator and the subreflector of the antenna are optimized by WIPL-D – the program package for rapid and precise analysis of metallic and/or dielectric/magnetic structures in the frequency domain [1].**

***Keywords* – reflector antenna, Cassegrain, WIPL-D**

I. Antenna Description

The antenna, *Fig. 1*, consists of:
-parabolic reflector
-primary feed
-coaxial-waveguide adaptor
-subreflector supporting
-hyperbolic subreflector
-link housing case
-positioning mechanism

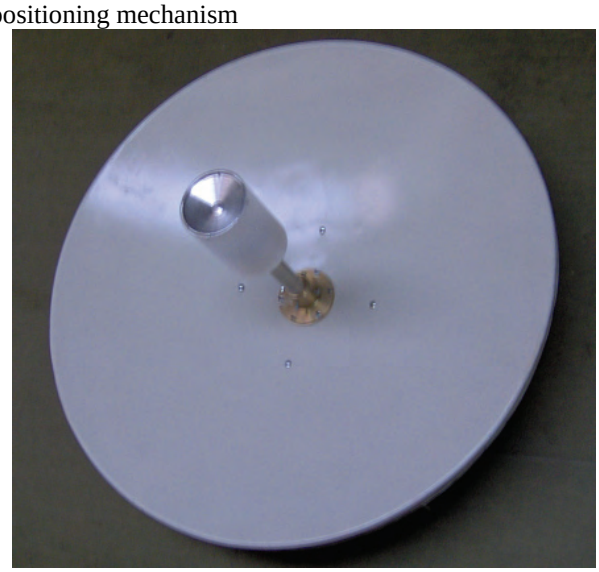


Fig. 1 Antenna parts

II Design Procedure

The antenna is designed using principles of microwave optics [2],[3],[4]:
According to known parameters:
D_{PR}- primary feed diameter
L_{FC}- distance from the phase center to the primary feed aperture
D_R- parabolic reflector diameter
F- focal length of the parabolic reflector
Θ-10dB radiation angle of the primary feed,
under condition of minimal blockade (equal shadow of the subreflector and the primary feed on the reflector), *Fig. 2*, we have determined the unknown parameters:
f- hyperbolic subreflector focal length
D_S- hyperbolic subreflector diameter
L_{RFC}- distance between the reflector center and the primary feed phase center
L_S- distance between the subreflector center and the primary feed phase center
a- major half-axis of the subreflector

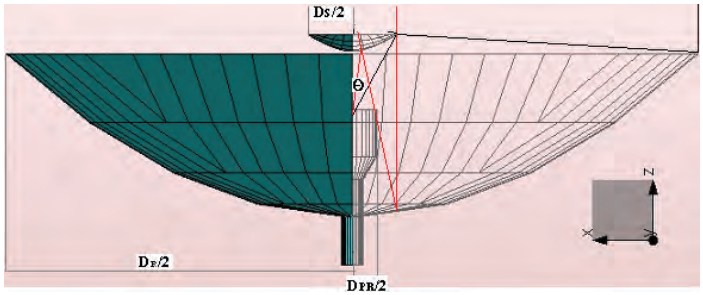


Fig. 2 *Position of the antenna parts.*

III Results of the WIPL-D Electromagnetic Analysis

Primary feed is a conical "dual mode" (TE_{11} i TM_{11}) horn antenna, *Fig. 3.*

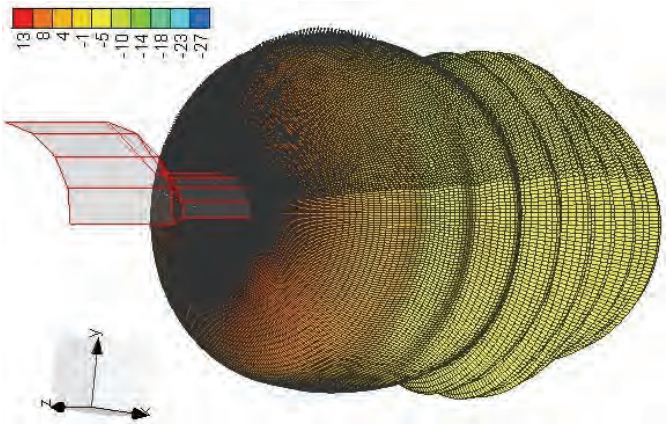


Fig. 3 – *WIPL-D model of the primary feed and its 3D radiation pattern.*

Zoran Mićić is with IMTELkomunikacije a.d., Bul. M. Pupina 165b, 11070 N. Belgrade, Serbia, zoran@insimtel.com
Vladimir Smiljaković is with IMTEL komunikacije a.d., Bul. M. Pupina 165b, 11070 N. Belgrade, Serbia, smiljac@insimtel.com
Ivan Jovanović is with IMTEL komunikacije a.d., Bul. M. Pupina 165b, 11070 N. Belgrade, Serbia, ivan@insimtel.com

Dimensions of the primary feed are optimized so to obtain the best possible symmetry correlation of the E- and H-plane radiation patterns, and the Θ-10dB radiation angle of about 55^0.

Position and length of the probe in the primary feed wavegude are determined by WIPL-D analyses resulting in return losses RL>19dB.

Complete antenna is modeled and optimized at 10.45GHz, using WIPL-D program package, *Fig.4*.

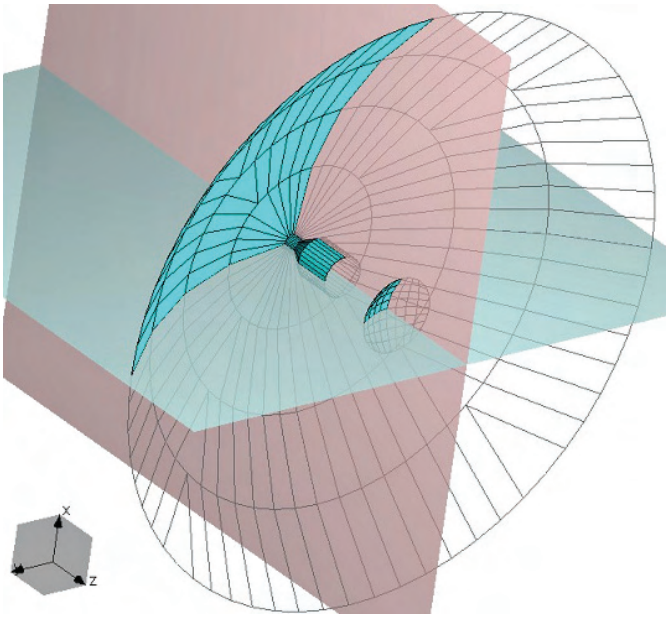


Fig. 4. *WIPL-D model of the antenna.*

Results of the WIPL-D analysis are shown in Table 1:

f [GHz]	***10.2***	***10.45***	***10.7***
G [dBi]	36.1	36.5	36.2
$FSLA^E$ [dB]	16.6	19.4	16.5
$FSLA^H$ [dB]	16.1	17.2	16.0
F/B [dB]	43.1	43.2	43.1
$HPBW^E$ [º]	2.1	2.1	2.05
$HPBW^H$ [°]	2.05	2.0	2.0
XPD [dB]	30	30	30

Table 1 *WIPL-D analysis results*

G- Antenna gain

FSLA- First side lobe attenuation

F/B- Front-to-back ratio

$HPBW^E$- Half-power beamwidth in E-plane

$HPBW^H$- Half-power beamwidth in H-plane

XPD –Cross polar discrimination

Spherical radiation patterns in E-plane (Φ=0) and H-plane (Φ=90) at the frequency f=10.45GHz are presented in *Fig. 5*.

First side lobe attenuations (FSLA) in H-plane (Φ=90), and E-plane (Φ=0) at the frequency of 10.45GHz are shown in *Fig. 6*. It can be seen that the position of first lobes are 3.3º/3.5º apart from the maximal radiation direction. HPBW angles in H- and E-plane are about 2.0º/2.1º (H/E).

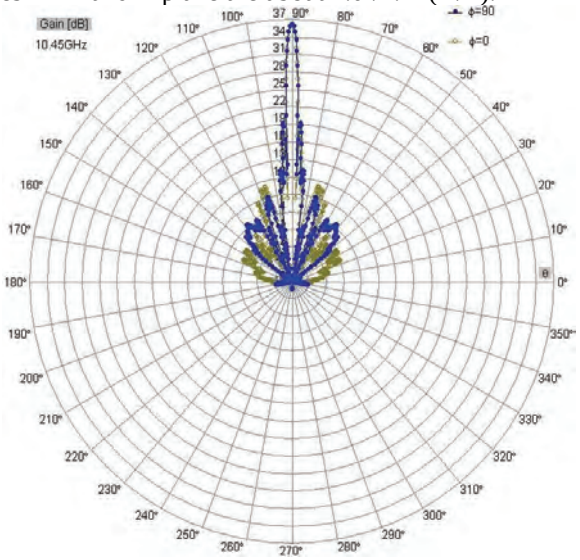


Fig. 5 *Spherical radiation patterns in H- and E-plane at f= 10.45GHz*

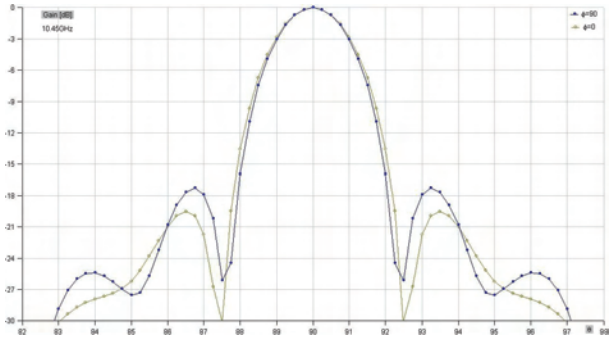


Fig. 6 *Main and first two side lobes in H- and E-plane at f= 10.45GHz*

Fig.7 (H-plane, Φ=90) and *Fig.8* (E-plane, Φ=0) present radiation patterns at the center as well as at lower and upper boundary of the operating frequncy range, in the left half-space of the antenna with ETSI RPE masks of class 1 and 2, [5], for frequency range 1. We can see that antenna belongs to class 1. For transferring the antenna into class 2, an absorber in the antenna shield is needed.

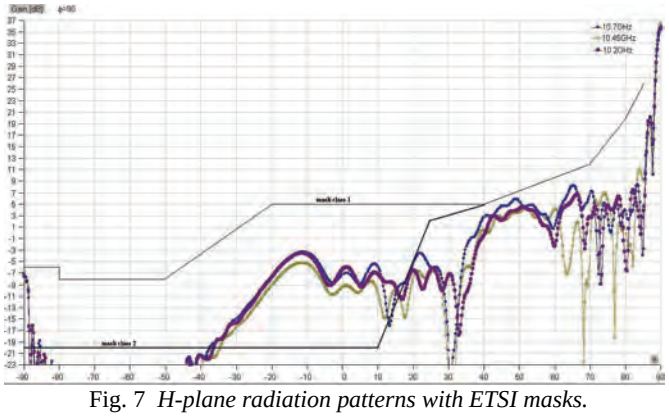


Fig. 7 *H-plane radiation patterns with ETSI masks.*

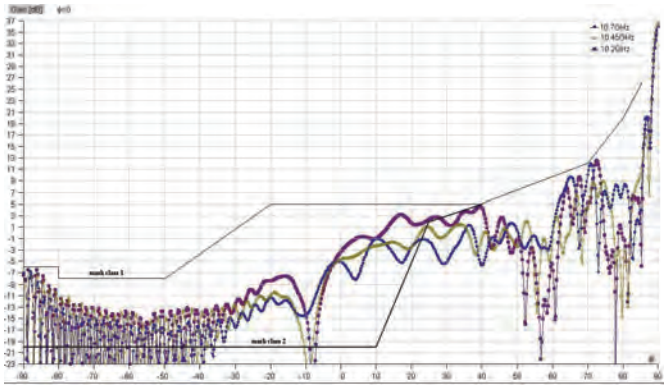


Fig. 8 *E-plane radiation patterns with ETSI masks.*

IV Technical Characteristics of the Antenna

Based on results of the analyses estimated technical characteristics of the antenna are:

- type: standard dual-reflectors Cassegrain antenna with 0.9m diameter, operating at 10.5GHz.
- manufacturer: IMTEL-Komunikacije a.d. Beograd
- parabolic reflector diameter D =0.9m
- reflector focal length/ diameter ratio (F/D)=0.4
- oparating frequency range: (10.2 - 10.7)GHz -ETSI frequency range 1 (3 -14GHz)
- G_{min}= 36.1dBi, ETSI Gain category 2 (high co-polar gain)
- ETSI RPE class 1, use in networks where there is a low interference potential
- XPD= 27dB, ETSI XPD category 1 (standard XPD)
- F/B =43dB
- FSLA = 16dB
- VSWRmax= 1. 38 (RL_{min}=16dB)
- HPBW=2.1º
- polarization: single or dual (H and/or V)
- inter port isolation ≥ 35dB for dual polarization
- cut off frequency $f_{cut\ off}$ =8.58GHz
- antenna input/output: 1/2 SMA 50Ω- connectors
- reactive near-field region: up to 2.1m in front of the antenna
- far-field Fraunhofer region >56.7m
- dimensions (mm) 950x950x700
- mass =20kg
- antenna color: white
- white aerodynamic cover
- cover attenuation: 1 dB
- tower / adapter requirements:
  - adapter (tube) ϕ 89
  - adapter deviation from the vertical position ≤0.2° (due to XPD)
  - under all conditions main beam axis deviation should be ≤0.3HPBW=0.6°
- mechanism for left-right azimuth tuning - coarse: 360°, fine: 10°
- mechanism for elevation tuning ±15° (optionally ±30°)
- the antenna operates under following environmental conditions:
  - temperature: from –45˚C to +60˚C
  - relative humidity: 100%
  - industrial atmosphere
  - UV radiation
  - wind speed: 200km/h with 25mm radial ice
  - rain, snow, hail, frosting, salt mist, condensation, fog, insects, birds
- possibility of ODU mounting on the back side of the antenna, in the link housing
- referent ETSI document EN 300833

V Conclusion

The paper presents analysis and simulation of standard axially symmetric dual-reflectors Cassegrain antenna with a minimal blockade, linear polarization (horizontal and/or vertical), having diameter of 0.9m and operating in the frequency range from 10.2 to 10.7GHz.

According to ETSI standard ETS 300 833 (antennas used in point-to-point digital radio relay systems), analyzed antenna belongs to:

1. Frequency range 1 (3-14GHz)
2. G category 2 (high co-polar gain G_{min} ≥32dBi)
3. RPE class 1 (antennas for use in networks where there is a low interference potential)
4. XPD category 2 (standard cross-Polar Discrimination)

Acknowledgements

Results presented in this paper are obtained during ralization of the project partialy finaced by Ministry of science of Republic of Serbia.

References

[1] B.Kolundžija, T.Sarkar, J.Ognjenović, R.Harison, *WIPL–D Pro V.5.1 - Electronic modeling of composite wire and plate structure*, Artech House, 2004.

[2] P. W. Hannan, “Microwave Antennas Derived from the Cassegrain Telescope”, *IRE Transactions on Antennas and Propagation*, March 1961, pp 140-153.

[3] K. A. Jensen, “Cassegrain Systems”, in A. W. Rudge, K. Milne, A. D. Olver, P. Knight (eds.), *The Handbook of Antenna Design*, *Volume 1* , London, Peter Peregrinus, 1982, Section 3.2.

[4] C.Granet, “Designing Axially Symmetric Cassegrain or Gregorian Dual-Reflector Antennas from Combinations of Prescribed Geometric Parameters”, *IEEE Antennas & propagation Magazine*, 40, 2, April, 1998, pp.76-82.

[5] ETS 300 833: Antennas used in point-to-point DRRSoperating in the frequency band 3 to 60 GHz

This page intentionally left blank.

POSTER SESSION PO III

PO III - Radio Communications, Microwave Technique and Antennas

Comparative Study of Circularly Polarized Microstrip Arrays Distinguished by their Feeding System

Georgi S. Kirov[1] and Desislava P. Mihaylova[2]

***Abstract* – A comparative study of 4x4 circularly polarized microstrip arrays designed to operate within the Ku-band is accomplished. Two different feeding networks combining in-phase excitation of the sub-arrays and sequential rotation principle are investigated. For the purpose, two antenna models distinguished by their feeding system are designed and comparative simulation results are presented. The main electrical features of the arrays and the benefits due to feed line modification are estimated. The final array structures employ inexpensive substrates and have the advantage of simplicity and optimal geometry.**

***Keywords* – Aperture coupled microstrip antenna, Circularly polarized microstrip array, Sequential rotation principle, Simulation results.**

I. INTRODUCTION

A comparative study of the characteristics of two 4x4 circularly polarized (CP) microstrip arrays is presented in this paper. The arrays are designed to operate in the Ku-band at around 12 GHz. The investigated structures utilize 2x2 sub-arrays of circularly polarized antenna elements and differ with their feeding networks. The obtained simulation results for two optimized 4x4 CP arrays are summarized and compared in order to ascertain their advance features.

It should be pointed out that most scientific references emphasize on the advantages of sequentially-rotated arrays (SRA) in terms of both Axial Ratio and impedance bandwidths [1 - 5]. In the current study, however, a combination of both sequential rotation principle and in-phase excitation of microstrip sub-arrays is proposed. Thus, a more comprehensive esteem of the radiation characteristics of circularly polarized microstrip arrays and design limitations as well, is provided. The EM 3D simulator CST Microwave Studio 2008 is employed.

II. ARRAYS DESIGN

The arrays investigated herein consist of circularly polarized 2x2 sub-arrays with sequential rotation (CP SRA) utilizing aperture-coupled microstrip elements. The geometry of the sub-array is displayed in Fig. 1.

The coupling cross-apertures etched in the ground of the two-layer structure have unequal shoulders defined according to the middle aperture length La as La1=2La.Ks/(1+Ks) and La2=2La/(1+Ks), so that 2La=La1+La2 when Ks=1.

.

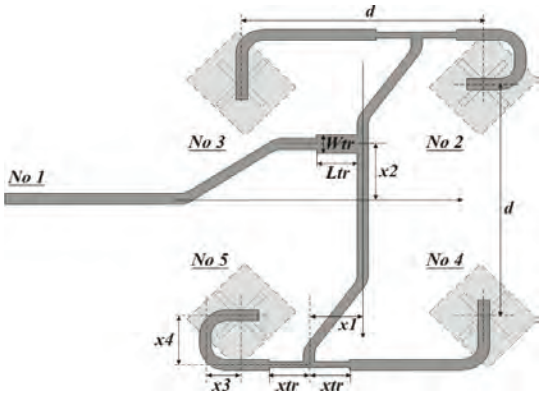


Fig. 1. Geometry of a circularly polarized 2x2 sub-array with sequential rotation (CP SRA) employing aperture-coupled microstrip elements.

Parameters of the optimized CP SRA 2x2 (Fig. 1): Antenna Size a = 53 mm, Array base d = 18 mm, Patch Length Lp = 6.1 mm, Patch Ratio Kp = Lp/Wp=1.14, Slot Ratio Ks = La1/La2 = 1.06, Stub Length Ls = 1.2 mm, Middle Aperture Length La = 4.8 mm, Impedance transformer Length Ltr(35.36Ω) = 3 mm, Impedance transformer Width Wtr = 1.3 mm, Distances xtr = 3 mm, x1 = 4 mm, x2 = 4.3 mm, x3 = 2.7 mm, x4 = 3.8 mm, Feed line Width Wf(50Ω) = 0.845 mm, Wf(70.7Ω) = 0.35 mm.

Substrates: 1) Patch substrate, Taconic TLX-7: $\varepsilon rp=2.60$, $\tan\delta p=0.0019$, Substrate thickness hp=1.575 mm, Cladding thickness tp=0.035 mm; 2) Feed substrate, Taconic RF-60A: $\varepsilon rf=6.15$, $\tan\delta f=0.0028$, Substrate thickness hf =0.635 mm, Cladding thickness tf=0.0175 mm;

The principle of sequential rotation is realized with 90^o rotation of the patches and corresponding coupling slots, and 90^o phase shift in the feeding part below the radiating patches. Particular phase shift is achieved due to different feed-line lengths, determined due to simulation so that to achieve optimal AR-bandwidth. Two types of bends 50Ω/2 x 70.7Ω and 35.36Ω/2 x 50 Ω are used in the feeding network. The choice of the first is explained with the lower susceptance, leading to a broader impedance bandwidth. The second bend is considered structurally suitable. The simulation approach as part of the current design procedure allows determination of the exact placement and dimensions of the applied transformers.

The main electrical characteristics of the optimized 2x2 CP sub-array are listed in Table 1. The bandwidth of the sub-array is limited by Directivity and ranges up to 10.98 %.

[1]Georgi S. Kirov is with the Department of Radio Engineering at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: gkirov@abv.bg

[2]Desislava P. Mihaylova is with the Department of Radio Engineering at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: de_c@abv.bg

TABLE I
SUB-ARRAY ELECTRICAL CHARACTERISTICS

Electrical characteristic	2x2 Sub-array
Impedance Bandwidth	
fmin / fmax [GHz]	10.66 / 12.55
fo [GHz]	11.6
bw %	16.29
Directivity Bandwidth	
fmin / fmax [GHz]	10.75 / 12
fo [GHz]	11.375
bw %	10.98
AR-bandwidth and characteristics within it	
$fmin_{AR}$ / $fmax_{AR}$ [GHz]	10.73 / 12.65
fo_{AR} [GHz]	11.69
bw_{AR} %	16.42
BRmin / BRmax [dB]	-25.0 / -11.76
Gmin / Gmax [dB]	8.15 / 11.15
Dmin / Dmax [dB]	8.9 / 11.9

Using the above mentioned sub-array another two 4x4 CP arrays distinguished by their feeding system are designed – Fig. 2 and Fig. 3. The first one is fully constructed on the sequential rotation principle, adjusting the radius r4 for appropriate phase shift. In the second in-phase feeding is used.

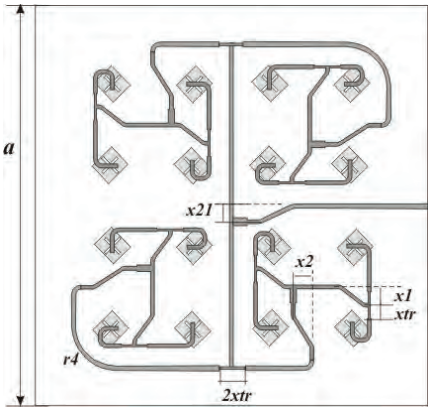


Fig. 2. Geometry of the circularly polarized microstrip array with sequentially rotated sub-arrays (CP ARRAY 1).

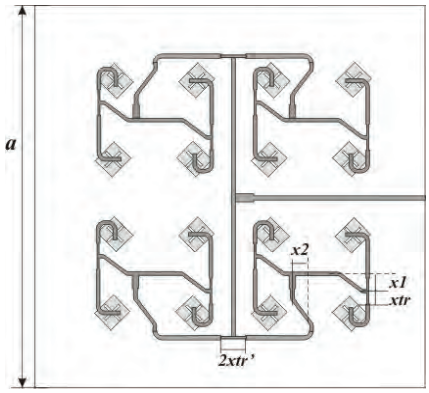


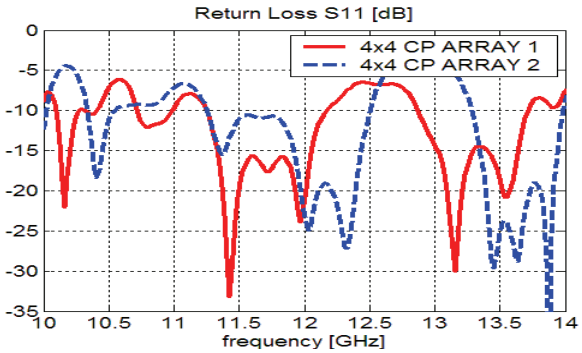
Fig. 3. Geometry of the circularly polarized microstrip array with in-phase excited sub-arrays (CP ARRAY 2).

Parameters of 4x4 CP ARRAY 1 (Fig. 2): Antenna Size a = 89 mm, Patch Length Lp = 6.1 mm, Patch Ratio Kp = 1.14, Slot Ratio Ks = 1.06, Stub Length Ls = 1.2 mm, Middle Aperture Length La = 4.8 mm, Transformer Length Ltr = 3 mm, Transformer Width Wtr = 1.3 mm, Distances xtr = 3 mm, x1 = 4 mm, x2 = 4.3 mm, x21 = 3.4 mm, x3 = 2.7 mm, x4 = 3.8 mm, radius r4 = 10 mm.

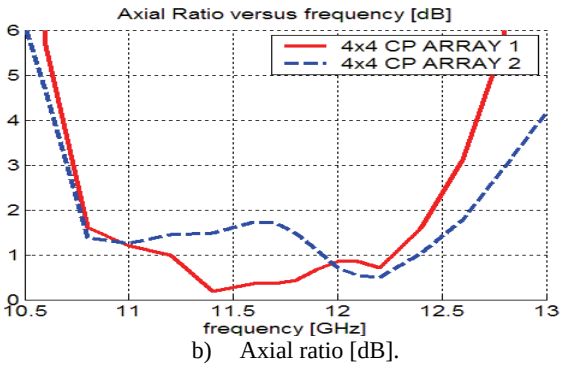
Parameters of 4x4 CP ARRAY 2 (Fig. 3): Antenna Size a = 89 mm, Patch Length Lp = 6.1 mm, Patch Ratio Kp = 1.17, Slot Ratio Ks = 1.06, Stub Length Ls = 1.2 mm, Middle Aperture Length La = 4.8 mm, Transformer Length Ltr = 3 mm, Ltr1 = 4 mm, Transformers Width Wtr = 1.3 mm, Distances xtr = 3 mm, xtr' = 2 mm, x1 = 4 mm, x2 = 4.3 mm, x3 = 2.7 mm, x4 = 3.8 mm.

III. COMPARATIVE STUDY OF ARRAYS

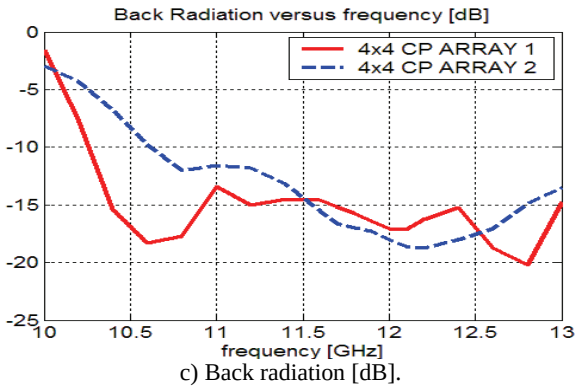
Comparative simulation results for the two 4x4 CP arrays are shown in Fig. 4. The arrays have been optimized in terms of AR-bandwidth due to CST Microwave Studio 2008 [6].



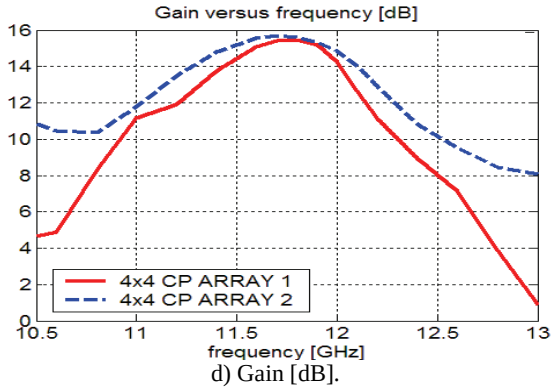
a) Return loss S11 [dB].



b) Axial ratio [dB].



c) Back radiation [dB].



d) Gain [dB].

Fig. 4. Characteristics of the CP sequentially rotated 2x2 sub-array: a) Return loss S11 [dB]; b) Axial ratio AR [dB]; c) Back radiation BR [dB]; d) Gain [dB] versus frequency.

TABLE II
ELECTRICAL CHARACTERISTICS OF ARRAY 1

Electrical characteristic	4x4 CP Array 1
Impedance Bandwidth	
fmin / fmax [GHz]	11.25 / 12.19
fo [GHz]	11.72
bw %	8.0
Directivity Bandwidth	
fmin / fmax [GHz]	11.25 / 12.13
fo [GHz]	11.69
bw %	7.52
AR-bandwidth and characteristics within it	
$fmin_{AR}$ / $fmax_{AR}$ [GHz]	10.73 / 12.6
fo_{AR} [GHz]	11.665
bw_{AR} %	16.0
BRmin / BRmax [dB]	-17.0 / -14.5
Gmin / Gmax [dB]	11.22 / 12.14
Dmin / Dmax [dB]	11.25 / 12.13

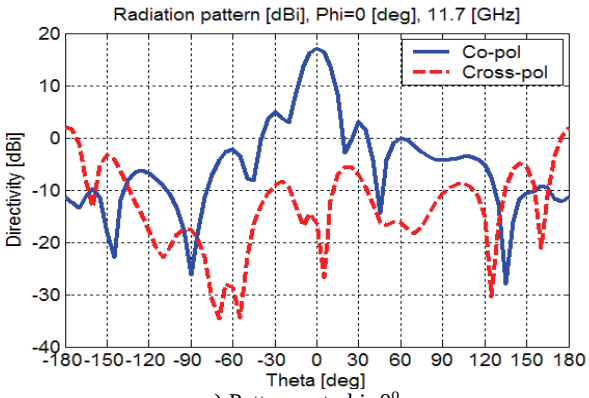
TABLE III
ELECTRICAL CHARACTERISTICS OF ARRAY 2

Electrical characteristic	4x4 CP Array 2
Impedance Bandwidth	
fmin / fmax [GHz]	11.25 / 12.5
fo [GHz]	11.875
bw %	10.53
Directivity Bandwidth	
fmin / fmax [GHz]	11.12 / 12.23
fo [GHz]	11.675
bw %	9.5
AR-bandwidth and characteristics within it	
$fmin_{AR}$ / $fmax_{AR}$ [GHz]	10.7 / 12.81
fo_{AR} [GHz]	11.75
bw_{AR} %	17.87
BRmin / BRmax [dB]	-18.7 / -12.0
Gmin / Gmax [dB]	12.6 / 15.6
Dmin / Dmax [dB]	14.15 / 17.15

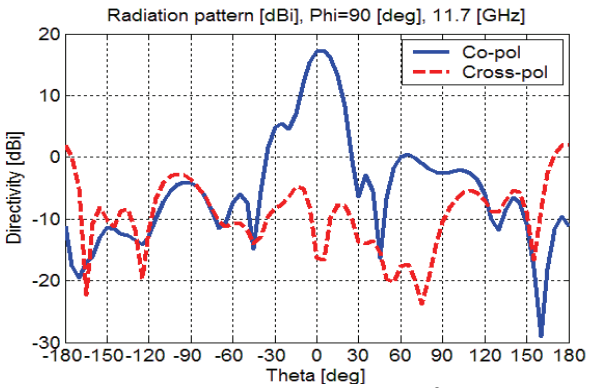
The obtained electrical characteristics can be compared by using Tables 2 and 3.

As seen in Fig. 4 a) the matching of the CP Array 2 is better within particular frequency range, but in the meantime it suffers at lower frequencies. Observing the Gain characteristic it is seen that the array with in-phase feeding (CP Array 2) has an advantage by means of Gain maximum and bandwidth (measured at level -3 dB from the maximum). The operational bandwidth of this antenna is fixed by the overlapping areas of impedance and Directivity bandwidths and reaches 8.3 %, while the relevant value of the other array is calculated 7.52 %. The Back radiation of the CP Array 2 distinguishes with one minimum, which equals -17 dB. The same characteristic of the CP Array 1 is broader with its values below -13 dB in the whole operational bandwidth.

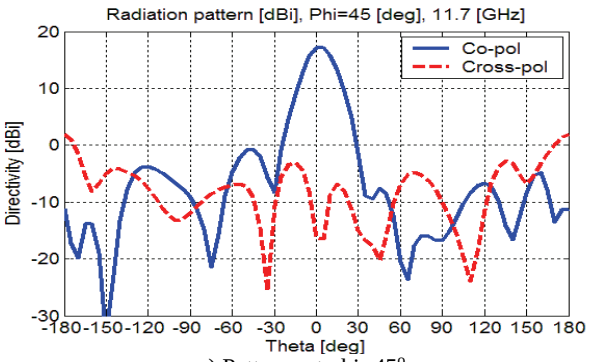
Co- and Cross-polarization patterns of the 16-element circularly polarized microstrip arrays are illustrated in Fig. 5 and Fig. 6.



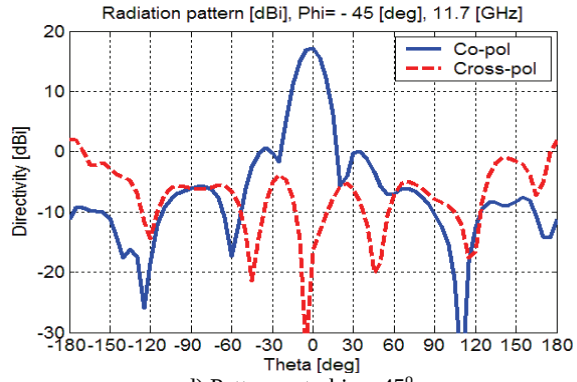
a) Patterns at phi=0°.



c) Patterns at phi=90°.

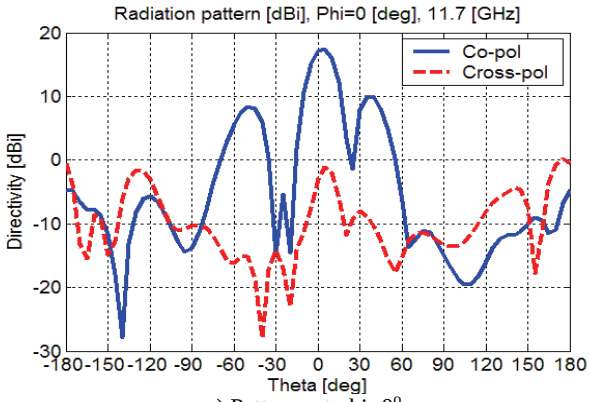


c) Patterns at phi=45°.

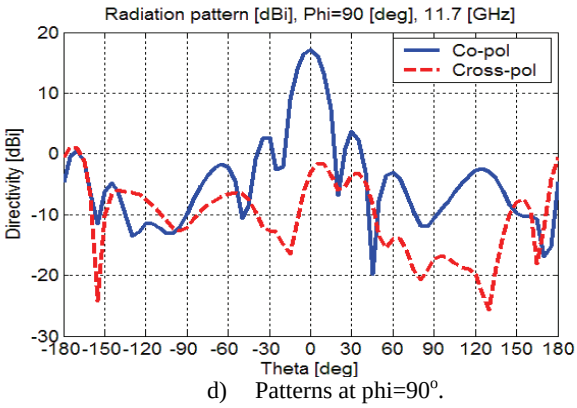


d) Patterns at phi= - 45°.

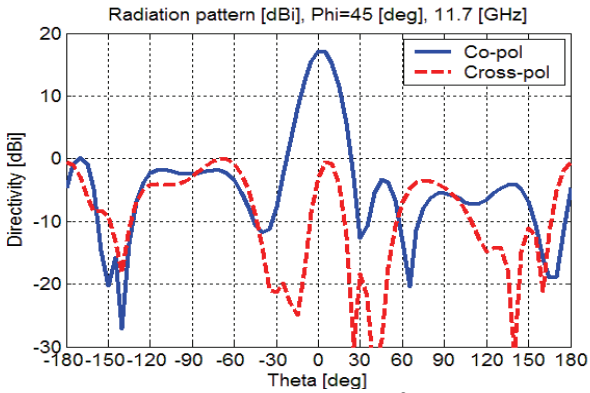
Fig. 5. Co- and cross-polarization patterns of the 4x4 CP ARRAY 1: a) Patterns at phi=0°; b) Patterns at phi=90°; c) Patterns at phi=45°; d) Patterns at phi= - 45°.



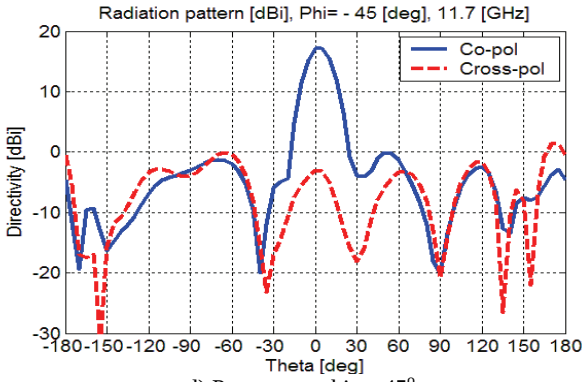
a) Patterns at phi=0°.



d) Patterns at phi=90°.



c) Patterns at phi=45°.



d) Patterns at phi= - 45°.

Fig. 6. Co- and cross-polarization patterns of the 4x4 CP ARRAY 2: a) Patterns at phi=0°; b) Patterns at phi=90°; c) Patterns at phi=45°; d) Patterns at phi= - 45°.

Both arrays exhibit right-hand circular polarization and have satisfying Co- and Cross-polarization characteristics in the principal planes (phi =0°, 90 °, 45 °, -45 °). Generally, the suppression of cross-polarization is more than 15 dB in comparison with the main lobe maximum.

IV. Conclusion

The characteristics of two circularly polarized 4x4 micrstrip arrays distinguished by their feeding system have been reported and compared. The simulation results show the advantage of the in-phase fed sub-array in terms of bandwidth and Gain. Both arrays are structurally simple and acquire good polarization properties. These arrays may be applied as receiver antennas in wireless communication systems.

Acknowledgement

The authors wish to acknowledge Bulgarian Ministry of Education and Science and Technical University of Varna for financial support under the Project № НП-7/2009.

References

[1] C. A. Balanis, *Antenna Theory: Analysis and Design*, John Willy & Sons, 3rd Ed., 2005.

[2] E. A. Soliman, S. Brebels, E. Beyne, G. A. E. Vandenbosch, "Sequential-rotation arrays of circularly polarized aperture antennas in the MCM-D technology", Inc. Microwave Opt. Technol. Lett. 44, pp. 581-585, 2005.

[3] Kumar, V.S. Srinivasan, V.V. Lakshmeesha, V.K. Pal, S., "Sequentially rotated microstrip array antenna at X-band for spacecraft", Applied Electromagnetics Conference, AEMC 2007. IEEE, pp. 1-4, 2007.

[4] H. Evans, A. Sambell, "Wideband 2 × 2 sequentially rotated patch antenna array with a series feed", Inc. Microwave Opt. Technol. Lett. 40, pp. 292-294, 2004.

[5] Y. Liu, M. Shahabadi, S. Safavi-Naeini, S. Suleiman, "Gain improvement of an array of sequentially rotated circularly polarized microstrip antennas using stacked parasitic patches", IEEE, pp. 614-617, 2003.

[6] CST Computer Simulation Technology, www.cst.com

Smart Antennas with Patch Elements. Modeling with Matlab

Vyara Y. Vasileva[1]

***Abstract* – This paper focuses on the antenna array modeling and adaptive estimation principle for antenna pattern creation of smart antenna composed of a number of uniformly distributed identical microstrip antenna elements. The utilities of adaptive beamforming and coupling compensation techniques in adaptive antenna modeling are depicted and accompanied by suitable simulation results. Numerical results are presented for the dependence of the radiation patterns of smart antennas on the properties of the estimation methods.**

***Keywords* – adaptive beamforming algorithm, antenna radiation pattern, patch element, smart antenna, uniform rectangular array.**

I. INTRODUCTION

In wireless communications larger capacity of the channel in the frequency-reused radio-communication system and better quality of the channel are the main conditions for the existing wireless system improving [1]. To meet these requirements, smart antennas with microstrip (patch) elements can be used.

A suitable adaptive array is a planar smart antenna because possesses the ability to scan the main beam in any direction of elevation and azimuth in 3-D space. In antenna array modeling, where size, cost, and performance are the main constraints, a number of uniformly distributed identical patch antennas may be required. These patch elements are low profile, low cost and simple to manufacture, mechanically robust, with a variety of impedance, polarization, and pattern characteristics [2].

In this paper, the least mean square (LMS) method and mutual coupling compensation technique are applied to antenna array modeling. The model descriptions are attended by simulation results obtained for a specific uniform rectangular array (URA).

II. THE RECTANGULAR MICROSTRIP ANTENNA ELEMENT MODELING. THE SMART ANTENNA PATTERN FORMATION

A. The rectangular microstrip antenna

The rectangular microstrip antenna is very easy to mathematical description and analysis. The patch element with a length L ($0.33\lambda_0 \le L \le 0.5\lambda_0$), illustrated in Fig. 1, consist of a very thin metallic patch with a thickness t ($t << \lambda_0$, λ_0 is the wavelength in the free space) placed on a dielectric substrate with a thickness h ($0.003\lambda_0 \le h \le 0.05\lambda_0$) above a perfectly conducting ground plane [3].

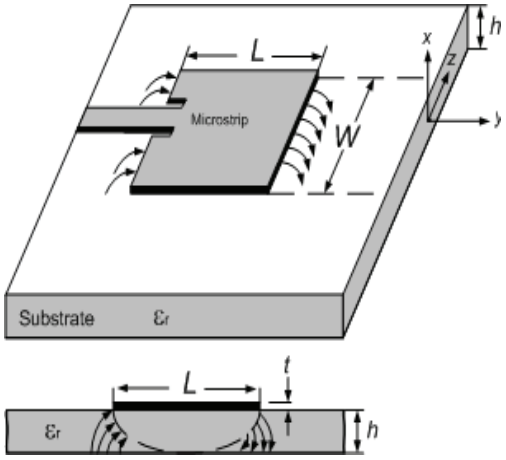


Fig. 1. Geometry of the rectangular microstrip element.

For the rectangular patch antenna design the most popular methods are: are used the transmission line model (TLM) and the cavity model. Here the TLM model is applied which gives accurate enough results avoiding complicated computations (like in the full-wave model). For the element modeling, the microstrip antenna is described as a configuration of two radiating slots with width *W*, height *h*, separated by a transmission line with length *L* into a dielectric with an effective dielectric constant ε_{reff} [3]. It is assumed that only the principal mode TM^{x}_{010} propagates in this line. For the design procedure is necessary to specify the substrate (ε_r - relative dielectric constant, *h*-height) and the resonant (operating) frequency f_r. The procedure has the following steps:

- width calculation:

$$W = \frac{1}{2 f_r \sqrt{\mu_0 \varepsilon_0}} \sqrt{\frac{2}{\varepsilon_r + 1}} \tag{1}$$

- effective dielectric constant:

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2\sqrt{1 + 12\,h/W}} \tag{2}$$

- length extension ΔL (Fig.2):

[1]Viara Y. Vasileva is with Department of Electrical Engineering, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: via_vas@abv.bg

$$\Delta L = 0.412\,h\frac{\left(\varepsilon_{r\,eff}+0.3\right)\left(W/h+0.264\right)}{\left(\varepsilon_{r\,eff}-0.258\right)\left(W/h+0.8\right)} \quad (3)$$

- effective microstrip length:

$$L = \frac{1}{2 f_r \sqrt{\mu_0 \varepsilon_0 \varepsilon_{reff}}} \qquad L_{eff} = L + 2\,\Delta L \quad (4)$$

-

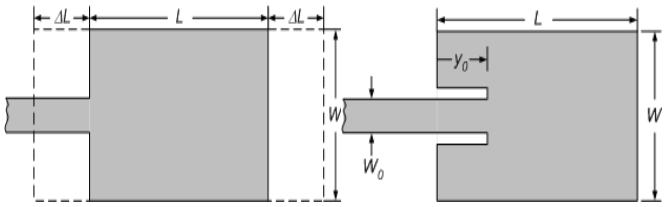


Fig. 2. The rectangular patch element modeling parameters.

In order to determine the input impedance we apply the following formula [3] (Fig.2):

$$R_{in} = \frac{1}{2\left(G_1 + G_{12}\right)}\ , R_{in}\left(y = y_0\right) = R_{in}\cos^2\left(\frac{\pi}{L}y_0\right) \quad (5)$$

where

$$G_1 = \frac{-2 + \cos\left(k_0 W\right) + \left(k_0 W\right)S_i\left(k_0 W\right) + \dfrac{\sin\left(k_0 W\right)}{k_0 W}}{120\pi^2} \quad (6)$$

is the conductance of a single radiating slot, k_0 is the free space phase constant, S_i (X) is an integral sine function, while G_{12} is the mutual conductance between the two radiating slots (often neglected in first approximation).

The next parameter of interest is the directivity of the patch antenna that can be calculated by the following expression [3]

$$D_{patch} = D_0\frac{2G_1}{G_1 + G_{12}} \quad (7)$$

where

$$D_0 = \frac{2\pi W/\lambda_0}{-2 + \cos\left(k_0 W\right) + \left(k_0 W\right)S_i\left(k_0 W\right) + \dfrac{\sin\left(k_0 W\right)}{k_0 W}} \quad (8)$$

is the directivity of a single slot. If we neglect the mutual coupling (in first approximation) the directivity of the antenna is $D_{patch} \approx 2D_0$.

B. The antenna array pattern creation

In the previous subsection, the singe patch element modeling procedure was represented. Because the single element radiation pattern is relatively wide and the directivity values is relatively low it is necessary to use a uniform rectangular array (URA) smart antenna to synthesize a required antenna pattern, to scan array beam in order to increase the directivity in the direction of SOI (signal of interest) and to decrease the directivity in the direction of SNOIs (signals not of interest).

The URA consists of N x M equally distributed identical microstrip antenna elements, is analyzed on the base of adaptive beamforming (ABF) estimation technique. The array element is modeled using the design procedure described above.

To achieve the desired array radiation pattern the LMS algorithm is used. It is applicable to estimate optimal weights of an antenna array. Briefly description of the method is presented below.

The expression of optimal weights is given by [4]

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\mathbf{g}(\mathbf{w}(n)) \quad (9)$$

where $\mathbf{w}(n+1)$ denotes a new computed weights vector at the (n+1)th iteration, μ is a gradient stepsize, and the array output is given by

$$y(\mathbf{w}(n)) = \mathbf{w}^H(n)\mathbf{x}(n+1) \quad (10)$$

where $\mathbf{x}(n+1)$ is an array signal vector computed at the (n+1)th iteration, and $y(\mathbf{w}(n))$ is an output signal.

In its standard form it uses an estimate of the gradient by replacing array correlation matrix **R** and correlation between array signals and reference signal *r* by their noisy estimates at the (n+1)th iteration [6]

$$\mathbf{g}(\mathbf{w}(n)) = 2\mathbf{x}(n+1)\mathbf{x}^H(n+1)\mathbf{w}(n) - 2\mathbf{x}(n+1)r^*(n+1) \quad (11)$$

where **g** is the gradient vector.

The error between the array output and the reference signal is given by [4]

$$\varepsilon(\mathbf{w}(n)) = r(n+1) - \mathbf{w}^H(n)\mathbf{x}(n+1) \quad (12)$$

and

$$\mathbf{g}(\mathbf{w}(n)) = -2\mathbf{x}(n+1)\varepsilon^*(\mathbf{w}(n)) \quad (13)$$

is the estimated gradient as a product of the error between the reference signal and the output of the array and the signals after the nth iteration.

In array configuration is important to take into account mutual coupling effect. It can be significant for the microstrip elements and the neglecting of this effect in the beamforming algorithm may produce inaccurate results. Here is applied a simple mutual coupling compensation technique [5], [6]. The new weights are chosen based on the matrix equation

$$\mathbf{w}_{comp} = \mathbf{C}^{-1}\mathbf{w} \quad (14)$$

where **C** is the coupling matrix determined by

$$\mathbf{C} = \mathbf{I} + \mathbf{Z}Z_L^{-1} \quad (15)$$

where Z_L is the load impedance of each element, **I** is the unit matrix, **Z** is a matrix with diagonal elements self-impedances, and off-diagonal elements – mutual impedances. In the analysis these elements can be easily computed by Matlab implementation [7].

III. Simulation Results

In this section simulation results are based on the theory described in Section 2.

Modeling results (mathematical calculations and Matlab simulations) are presented in Table I and Fig. 3. The design

TABLE I

Microstrip parameters	Mathematical calculation	Matlab simulation results
Physical width	4.9411 cm	4.9411 cm
Physical length	4.1373 cm	4.1356 cm
Effective length	4.3047 cm	4.3030 cm
Resonant input resistance	244.5795 omhs	244.7745 omhs
Feed point position	1.4504 cm	1.4504 cm
Directivity	5.0295	5.2118
Directivity [dB]	7.0152	7.1699

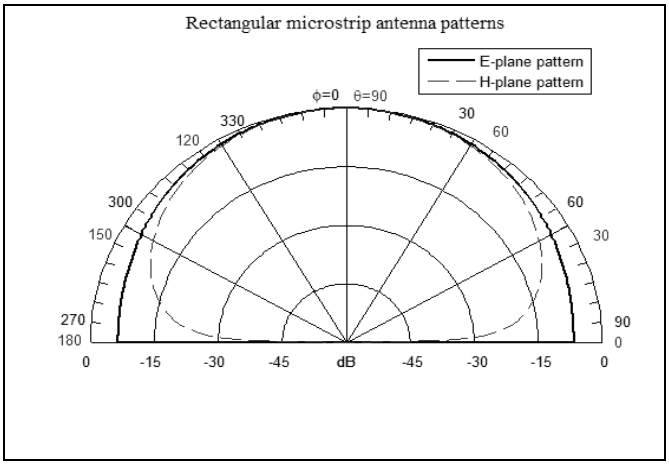


Fig. 3. The rectangular microstrip element radiation pattern.

procedure was realized for rectangular patch antenna operating at 2.4 GHz and using substrate with parameters $\varepsilon_r = 2.2$, $h = 0.1588$.

Simulation results, utilizing the LMS algorithm give precise data when adapt the beamforming pattern. To illustrate the ABF algorithm applicability for URA, we consider two cases where LMS algorithm is used: a) the URA with N=8 and M=8 elements and interelement spacing (the center-to-center separation between elements) $s_x = s_y = 0.45\lambda$ (Figs. 4 and 5); b) URA with N=8 and M=8 elements and interelement spacing $s_x = s_y = 0.45\lambda$ when using coupling compensation (Figs. 6 and 7). These figures present the Matlab simulation results. The URA is examined about following scenario: the SOI impinges from direction $(\theta = 135^\circ, \phi = 70^\circ)$ in the presence of one SNOI from direction $(\theta = 105^\circ, \phi = 95^\circ)$, that is an Additive White Gaussian Noise (AWGN) with a zero mean, and a variance 0.1. All simulation results are based on 100 times Monte Carlo runs. A stepsize is μ = 0.001 and a signal with uncoded BPSK modulation are used in the numerical examples to simplify the simulations. These figures illustrate the resulting beamforming pattern with respect to $\theta_0 = 0^\circ$. The results demonstrate its good performance, characterized by accurate estimation ability, and it is evident that after compensation the radiation pattern (Fig. 7) has better ability to distinguish the SOI from the SNOI.

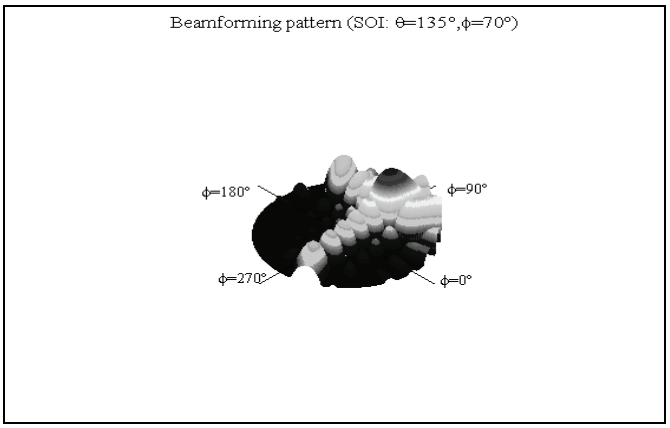


Fig. 4. Beamforming pattern 3D plot.

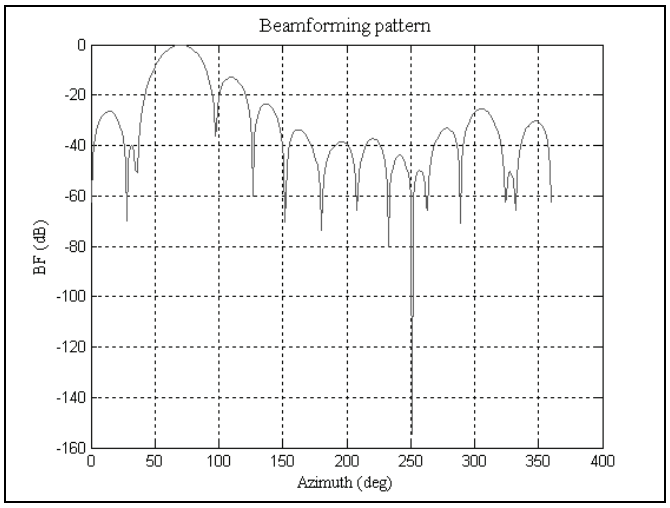


Fig. 5. Beamforming pattern 2D plot.

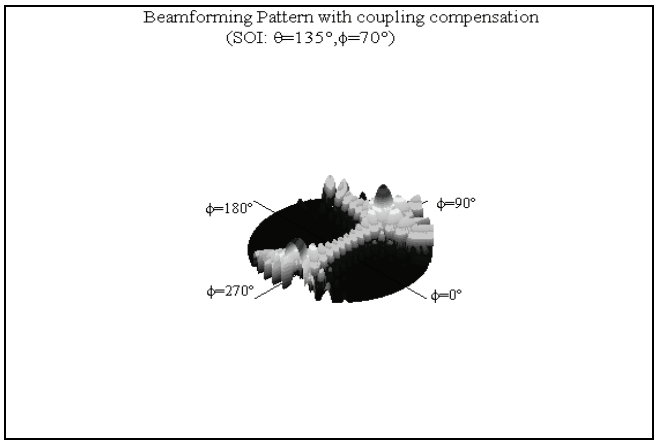


Fig. 6. Beamforming pattern after coupling compensation 3D plot.

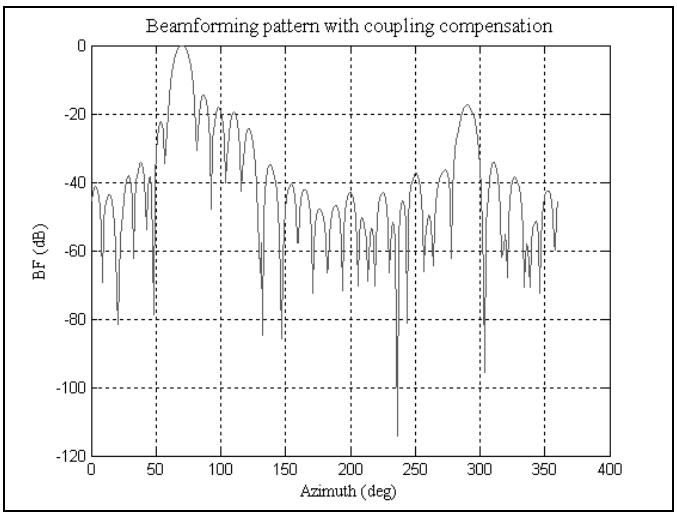


Fig. 7. Beamforming pattern after coupling compensation 2D plot.

IV. CONCLUSION

This paper investigates the issues for a single rectangular microstrip element and a uniform rectangular smart antenna array with patch elements modeling. The adaptive beamforming (ABF) and mutual coupling compensation were examined. The LMS algorithm is used after the ABF technique. As concerns to beamforming the examination array has an accurate and stable enough radiation pattern regarding both: desired signal - the beamforming pattern maximum to the SOI and interfering signals – deep nulls towards the SNOI.

Numerical simulation results are illustrated that the patch element design gives accurate results, and that the antenna geometry configuration with M=N=8 microstrip elements is an optimal scenario, because the ABF estimations are proved to be accurate and stable enough, and the ability of the smart antenna to reject undesired signals is affected by the type of elements, size and geometry of the antenna array. Consequently, it is observed that the array element choice and the smart antenna design impact on the overall wireless communication network efficiency.

The performance of the smart antenna (beamforming pattern) is examined in two cases: a) without a mutual coupling; b) with a mutual coupling compensation. Matlab programs are used for simulations. The numerical results presented here clearly display the importance of the coupling compensation.

ACKNOWLEDGEMENT

Project no. 21 "Improvement of the research potential in the area of engineering and information technology" in the frame of Program "Development of human resources" and Ministry of Education of Bulgaria.

Project in the frames of the Research Program, financed from Ministry of Education of Bulgaria.

REFERENCES

[1] S. Bellofiore et al., “Smart-antenna systems for mobile communication networks, part 1: overview and antenna design”, IEEE Antennas and Propag. Magazine, vol. 44, pp. 106-114, 2002.

[2] J. James, P. Hall, “Handbook of Microstrip Antennas”, vols. 1, 2, P.Peregrinus Ltd., London, 1989.

[3] C. A. Balanis, “Antenna Theory”, John Wiley & Sons, New York, 2005.

[4] J. Foutz et al, “Adaptive eigen-projection beamforming algorithms for 1-D and 2-D antenna arrays”, IEEE Antennas and Wireless Propag. Letters, vol. 2, pp. 62-65, 2005.

[5] I.Gupta, A. Ksienski: “Effect of mutual coupling on the performance of adaptive arrays”, IEEE Trans. Antennas and Propag., vol. 31, pp. 785-791, 1983.

[6] H. Steyskal, J. Herd: “Mutual coupling compensation in small array antennas”, IEEE Trans. Antennas and Propag., vol. 38, pp. 1971-1975, 1990.

[7] K. Lonngren, S. Savov, R. Jost, “Fundamentals of Electromagnetics with Matlab”, (2nd ed.), SciTech, 2007.

Design of a Compact Hexagonal Monopole Antenna for Ultra-Wideband Applications

A. A. Shaalan[1], and M. I. Ramadan[2]

***Abstract* – This paper presents two design compact hexagonal monopole antennas for ultra-wideband applications. The two antennas are fed by a single microstrip line . The Zeland IE3D version 12 is employed for analysis at the frequency band of 4 to 14 GHz which has approved as a commercial UWB band. The experimental and simulation results exhibit good agreement together for antenna 1. The proposed antenna1 is able to achieve an impedance bandwidth about 111%. The proposed antenna2 is able to achieve an impedance bandwidth about (31.58%) for lower frequency and (62.54%) for upper frequency bandwidth. A simulated frequency notched band ranging from 6.05 GHz to 7.33 GHz and a measured frequency notched band ranging from 6.22 GHz to 8.99 GHz are achieved and gives one narrow band of axial ratio (1.43%).**

***Keywords* – Ultra-wideband antenna, Circular polarization, Notch band.**

I. INTRODUCTION

Ultra Wideband (UWB) technology has become a major interest to researchers and scientists. The commercial usages of frequency band from 3.1 GHz to 10.6 GHz, was approved by Federal Communications Commission (FCC) in 2002 [1]. However, there is always an increasing demand for smaller size, and greater capacities and transmission speeds, which will certainly require more operating bandwidth in the near future. In the last few years, researchers have investigated several kinds of microstrip slot and printed antennas for UWB applications [2–12]. Ultra-wideband (UWB) communication systems are currently under investigation and have been widely adopted in commercial and military domains. In this paper, a compact planar monopole antennas for UWB applications are proposed, The proposed antennas not only occupy a small size but also preserve a very single structure which is easy to be fabricated. The input impedance matching over a wide frequency range is achieved (4 GHz–14 GHz) .

Two design compact hexagonal monopole antenna for ultra-wideband applications are:

Antenna 1: Hexagonal monopole antenna,

Antenna 2: Hexagonal monopole antenna with asymmetrical U-slot

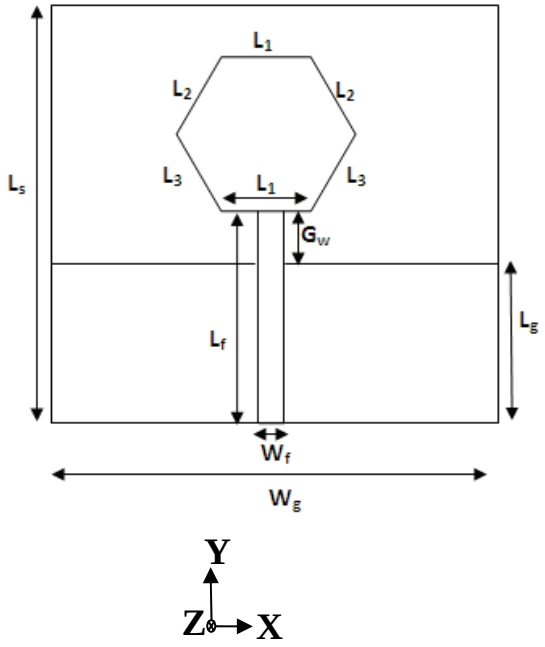
The detailed design and experimental results are presented and discussed below.

.

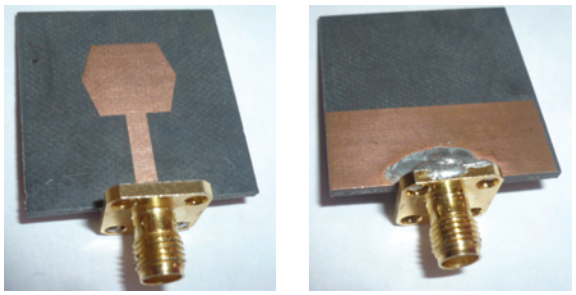
[1]Email: shaalan_zag2010@yahoo.com
[2] Email: mohram_2002@yahoo.com

II. ANTENNA GEOMETRY

The configuration and the photo of the proposed UWB antennas are illustrated in Figure 1 and Figure 2, respectively. The proposed antenna have a compact size of 30×30mm and it is printed on conventional RT5880 substrate with thickness of 1.57mm, and relative permittivity =2.2.The radiating element is a hexagonal patch and the ground is a rectangular. The radiator and the 50Ω feed line are printed on the same side of the substrate and the ground plane is located on the other side. An asymmetrical U-slot is used to obtain dual circular polarization for ultra wideband applications.

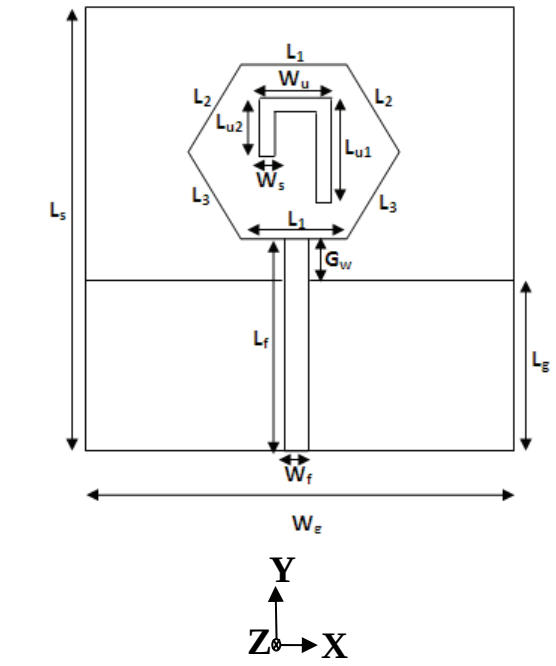


(a)

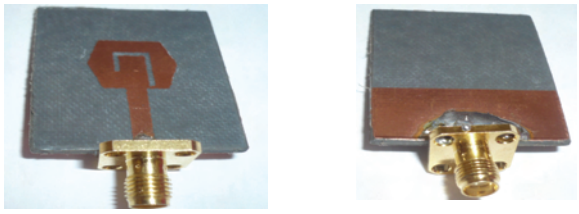


(b)

Fig. 1. Antenna 1: Hexagonal monopole antenna (a) Antenna geometry (b) Photograph



(a)



(b)

Fig. 2. Antenna 2: Hexagonal monopole antenna with asymmetrical U-slot (a) Antenna geometry (b) Photograph

The geometry and photograph of the proposed antenna1 with its parameters is depicted in Figure 1. The antenna is located in the xy plane and the normal direction is parallel to the z axis. The dimensions of the proposed antenna1 including the substrate L_s X W_g = 30mm x 30mm ,The radiation element is a hexagonal patch with dimensions L_1 = 10 mm, L_2 = 5.5 mm, L_3 = 6 mm, A 50Ω microstrip feed line with width of W_f =3 mm and the length of L_f = 13 mm, The gap G_w= 1 mm is the distance between the ground plane on the back side and the hexagonal patch on the front size. The ground plane size of L_g x W_g = 12 mm X 30 mm.The geometry of the proposed antenna 2 with its parameters is depicted in Figure 2. The dimensions of proposed antenna in figure 2 is similar to dimensions of the proposed antenna in figure 1 except the radiation element is a hexagonal patch with asymmetrical U-slot with dimensions L_1 = 10 mm, L_2 = 5.5 mm, L_3 = 6 mm ,L_{u1}= 8.5 mm, L_{u2}= 5mm, W_u = 6 mm, W_s = 1 mm.

III. MEASUREMENT AND SIMULATION RESULTS OF ANTENNA 1

The structures are simulated with IE3D which utilized the moment method for electromagnetic computation.

A. Return Loss

The measured and simulated return loss S_{11} of antenna 1 are shown in Figure 3. From the Figure, it is predicted that there is a good agreement between numerical and experimental results. The experimental and the simulated band frequencies from 3 GHz to 14 GHz are predicted, The slight frequency shift and discrepancy are achieved due to the probable deviation on the substrate permittivity at high frequency, Otherwise, a good agreement is achieved.

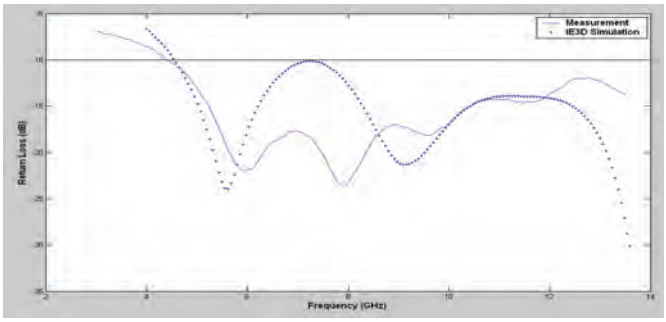


Fig. 3. Simulated and measurement return loss of Antenna 1

B. Gain

The Simulated maximum gain of the proposed antenna 1 is performed by using IE3D and presented in Figure 4. The antenna gain varies from 2.32 dB to 4.4 dB over the operating UWB frequency range. It can be concluded that the gain variation is not less than 1.25 dB over the entire operating frequency range from 4 to 14 GHz.

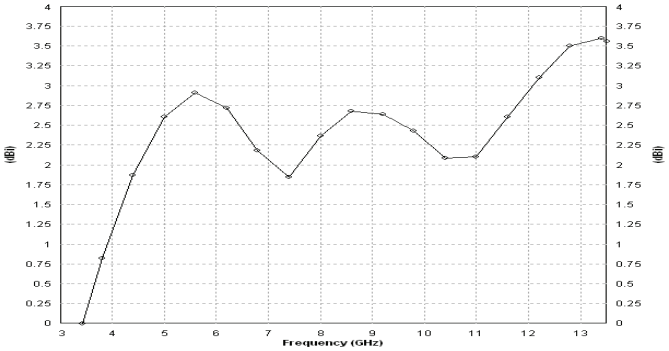


Fig. 4. Simulated maximum gain of Antenna 1

IV.MEASUREMENT AND SIMULATION RESULTS OF ANTENNA 2

A. Return Loss

The measurement and Simulated return loss S_{11} of antenna 2 are shown in Figure 5. The simulated impedance bandwidth for return loss (RL) less than -10dB (|S11| ≤ -10 dB) is obtained. It is seen that the impedance bandwidth for return loss of less than -10 dB ranges from 4.4 GHz to 14 GHz, in which a simulated frequency notched band ranging from 6.05 GHz to 7.33 GHz is achieved, a measured frequency notched band ranging from 6.22 GHz to 8.99 GHz is achieved. Due to this notch frequency, there are two bandwidth of the UWB antenna. The Simulated Lower frequency bandwidth (BW1) is

starting from 4.4 GHz to 6.05 GHz and the simulated upper frequency bandwidth (BW2) is starting from 7.33 GHz to 14 GHz. The measured Lower frequency bandwidth (BW1) is starting from 5.19 GHz to 6.22 GHz and the measured upper frequency bandwidth (BW2) is starting from 8.99 GHz to 14 GHz.

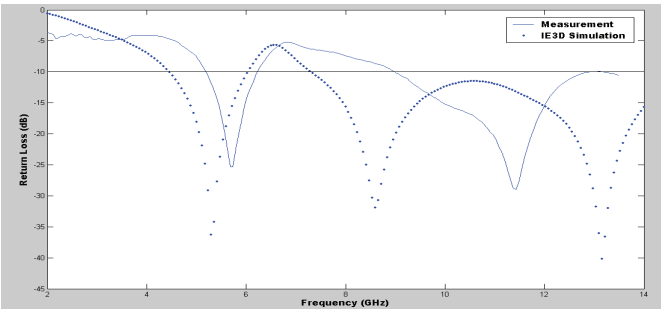


Fig. 5 Simulated and measurement return loss of Antenna

B. *Axial Ratio*

The simulated axial ratio (AR) of antenna 2 is shown in Figure 6. The simulated axial ratio bandwidth (AR) ≤ 3 dB provides single circular polarization. It is seen that the (AR) bandwidth have ranges (6.92 – 7.02) GHz , BW = (1.43%) .The circularly polarized (CP) antenna could have many different types and structures where the basic operation principle is to radiate two orthogonal field components with equal amplitude but in phase quadrature. The CP is generated by the unequal arms of the U-slot (Asymmetrical U-slot). Also Antenna 2 gives Left Hand Circular Polarization (LHCP) at one frequency 6.9 GHz which presented in Figure 7, [13, 14].

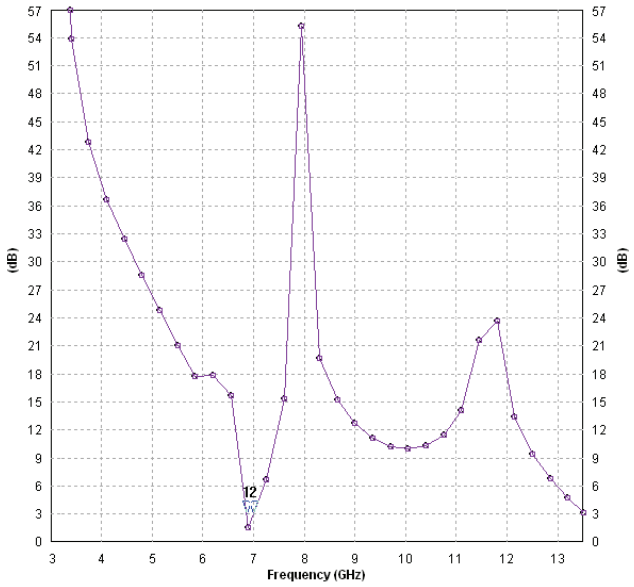


Fig. 6. Simulated AR vs. Frequency of Antenna 2

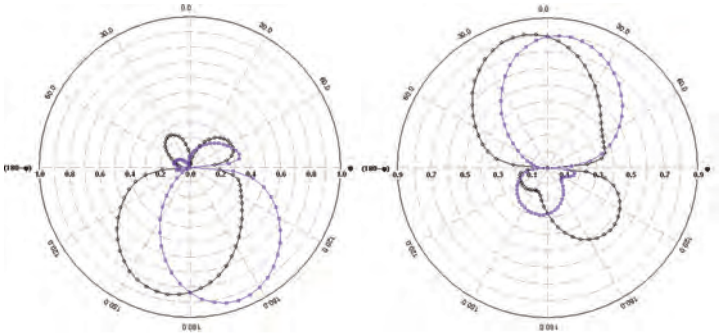


Fig. 7. Radiation Pattern at 6.9 GHz of Antenna 2

C .Gain

The Simulated maximum gain of the proposed antenna 2 is performed by using IE3D and presented in Figure 8. The antenna gain varies from 2.2 dB to 3.6 dB over the operating UWB frequency range. It can be concluded that the gain variation is not less than 1.38 dB over the entire operating frequency range from 4 to 14 GHz.

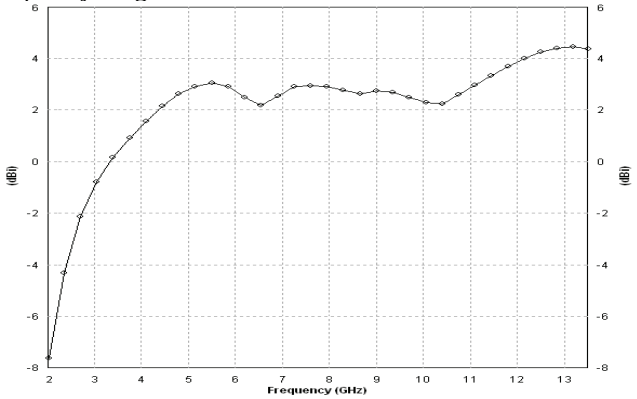


Fig. 8. Simulated maximum gain of Antenna 2

CONCLUSION

This paper presented two design of a compact hexagonal monopole antenna for ultra-wideband applications. Antenna 1 (Hexagonal monopole antenna) gives an return loss bandwidth about 111% with acceptable gain over the entire frequency band of operation. Antenna 2 (Hexagonal monopole antenna with asymmetrical U-slot) gives the lower frequency bandwidth (BW1) is starting from 4.4 GHz to 6.05 GHz (31.58%), the upper frequency bandwidth (BW2) is starting from 7.33 GHz to 14 GHz (62.54%) and the frequency notched band ranging from 6.05 GHz to 7.33 GHz. Antenna 2 gives single axial ratio (AR) band have ranges (6.92 – 7.02) GHz, BW = (1.43%). Antenna 2 gives Left Hand Circular Polarization (LHCP) at one frequency 6.9 GHz. Also, the gain of this antenna is presented. It is noticed that the gain is accepted for the band of operation.

REFERENCES

[1] Anon, “FCC first report and rder on Ultra-Wideband Technology”, February 2002.

[2] Lin, C.-C. and H.-R. Chuang, "A 3–12 GHz UWB Planar Triangular Monopole Antenna with Ridged Ground-Plane", *Progress In Electromagnetic Research*, PIER 83, 307–321, 2008.

[3] Yang, Y., Y. Wang, and A. E. Fathy, "Design of Compact Vivaldi Antenna Arrays for UWB Through Wall Applications", *Progress In Electromagnetic Research*, PIER 82, 401–418, 2008.

[4] Dehdasht-Heydari, R., H. R. Hassani, and A. R. Mallahzadeh, “A New 2–18 GHz Quad- ridged Horn Antenna”, *Progress In Electromagnetic Research*, PIER 81, 183–195, 2008.

[5] Dehdasht-Heydari, R., H. R. Hassani, and A. R. Mallahzadeh, “Quad Ridged Horn Antenna for UWB Applications”, *Progress In Electromagnetic Research*, PIER 79, 23–38, 2008.

[6] Sadat, S., M. Houshmand, and M. Roshandel, “Design of a Microstrip Square-ring Slot Antenna Filled by an H-shape Slot for UWB Applications”, *Progress In Electromagnetic Research*, PIER 70, 191–198, 2007.

[7] Sadat, S., M. Fardis, F. G. Kharakhili, and G. R. Dadashzadeh, "A Compact Microstrip Square-ring Slot Antenna for UWB Applications”, *Progress In Electromagnetic Research*, PIER 67, 173–179, 2007.

[8] Hosseini, S. A., Z. Atlasbaf, and K. Forooraghi, “Two New Loaded Compact Planar Ultra-Wideband Antennas Using Defected Ground Structures”, *Progress In Electromagnetic Research B*, Vol. 2, 165–176, 2008.

[9] Yin, X.-C., C.-L. Ruan, C.-Y. Ding, and J.-H. Chu, “A Planar U Type Monopole Antenna for UWB Applications”, *Progress In Electromagnetic Research Letters*, Vol. 2, 1–10, 2008.

[10] Rajabi, M., M. Mohammadirad, and N. Komjani, “Simulation of Ultra Wideband Microstrip Antenna Using EPML-TLM”, *Progress In Electromagnetic Research Letters*, Vol. 2, 115–124, 2008.

[11] Wang, F.-J., J.-S. Zhang, X.-X. Yang, and G.-P. Gao, “Time Domain Characteristics of a Double-Printed UWB Dipole Antenna”, *Progress In Electromagnetic Research Letters*, Vol. 3, 161–168, 2008.

[12] B. Allen, M. Dohler, E. E. Malik, A .k .Brown, and D. J. Edwards, “Ultra-Wideband Antennas and Propagation for Communications, Radars ,and Imaging”, John Wiley and Sons ,LTD, 2007.

[13] A. A. Shaalan, "A Compact Broadband H -slot and Horizontal H-slot Patch Antenna for Circular Polarization”, International Journal of Infrared, Millimeter and Terahertz Waves, Springer, July, 2009.

[14] A. A. Shaalan, “Design and Simulation of a Single Feed Dual-Band Symmetrical/ asymmetrical U-Slot Patch Antenna”, XLIV International Scientific Conference, Information, Communication and Energy Systems and Technologies (ICEST 2009), Veliko, Tarnavo, Bulgaria,2009.

Design of an Anechoic Chamber at the Faculty of Electronic Engineering in Niš

Zoran Stanković, Bratislav Milovanović, Nebojša Dončov, Marija Agatonović

Abstract **- The project of an anechoic chamber realization at the Faculty of Electronic Engineering in Niš, is presented in this paper. The chamber is intended for measurements in the frequency range of 0.8 – 20 GHz. Selection of suitable geometry is performed, dimensions are calculated and the strategy of covering the chamber internal surface with RF absorbing material is analyzed. The most important results are discussed.**

Keywords **– Anechoic chamber, Antenna measurements, EM propagation, EMC characterization**

I. INTRODUCTION

Recent development in wireless communication systems requires efficient electromagnetic characterization. Methods providing high - accurate measurements of antenna radiation pattern, radar cross section, EM immunity and the leaking radiation intensity are the most important. By reducing interference signals, they must enable precise evaluation of the direct signal level at the receiver. All of the signals reflecting from the obstacles near the radiation source and the test region, reaching the receiver at the same time, are causing undesirable interference. However, this effect can also be a result of receiving a number of EM signals that belong to the other radiation sources, which radiate at the approximately the same frequency, as the observed source. Several methods for characterization of EM radiation are able to significantly reduce interference and to provide the required accuracy of the performed measurements. Depending on the conditions for EM characterization, these methods can be divided into outdoor and indoor measurements.

Outdoor measurements are performed far away from the urban settlements and sources of the interfering electromagnetic radiation [1]. In the wide area surrounding the radiating source and the test region, there should be no obstacles that could cause signal interference or diffraction. Outdoor measurements provide long distance between transmitting and receiving antenna that is essential condition for accurate measurements and good approximation of EM plane wave at the receiver. This is very important in cases when the large antennas are evaluated and/or antenna radiation pattern at high frequencies is measured. It is considered that the real spherical wave illuminating the antenna can be approximated by an ideal plane wave at the receiver, if the distance between antennas is equal to, or greater than the Rayleigh range.

The authors are with the Department of Telecommunications, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, E-mails: zoran.stankovic@elfak.ni.ac.rs, batam@pogled.net, nebojsa.doncov@elfak.ni.ac.rs and marija.agatonovic@elfak.ni.ac.rs

The boundary between the near-field and the far-field zone is given by

$$r \le \frac{2D^2}{\lambda}, \quad (1)$$

where D is the aperture of the largest antenna under test and λ is the wavelength of the EM wave. For instance, radiation pattern of a 0.5 m diameter parabolic antenna can be measured if the antenna is located at least 34 m away from the transmitting antenna at the operating frequency of 20 GHz. Outdoor measurements are strongly influenced by atmospheric conditions and undesirable radiations from distant transmitters or mobile communication base stations. Also, the flat areas used for performing these measurements are quite far from the laboratories and development centers where components for wireless communication systems are fabricated. On the other hand, specialized anechoic chambers are able to eliminate all of these disadvantages [1, 2].

The research at the Department of Telecommunications focuses on development of devices for application in wireless communication systems. Since there is no specialized anechoic chamber at universities in Serbia, required measurements cannot be performed. Finally, it is decided to design and construct an anechoic chamber at the Faculty of Electronic Engineering.

II. RF ANECHOIC CHAMBER

RF anechoic chamber is a shielded room intended to simulate a free-space environment. Reflected waves inside the chamber are attenuated by covering the chamber walls with special material having significant absorbing effects at the radio frequencies (RAM – Radio Absorbing Material). Although anechoic chambers of different geometry are present, the rectangular chamber is mostly used. It provides measurements of antennas and antenna arrays, radar cross section and EM compatibility. In addition to this type of chamber, there are chambers of specialized architectures which are utilized for measuring antennas (tapered chamber, TEM chamber and double horn chamber [2, 4]), radar cross section of the military missiles, and more.

At one end of the chamber, a source of electromagnetic radiation is placed, and at the other end, an antenna which has to be evaluated (receiving antenna), is located. The antenna data acquisition system enables one to plot the radiation pattern of antenna and determine its important parameters. The antenna to be tested is usually mounted on the pedestal that can be rotated under the control of the positioner. By rotating the antenna, the incidence angle of EM wave illuminating the antenna surface is changing and the level of

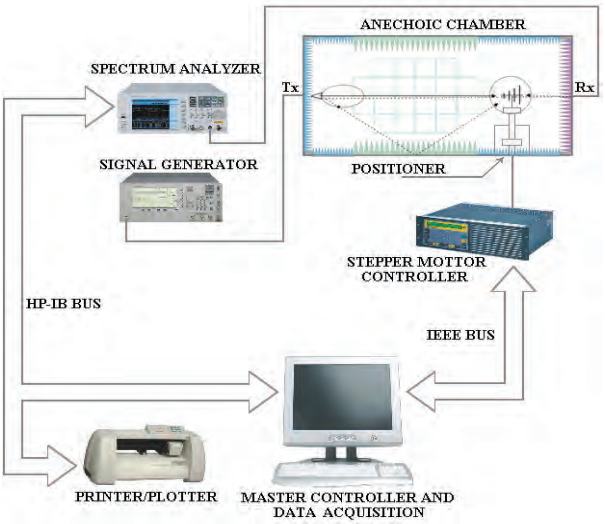


Fig. 1. Antenna data acquisition block diagram

the received signal is measured (Fig 1). The resulting signal amplitude data is plotted against the angle-of-arrival to provide a graphical representation of the signal components received by the AUT (Antenna Under the Test).

III. Radio Absorbing Material

In order to attenuate EM waves reflected from the walls, floor and ceiling, the entire internal surface of the anechoic chamber is covered with the radio absorbing material. There are two basic types of RAM elements: the first one is used to attenuate reflections at low frequencies while the second one absorbs high-frequency EM waves. Low-frequency RAM elements (30 MHz – 1000 MHz) are typically made of ferrite materials. At the higher frequencies (0.1 GHz – 50 GHz), absorbers made of the carbon impregnated polyurethane foam are used [2-6].

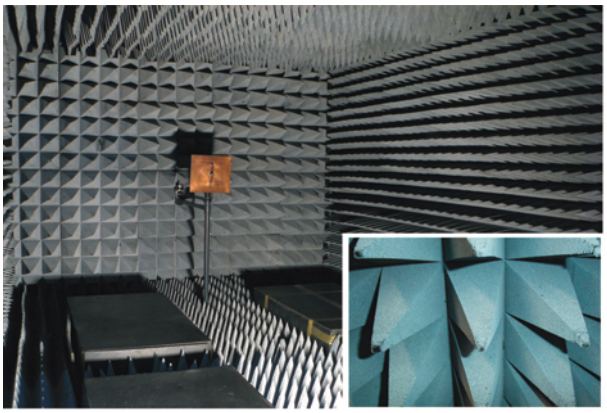


Fig. 2. The internal appearance of the anechoic chamber

Typical physical realization of high-frequency RAM elements is in the form of square based pyramids. A number of pyramids are fastened to the square panel which is also made of the same radio absorbing material. Depending on the manufacturer, standard panel dimensions are 0.5 m x 0.5 m or 0.61 m x 0.61 m. The height of the pyramids depends on the

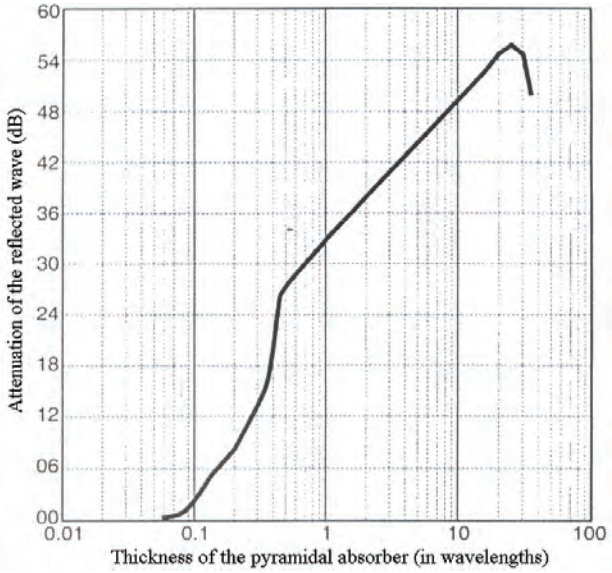


Fig. 3. Attenuation of the reflected wave as a function of the thickness of pyramidal RF absorber at normal incidence

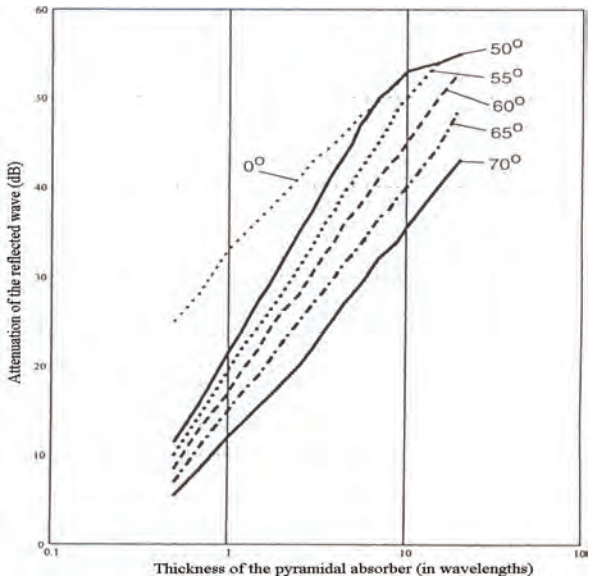


Fig. 4. Attenuation of the reflected wave as a function of the thickness of pyramidal RF absorber at various incidence angles

lowest operating frequency and the amount of the absorption required. Generally, RAM elements have good performance when the absorber thickness is greater than the wavelength of the incident EM wave. RAM elements are mounted on the metal surface of the internal chamber walls. For that purpose, low-permittivity adhesives are used.

RAM performance depends on the pyramidal absorber thickness and the EM wave angle-of-arrival. Fig. 3 shows attenuation of the reflected EM wave against the thickness of the pyramidal absorber (in wavelengths), at normal incidence. The same dependence, for various angles of incidence, is shown in Fig. 4. If the higher pyramids are used, the chamber reflectivity performance is improved. However, RAM elements are very expensive and their cost depends on the amount of RAM material used for their fabrication. Depending on the manufacturer, pyramid–shaped RAM elements 200 mm high cost 95$ - 120 $ per square meter,

while the pyramidal absorber height of 400 mm costs 160$ - 220$ per sq meter. The RAM performance deteriorates rapidly with increasing incident angle.

IV. Rectangular Anechoic Chamber

Rectangular anechoic chamber is mostly utilized in the measurements of antenna radiation pattern, radar cross-section and EMC. Electromagnetic characteristics of this architecture are not too difficult to analyze, allowing efficient planning and construction of the chamber. When starting the process of the rectangular anechoic chamber design, the most important information is related to the minimum and maximum operating frequency, and the aperture size of the largest object under test. The text that follows analyzes a scenario of an antenna radiation pattern measurement, in order to evaluate chamber dimensions. For that purpose, a directional antenna is used as a transmitter (usually a horn antenna at high frequencies or a log-periodic antenna at low frequencies) and an antenna whose characteristics are measured, as a receiver (AUT). Usually, Ray Tracing method is utilized in the EM analysis. This method is based on the evaluating of all possible paths of EM wave propagation from the transmitting to the receiving antenna. The signal level at the receiver is determined. In the measurement process, only direct signal is valuable for AUT characterization. Signals received by AUT after one or a few reflections from the side walls, floor and ceiling of the chamber, are unwanted signals which cause interference in the measurement process. These signals are usually called the anechoic chamber noise. In Fig. 5, direct signal is drawn with the solid line (DW), while the reflected signals are presented with the interrupted lines. There are six reflected signals reaching AUT after only one reflection. In Fig. 5, four of them are shown: reflection from the wall at the back of the transmitting antenna (SRT), reflection from the wall at the back of receiving antenna (SRB), reflection from the right side wall (SSR1) and reflection from the left side wall (SSR2). Reflections from the floor and ceiling could not be presented in this two-dimensional cross section. However, they are identical to the reflections from the side walls. Fig. 5 also presents a signal which is reflected from the both side walls (MSR) and those one reflected from the side wall and then, from the wall at the back of the receiver (MRB), into the quite zone. As it can be seen, beside these two reflections, there are a large number of multiple reflections. Since levels of these signals are very low at the receiver (AUT), they are not included into the consideration. Reflection of the signal radiated by the back lobe of antenna radiation pattern (SRT), has low level at the receiver, due to the use of very directional antennas and it does not affect the evaluation of AUT, too.

If the chamber geometry is designed such that an EM signal from the transmitter reaches the receiving antenna after a few reflections, the EM energy may be sufficiently decreased after a few bounces from the absorbing walls, floor and ceiling. Since the single reflections (except SRT), entering the test region, can affect the measurement accuracy, they should be maximally attenuated by lining the chamber walls with the appropriate RAM elements.

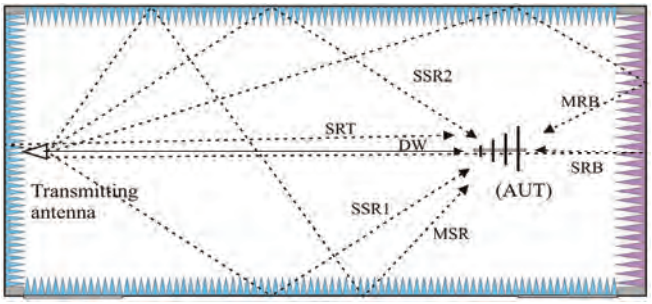


Fig. 5. Propagation of EM waves inside the rectangular anechoic chamber

The quite zone of the anechoic chamber is a defined volume within the chamber where AUT is to be placed for evaluation. Dimensions of the quite zone are equal to the aperture of the largest AUT, or greater. Suitable architecture of the chamber will result in the lowest noise level in this volume. In order to reduce reflections from the side and back walls, the quite zone diameter Q is usually determined as follows

$$Q \approx \frac{1}{3} \cdot W \tag{2}$$

In addition, distance between the quite zone and the back wall at the receiver should be equal to the quite zone width ($G \approx Q$).

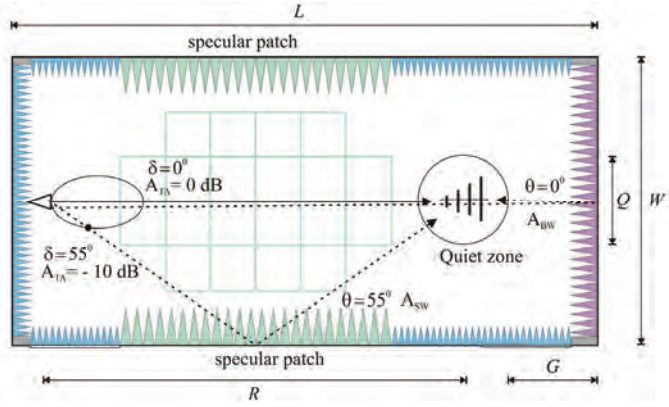


Fig. 6. Effect of EM signal specular reflection into the quite zone in the rectangular anechoic chamber

The first phase of an anechoic chamber design is related to evaluation of its size. For a given operating frequency and maximum antenna aperture, distance between transmitting and receiving antenna must meet the far-field condition according to Eq. (1). If the transmitting antenna is placed very close to the back wall, the minimal length of the anechoic chamber can be calculated as follows

$$L_{\min} \approx R + \frac{3}{2} \cdot Q = R + \frac{W}{2} \cdot \tag{3}$$

Good performance in attenuation of reflected signals can be achieved for angles of incidence less than 60º. This condition can be satisfied if the width and the height of the chamber are equal to, or greater than the half of the chamber length (W, $H \geq L/2$). Therefore, the minimal anechoic chamber length can be written by

$$L_{\min} \approx \frac{4}{3} \cdot R \tag{4}$$

The second phase of the anechoic chamber design is related to the evaluation of RAM elements to cover the chamber walls. A great deal of care must be taken to reduce the effects of the reflected signals. For every part of the chamber, type of RAM elements to be used is determined. In order to achieve satisfying performance with reasonably saving of cost, complete internal surface must be lined. The noise level in the quite zone, which is a result of reflected and diffracted signals inside the chamber, should be below the specified value. Attenuation of the reflected signals is better if the higher pyramids are utilized. Since the use of these RAM elements results in the higher cost, specified signal-to-noise ratio in the quite zone must be obtained using pyramidal absorber whose thickness is reduced as much as it is possible. Calculation of the minimal height of the pyramid elements is performed at the chamber lowest operating frequency. The RAM performance improves as the frequency goes higher due to the relative increase of the pyramidal absorber thickness (in wavelengths).

The accurate measurements in the anechoic chamber are obtained when the reflectance level of the signals entering the quite zone is 30-40 dB below the direct signal level (at the lowest operating frequency). Since single reflections have the greatest influence on the quite zone noise, they must be sufficiently reduced by RAM elements. In this case, multiple reflections would be negligible. The attenuation of the direct signal, propagating from the radiating source to the receiver, is equal to the free-space path loss $A(R)$ (Fig 6). The attenuation of the signal, reflected from the back wall at the receiver, is the sum of free-space path loss $A(R+3Q)$ and the attenuation of the back wall RAM elements A_{BW}. The attenuation of the signals reflected from the side walls, floor or ceiling, is determined as a sum of the attenuation of antenna radiation in the δ direction A_{TA} (which is different from the maximum antenna radiation), free-space path loss $A(\sqrt{R^2+W^2})$ and the attenuation of RAM elements of the side walls, floor and ceiling A_{SW}. If the differences in the free-space path losses, caused by the various propagating paths, are small, then the relative attenuation of the signal reflected from the back wall into the quite zone has value of A_{BW}. The relative attenuation of the signals reflected from the side walls into the quite zone is equal to $A_{TA}+A_{SW}$ (relative to the direct signal). The required values for A_{BW} and A_{SW} are determined in regard to the relative attenuation of the specular reflections entering the quite zone and the antenna radiation pattern. In Fig. 3, the height of the back wall RAM elements can be determined.

In order to evaluate the thickness of the pyramidal absorber needed to cover the side walls, floor and ceiling, the angles of incidence must be determined. For the side walls, these angles have value of $arctan(R/W)$, while for the floor and ceiling the incidence angles can be calculated as $arctan(R/H)$. Required height of the pyramidal absorber can be evaluated on the basis of data in Fig. 4.

V. CALCULATIONS OF THE SPECULAR ZONES OF THE RECTANGULAR ANECHOIC CHAMBER

When the minimum height of RAM pyramidal elements is determined, the complete interior surface of the chamber is usually not covered with one type of the pyramidal absorber. If these elements are used for lining the chamber surfaces that cause single reflections into the quite zone, significant cost savings can be achieved. Less expensive absorber can be used to cover the rest of the chamber surface that can only cause multiple reflections into the test region. The height of these pyramids is two times smaller than of those in the specular zone [2].

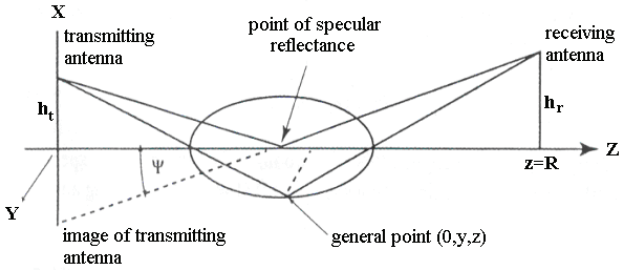


Fig. 7. The Fresnel zone boundary at the surface of the chamber wall

Techniques of geometrical optics are used to evaluate the size of specular zones. Their boundaries are determined by single reflections that are tangent to the quite zone. However, since diffraction effect is not included in the calculation, the specular zones surfaces should be larger. Due to these effects, RAM elements close to these zones are able to send some amount of EM energy into the quite zone.

Applying the concept of Fresnel zones at the area of the wall illuminated by the transmitting antenna, boundaries of the Nth Fresnel zone can be determined as follows

$$\left[h_t^2+y^2+z^2\right]^{1/2}+[h_r^2+y^2+(R-z)^2]^{1/2}-[R^2+(h_r+h_t)^2]^{1/2}=N\lambda/2 \quad (5)$$

The successive outer boundaries of the Fresnel zones on the surface of the chamber wall describe a set of ellipses whose major axis lies along the longitudinal chamber axis (Fig. 7).

Introducing new parameters

$$\varphi=\tan^{-1}[R^2+(h_r+h_t)^2]^{1/2}, \quad (6)$$

$$F_1=(N\lambda/2R+\sec(\varphi)), \quad (7)$$

$$F_2=(h_r^2-h_t^2)/[F_1^2-1)R^2], \quad (8)$$

$$F_3=(h_r^2+h_t^2)/[(F_1^2-1)R^2], \quad (9)$$

expressions for calculating the center of the Nth Fresnel zone, the zone length and width can be derived

$$C_N=R(1-F_2)/2 \quad (10)$$

$$L_N=RF_1(1+F_2^2-2F_3)^{1/2} \quad (11)$$

$$W_N=R[(F_1^2-1)(1+F_2^2-2F_3)]^{1/2}. \quad (12)$$

Experience has shown that minimum six Fresnel zones (N=6) need to be covered to achieve good reflectivity in the anechoic chamber.

VI. ARCHITECTURE OF THE ANECHOIC CHAMBER

In order to perform required measurements for the Laboratory for antennas, propagation and EM compatibility, a rectangular anechoic chamber will be designed and implemented at the Faculty of Electronic Engineering. High - accuracy measurements of a large number of antennas for modern wireless communication systems will be performed in the frequency range of 0.8 – 20 GHz.

Considering the available space inside the building of the Faculty of Electronic Engineering and the budget limitations, distance between transmitting and receiving antenna is selected to be 4.6 m. In Fig. 8, the antenna aperture size is shown as a function of the operating frequency (which is calculated based on the condition for far-field radiation). As it can be seen, at the lowest operating frequency of the chamber, antenna with the maximum aperture of 0.93 m can be evaluated, what is suitable for measurements of antennas for mobile communication base stations. According to Eqs. (3) and (4), the anechoic chamber dimensions are determined as follows: length 6.5 m (L), width 3.2 m (W) and height 3.2 m (H). Since the maximum antenna aperture decreases as the frequency goes higher, AUT can be placed closer to the back wall and still remain in the test region.

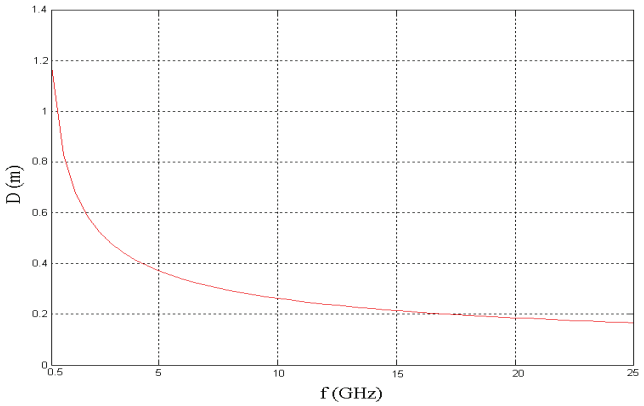


Fig. 8. Frequency dependence of the AUT aperture for R = 4.6m

The level of reflected signal entering the quite zone is the initial criteria for the selection of radio absorbing material. In the quite zone, this level should be at least 30 dB below the direct signal level. This condition is completely satisfied at the higher frequencies.

As it is stated before, noise level in the test region mainly depends on the single reflections from the side walls, floor and ceiling. Reflection from the back wall at the incidental angle of θ=0° propagates from the transmitting antenna in the direction of the antenna maximum radiation. Since δ=0°, relative attenuation of the signal is given by

$$A_{TA}(0°)+A_{BW}(0°) = 0\ \text{dB}+A_{BW}(0°) = A_{BW}(0°) \geq 30\ \text{dB} \quad (13)$$

Reflections from the side walls, floor and ceiling are carried out under the incidental angle of $\theta = \arctan(4.6/3.2) \approx 55°$. Reflected wave is previously radiated by the transmitting antenna in the direction $\delta = 90°-55°= 35°$. Relative attenuation of the wave can be written as

$$A_{TA}(35°) + A_{SW}(55°) = 10\ \text{dB} + A_{SW}(55°) \geq 30\ \text{dB} \quad (14)$$

For the chamber operating frequency range, the source antenna is usually a standard horn antenna, having attenuation of radiation around 10 dB and greater (in E and H plane), in the direction of 35°. In Eq. (14), $A_{TA}(35°)$ is assumed to be 10 dB.

According to (13), required attenuation must be $A_{BW}(0°) \geq$ 30 dB. From the graphic in Fig. 3, the minimal height of pyramidal absorber that meets this condition is equal to 0.7λ, or 265 mm, at the operating frequency of 0.8 GHz. Therefore, RAM elements, whose base dimensions are 0.5 m × 0.5 m and height of 300 mm, are selected to cover the back wall of the anechoic chamber.

Attenuation of specular zones at the side walls, floor and ceiling is $A_{SW}(55°) \geq 20$ dB. From the graphic in Fig. 4 (angle of incidence is 55°), the minimum height of the pyramidal absorber is calculated. At the operating frequency of 0.8 GHz, it is equal to 1.05λ, or approximately 393 mm. For lining the specular zones, RAM elements with base dimensions of 0.5 m × 0.5 m and height of 400 mm are selected.

TABLE I. THE ANECHOIC CHAMBER SPECIFICATIONS

Frequency range	0.8 GHz – 20 GHz
Dimensions of the chamber	6.5m (L) × 3.2m (W) × 3.2m (H)
Distance between antennas	4.6 m @ 0.8 GHz 5.0 m @ 20 GHz
Aperture of the largest AUT	0.93 m @ 0.8 GHz for R = 4.6 m 0.19 m @ 20 GHz for R = 5.0 m
Quite zone dimension	diameter 1 m @ 0.8 GHz
Quite zone reflectivity performance	- 30 dB and below @ 0.8 GHz - 33 dB and below @ 1.0 GHz - 35 dB and below @ 1.5 GHz - 38 dB and below @ 2.0 GHz - 40 dB and below @ 3.0 GHz - 44 dB and below @ 4.0 GHz

TABLE II. SPECULAR ZONE DIMENSIONS FOR N=6 AT DIFFERENT OPERATING FREQUENCIES

Frequency[GHz]	Specular zone length [m]	Specular zone width [m]
0.8	5.1	3.7
1.6	3.9	2.6
2.4	3.3	2.0
5	2.4	1.4
10	1.7	1.0
20	1.2	0.7

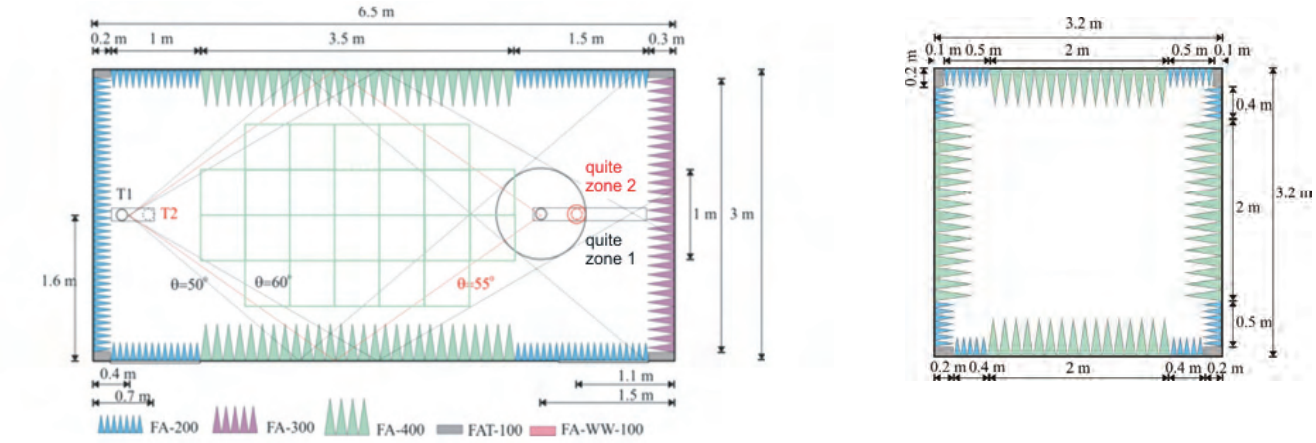


Fig. 9. Horizontal and vertical cross-sectional views of the anechoic chamber to be constructed at the Faculty of Electronic Engineering

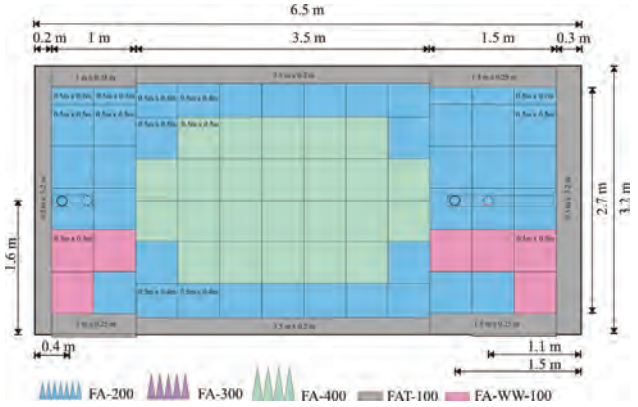


Fig.10. The anechoic chamber floor covered with RAM elements

TABLE III. REFLECTION COEFFICIENTS OF RAM ELEMENTS AT DIFFERENT FREQUENCIES

f [GHz]		0.8	1	1.5	2
Absorber	θ ()				
FA-400	0°	- 33 dB	- 36 dB	- 40 dB	- 44 dB
	55°	- 20 dB	- 23 dB	- 28 dB	- 32 dB
FA-300	0°	- 31 dB	- 33 dB	- 36 dB	- 40 dB
	55°	- 16 dB	- 19 dB	- 25 dB	- 28 dB

In Table II, calculated dimensions of the specular zone covering the first six Fresnel zones are given. FA – 400 RAM elements will be used to line the specular patch with dimensions of 3.2 m x 2 m. At the lowest operating frequency, this patch will cover the first two Fresnel zones. Covering all six Fresnel zones will be provided at the operating frequency of 2.4 GHz and higher. Figs. 9 and 10 illustrate specular patch at the side walls, floor and ceiling. The rest of the chamber internal surface can be lined with RAM elements that have two times smaller height of the pyramids (Fig. 9). Standard 100 mm high flat absorber, FAT -100, is utilized to cover the chamber corners. Fig. 10 illustrates the floor lined with special flat absorber FAT-WW-100 (100 mm thick), in the vicinity of the antenna. This is a solid absorber which allows approach to the antennas. At the selected operating frequency, the quite zone performance depends on the reflected signal maximum level (evaluated in relation to the DW level). Attenuation of SRT, SSR and SRB signals can be evaluated on the basis of reflection data from Table 3 (or taken from Figs. 3 and 4 or manufacturers specifications). It can be concluded that these results completely correspond to the quite zone performance given in Table I.

VII. CONCLUSION

Specialized electromagnetic anechoic chambers are widely used for efficient EM characterization One RF anechoic chamber, having a performance and construction as well as the other modern microwave chambers, will be realized at the Faculty of Electronic Engineering, in Niš. This implementation is very important for intensive research activities in the area of antennas, propagation and EM compatibility. In addition, the anechoic chamber will significantly improve the teaching process through the modern laboratory exercises and active participation of students in the research projects. Finally, the Laboratory for antennas, propagation and EMC will take a great part of the activities of the National laboratory for electromagnetic compatibility, which is now in the process of establishment.

REFERENCES

[1] S. Drabowitch, A. Papeiernek, H. D. Griffiths, J. Encinas and B. L. Smith, "*Modern Antennas*", Springer, 2005.

[2] L. H., Hemming, *Electromagnetics Anechoic Chambers – A Fundamental Design and Specification Guide*, John Wiley & Sons, 2002.

[3] B. K. Chung and H. T. Chuah, "Modeling of RF Absorber for Application in the Design of Anechoic Chamber", *Progress In Electromagnetics Research, PIER 43*, 273–285, 2003.

[4] Emerson, W. H., "Electromagnetic wave absorbers and anechoic chambers through the years," *IEEE Transactions on Antennas and Propagation*, 484–490, July 1973.

[5] Janaswamy, R., "Oblique scattering from lossy periodic surfaces with application to anechoic chamber absorbers," *IEEE Transactions on Antennas & Propagation*, Vol. 40, 162–169, 1992.

[6] Holloway, C. L., R. R. DeLyser, R. F. German, P. McKenna, and M. Kanda, "Comparison of electromagnetic absorber used in anechoic and semi-anechoic chambers for emission and immunity testing of digital devices," *IEEE Transactions on EMC*, Vol. 39, No. 1, 33–47, February 1997.

POSTER SESSION PO IV

PO IV - Signal Processing

Extracting Cuts From Video Streams in Real Time

Jugoslav Joković, Danilo Đorđević

Abstract – **An approach for extracting scene boundaries for streaming videos is proposed. Presented approach extends the mainstream hard cut scene detection algorithm based on the color-histogram difference by adding new criteria to increase precision and improve computational efficiency, by using additional detection criterion based on spatial distribution of luminance blocks difference during scene cut (and also by utilizing the temporal properties of video by sampling at different time resolutions). We note the speed of extraction which is crucial in streaming tasks and give results on precision and reliability.**

Keywords – **Cut detection, Video Summarization, real-time**

I. Introduction

A scene is defined as sequence of successive video frames captured with one continous operation of capture device. Isolating the original structure of the video by dividing the original sequence to separate scenes is called temporal video segmentation. It strives to detect transitions created by author and find boundary of original scenes that video consist of.

Automatic scene cut detection as a primary method in temporal video segmentation initiates starting and basic step of every video signal processing for displaying or further manipulation. Among primary application of video segmentation in creating the new content (*video editing*), automatic scene cut detection finds its application in algorithms for improving video quality (*video enhancement*), as well as in systems for detection of copyrighted videos and repeated sequnces. For the first case it is desirable to know scene boundaries for complex object tracking algorithms to stop running on the position where scene breaks. In the second case the exact times where scene change occured is used as a *fingerprint* for the video sequence, for later matching with other fingerprints in the database. Here it is also necessary to keep the precision of scene change detection to the videos that are transformed (purposefully or not) with changes in frame rate, bitrate, aspect, contrast etc.

Thanks to the advance of computer science and video technologies there are plenty of effects in video editing and, consequently, a lot of transition types. Majority of these effects can be classified into three base categories, where each transition type can be mathematicaly modelled:

1. Abrupt transitions (*cuts)*, defined as direct concatenation of two scenes without adding any kind of transitional frames, so that transition sequence is empty.

2. Blended transitions (dissolves), where the sequence I(t,x,y) of duration T results from combining two sequnces,

Authors are with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: jugoslav.jokovic@elfak.ni.ac.rs, djdane81@gmail.com

where the first sequence vanishes and second is forming:

$$I(t,x,y) = f_1(t) * I_1(t,x,y) + f_2(t)*I_2(t,x,y),\ t \in [0,T]$$

3. Nonlinear transitions (without specific pattern), where the sequence I(t,x,y) is created by combining two sequences $I_1(t,x,y)$ and $I_2(t,x,y)$ without mathematicaly specified pattern.

Temporal video segmentation is widely researced in video processing field. Resulting methods concentrate mainly on first two transition types, which are in majority of all transitions that are used in real video streams, while computationaly generated effects are mostly ignored. Alternative clasification of scene transitions [1] is based on how two scenes are spatialy and temporaly divided. Their detection is based on identification of new statistical process that the video sequence is being exposed (hard cuts), or that the video sequence was scaled with mathematicaly simple and well defined function (dissolves).

Basic idea of transition detection (scene cut) detection results from assumption that some video characteristics are different during transitions than during scene itself. This is the reason why algorithms for scene cut detection select some features and analyze their change. When significant change of monitored video features occurrs, a scene cut is declared. Color histogram, edges [2] or motion are the main types of video features used for transition detection [1,3]. Usually, smallest units of video – frames - are monitored, where each frame is characterised individualy with its feature [1,2,3,4]. Another (seldom) approach is monitoring the group of frames with some common feature, such as *motion* [5]. Scene change expresses locality principle, meaning that reliable detection is based on some local frame (frame feature) group where discontinuity occurs [6]. Generally, the steps for scene cut detection typical algorithms are frame feature selection, forming local feature group and feature change detection.

This paper presents results of use of algorithms for hard cut scene detection based on on some subsistent examples. For efficiency improvement address approach with additional criterion of spatial distribution of luminance blocks during scene change and time sampling in smaller time intervals and detection on larger time sampling intervals where we used.

One of the possible applications of scene detection is in video summarization area. Typically, video is continuously being read and segmented by using transition detection techniques and than saved back on (other) storage. Once cuts are extracted, it is possible to choose representative by selecting one frame from each cut (usually first or the middle frame) but other combinatations exists as well [8]. It is crucial for such systems to have very fast, possibly real-time transitions detection. A simplified diagram for this systems is shown in Fig 1. For performing transition detection we do not assume in our work that the video stream is MPEG encoded stream [9], although this assumption can simplify some processing steps since in these cases frames are already divided into blocks.

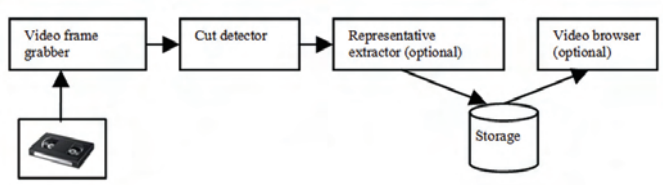


Fig. 1. Typical video summarization system diagram

II. Base Algorithm For Hard Cut Scene Detection

For all frame features used in hard cut scene detection, in literature, as well as in practice, the ones that treat spatial pixel distribution (luminance and chrominance) dominate among others. For its fingerprint compactness and amount of information in it, colour histogram highlights here. It can be crated by concatenation of component histograms to one integral fingerprint, tending to create such fingerprint representation that can be suitable for later comparison with other fingerprints with proper metrics. Metrics most used are simple (*Manhattan*) distance, Euclidean distance and histogram intersection. Regardless on the metrics being used, it is necessary to perform histogram normalization according to the frame size before comparison with other histograms. When scaling the frame to larger spatial resolution the histogram is also being scaled. Moreover, histogram in YC_bC_r domain is used much more than histogram in RGB domain, for its dominance in video format representation.

Abrupt scene changes are detected by finding the discontinuity in appropriate video frame representation and common approach is by calculating differences between frame features (color histograms) among consecutive frames and monitoring their difference - CHD (Color Histogram Difference). Scene changes (cuts) are declared when this difference surpasses some determined threshold. Typical CHD shift is depicted in Figure 2. It is obvious that scene change occurs where CHD plot peaks.

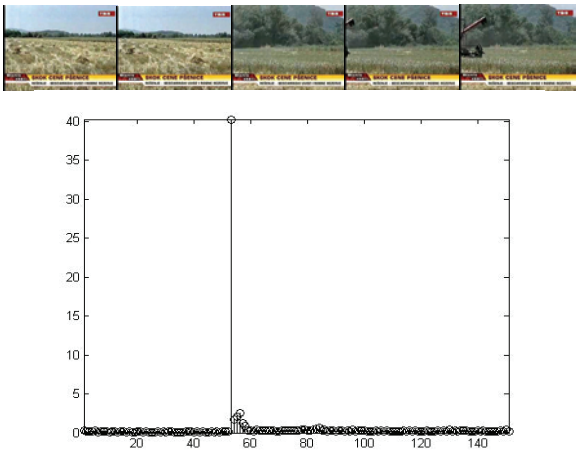


Fig. 2. Color Histogram Difference (CHD) for the video sequence »*News*«

The simplest approach for hard cut scene detection is by taking the global threshold (for example ratio between maximum peak intensities or a function between average values of CHD). Nevertheless, this approach is inefficient because sometimes small changes in CHD can seem as scene cut and vice versa – scene cut indeed occurred but change in CHD was not substantial to surpass the fixed threshold, what depends greatly on the video content being monitored. By introducing the adaptive threshold [6] that establishes itself on average CHD intensity parameter in the window of diameter w. In that case for every frame in the sequence of *2w+1* frames in width, where *w* ranges from 3 to 7 frames, and for all frames inside the window (except the central one) the average value of chosen feature (CHD) is calculated. The hard scene cut over central frame in window is declared when following conditions are fulfilled: its CHD (or $f(x)$) is maximal inside widow:

$$f(t) \geq f(x) \ \forall x \in [t-w, t+w] \qquad (1)$$

and ratio between its CHD value and average CHD values in the window (excluding central frame) crosses some empirically defined threshold ratio:

$$ratio = \frac{(f(t) + c)}{c + \sum_{x \in [t-w, t+w] \setminus \{t\}} f(x)} \qquad (2)$$

As a consequence of characteristic appearances in video relating to frequent motion, a new condition is added – CHD has to be significantly greater (ranging from 2 to three times) than the next maximum CHD in the window. Parameter *c* in expression (2) is used as a correction to avoid situations when scene changes around central frame are neglectingly small, and where a larger motion in the scene (central frame) is manifested as a peak in CHD plot leading to false detection. This method for hard cut scene detection confirmed greatest accuracy and precision [2,4,6].

Functional limits of this algorithm is that it cannot detect two scene cuts in one window, what can be a drawback in the action movies where scene changes can occur even on 3 to 5 frames, but because of great speed human audio-visual system treats these changes as a part of the one same scene.

III. Algorithm For Scene Cut Detection With Additional Spatial-Temporal Criterion

By testing the base algorithm on large video test sequences it can be noticed that previously described approach has drawbacks especially with encoded videos with large difference between I- and P- frames, what can defect the CHD between successive frames. Also, CHD can be disturbed with previously described insignificant motion on the scene which is perceptually irrelevant. Both cases exhibit very little luminance change.

Following that point, an additional criterion is introduced – *luminance difference* (YD), which is based on frame luminance block division and tracking the luminance change on *kxk* blocks for the frames where cut is suspected to occur, where YD is calculated as a sum of absolute difference between each of *kxk* corresponding luminance block average for two frames.

Algorithm for detection can be described by using the following steps:

1. Decode each video frame in video stream by using the *ffmpeg* video library [7] for frame decoding. For each frame calculate color histogram (convert the 3D color histogram to 1D representation), by performing quantization of each pixel intensity value and than form 2D matrix which will sum pixel intensities on *x* and *y* coordinates for each video component (RGB or YC_bC_r).

2. Calculate difference in color histograms between two consecutive frames (CHD) by using the sum of absolute differences of proper array elements (*Manhattan* distance).

3. For each CHD array element (for each frame feature) create a window of size *2w+1* where mean value of CHD will be calculated for elements within window, as well as ratio of this mean value with window-central CHD value.

4. If following conditions are met, declare scene cut:

- Central frame has maximum value inside window
- It is *n* times greater than the next maximum CHD value within window
- The ratio between CHD of the central frame and the mean value of CHD values within window is greater than some empirically declared threshold.
- Luminance difference (YD) on *kxk* blocks for the frames surpasses empirically declared threshold (global threshold)

This approach can help to eliminate flaws of base algorithm, to eliminate compression effects and to lower detection of motion as cuts, with lowering empirically chosen thresholds. Introducing the new criterion further improves base algorithm for cuts being missed in the cases of high motions scenes (action movies) where motion inserts high CHD values very close to the actual cuts, causing the average CHD within window to be much higher than in one at still scenes. Similar problem occurs also in videos that have smaller frame rate and high motion on the scene that cover larger portion of the frame.

This implementation of the algorithm is especially suitable for streaming performance. For both CHD and YD we only need one incoming frame from the frame grabber to extract the color histogram and average per-block luminance, and than calculate differences between those features from previous frame. After that we operate only on the array of feature vectors of the window size *2xk +1* for the cut detection phase. Thus we managed to minimize the memory operations.

Although the explained algorithm proved itself as the most reliable among other algorithms, some improvements are possible. Usage of every frame in detection can be avoided by utilizing the temporal characteristics of video. Namely, we assume that two scene changes aren't possible within some time period of t_s, where 0.5 seconds covers over 99% of possible video content for different nature. As changes are noticeable mainly in the luminance, we will use luminance difference (YD) over *kxk* blocks for two frames sampled at t_s, as described earlier in this section. For the temporal characteric we use sampling time t_s of 0.5seconds, where we are sure that double scene change won't occur. When noticeable change above fixed threshold occurs, we treat such changes either a as a cut or as a high motion. In order to reduce motion, we use higher sampling resolution and calculate YD over higher block number. When significant change is found we declare a cut.

Modified algorithm for cut detection can be described by using the following steps:

1. Decode video frames sampled at temporal resolution of t_s. Use *ffmpeg* video library for frame decoding.

2. Calculate YD between two time sampled frames, by using the sum of absolute differences over chosen block size. If this difference doesn't cross the defined «critical» threshold repeat previous steps, if it is not the case:

3. Start with frame by frame decoding from the previously sampled frame and apply the improved base algorithm, described on the beggining of this chapter.

IV. Results and Analyses

For the video sequences »*News*« luminance difference (YD) between successive frames is tested, for the whole frame as well as for blocks of size *kxk*. First, for illustrational purposes, Fig. 3 depicts results of hard cut scene detection for the video sequence »*News*«, through alongside view of YD feature with time sampling of t_s=0.5 seconds. It can be noticed how algorithm with higher time sampling enables to find periods in a video sequence (frame groups) for which we can say that they don't have cuts.

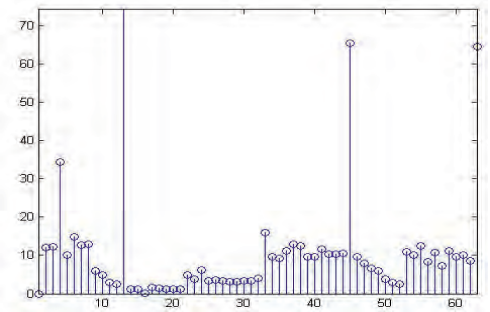
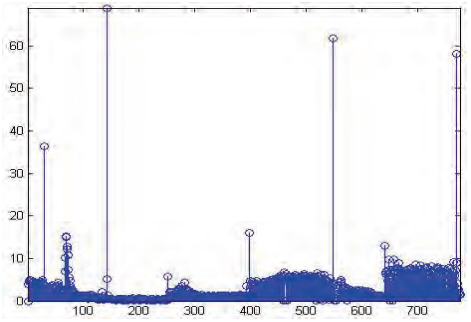
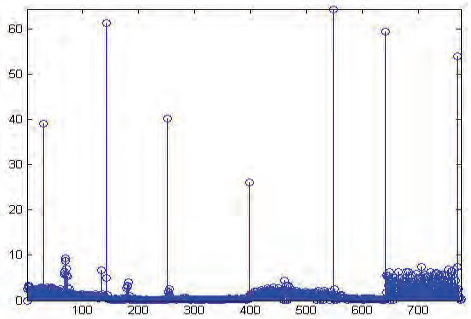


Fig. 3. Cut detection for the video sequence „News“ based on YD for the whole frame, with time sampling t_s=0.5s,

Further, Fig. 4. presents YD with frame time sampling of the video's frame rate (25*fps*), for the whole frame and for the 8x8 blocks within frame. Figure 4.a shows YD plot which is the supporting feature for reliable cut detection where this plot peaks. The result on Figure 4.b, where frame luminace difference is formed by block division, illustrates improvement of the algorithm in increasing reliability of cut detection and lowering the possibility of detecting the motion on the scene as a cut. Comparing the plots in Fig. 4, it can be seen that cuts are more noticeble in the when frame divided into blocks, while values of YD in case frames representing motion are less than YD based on whole frame. As an illustration, in Table 1 are presented results on YD versus number of blocks that frame divided into, for characteristic frames in the video sequence „News“. It can be noticed that in the case of whole frame values YD of some cut frames are less than other representing motion. However, dividing frames into blocks provides possibility of establishing of unique threshold and avoiding false cut detection.



a)



b)

Fig. 4. Cut detection for the video sequence „News“ based on YD for the whole frame, a) for the whole frame, b) for the frame blocks 8x8,

TABLE 1. YD VERSUS NUMBER OF BLOCKS FOR THE SEQUENCE „NEWS“

Frame	*Type*	*Number of blocks*			
		1	2	4	8
30	Cut	36.38	19.95	33.81	39.03
70	Motion	15.27	2.24	4.15	9.36
142	Cut	68.89	61.09	61.09	61.25
252	Cut	5.76	37.22	38.58	40.26
282	Motion	4.39	0.25	0.39	0.58
398	Cut	16.05	13.33	18.40	26.11
460	Motion	6.73	0.91	3.55	4.49
548	Cut	16.05	48.41	51.40	64.37
641	Cut	16.05	53.68	55.79	59.37
656	Motion	9.86	1.42	2.29	2.94
705	Motion	8.38	1.53	5.02	7.41
769	Cut	16.05	53.55	53.55	53.87

When complete frame is in the memory, total processing time depends on complexity of the algorithm for extracting features from each frame (CHD and YD at the same pass over frame) which is in our case O(NxM), where NxM is frame size. Cut detection in the feature space is with uncomparably smaller complexity to the feature extraction. In the terms of real time processing with frame grabbing we are limited to performance of frame grabber (in our case *ffmpeg* as software implementation) that also depends on installed hardware.

The speed of described algorithm depends mostly from the frame resolution of the video file. For the sequence of different resolution, in the Fig. 5. average execution time for clip in length of 200s at 25fps is presented. About one third of total time of algorithm execution was spent for frame decoding from *ffmpeg* library. Testing was performed on comodity PC configuration Intel P4 2 GHz with 2GB RAM.

It can be seen that for frame resolutions of double CIF we have real-time performance, while for full DVD resolution of 720x576 we are two times slower than real-time. If we can consider frame decoding as a significant time loss, our performance is near real-time.

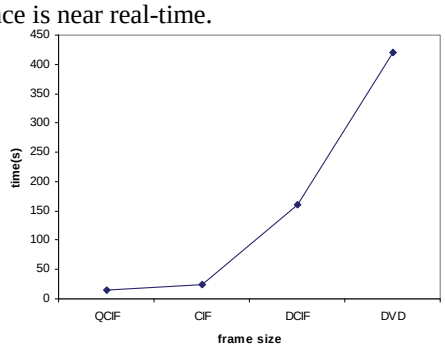


Fig. 5. Average execution time for clip in length of 200s at 25fps

V. CONCLUSIONS

Based on obtained results, the conventional algorithm for scene cut detection by using differences in color histograms of adjacent frames was strengthen with additional detection criteria oriented at spatial and temporal characteristics of a video signal, where the changes in mean luminance difference are used to distinguish cuts from motion for the given scene (to some extent). To increase the speed of algorithm, frame extraction is performed on specific fixed time intervals and afterward the mean block luminance change of adjacent frames is used as a spatial criterion. For a characteristic sequence, together with increased efficiency in processing, the results with lowered false detection rate and higher detection precision rate were attained. Also, speed of the calculation algorithm is considered in terms of real-time performance.

REFERENCES

[1] R. Lienhart, *"Reliable Transition Detection in Videos: A Survey and Practitioner's Guide"*, International Journal of Image and Graphics, vol. 1, no. 3, 2001.

[2] R. Lienhart, *"Comparison of Automatic Shot Boudary Detection Algorithms"*, Proceedings of Storage and Retrieval for Still Image and Video Databases VII, SPIE vol. 3656-29, January, 1999.

[3] I. Koprinska and S. Carrato, *"Temporal video segmentation: A survey"*, Signal Processing: Image Communication, 2001.

[4] Timothy C. Hoad, Justin Zobel: Video Similarity Detection for Digital Rights Management. ACSC 2003

[5] http://www.mathworks.com/matlabcentral/fileexchange/, Block Matching Algorithms for Motion Estimation by Aroh Barjatya

[6] Ba Tu Truong, Chitra Dorai, Svetha Venkatesh, "*New Enhancements to Cut, Fade, and Dissolve Detection Processes in Video Segmentation*", ACMM 2000.

[7] FFmpeg - http://sourceforge.net/projects/ffmpeg

[8] Borko Furht, Video Databases, book, 2003.

[9] A. Fouad, M. Bayoumi, M. Onsi, "*Shot Transition Detection with Minimal Decoding of MPEG Video Streams*", 2006.

Shape Similarity for Biometrical Analyses

Mariana Tsv. Stoeva[1]

Abstract – **This paper presents an approach for evaluation the shape similarity of image objects. The approach was developed for use in biometrical and medical image databases. This approach has the advance to achieve results that are invariant with respect of arbitrary compositions of graphical transformations and point of view change of the images processed. The approach includes an original objects image shape description in the type of multi-sized attribute stored in image data base and a technique to find and determine in the database an image that is equal in shape to the target image – query.**

Keywords – **Biometrics, Biometrical Systems, Content Based Image Data bases, Shape Similarity.**

I. INTRODUCTION

The presented approach was developed for use in biometrical and medical Image Data Bases (IDB). Both areas are developing rapidly especially at the moment.

The growth of the biometrical analysis systems applications needs great requirements, systematized in [5] to the systems for identification and verification based on biometrical analysis of fingerprint images, palm images, signatures, eye retina, cochlea, etc. In the medical field, images, and especially digital images, are produced in ever-increasing quantities and used for diagnostics and therapy. Three large domains can instantly be stated for the use of content-based access methods: teaching, research and diagnostics. Other very important fields are the automatic annotation/ codification of images and the classification of medical images.

There are a number of fields close to the medical domain where the use of content-based access methods to visual data have been proposed as well or are already implemented. The article [3] gives an overview of available literature in the field of content-based access to medical image data and on the technologies used in the field. The contemporary investigation interest is concentrated on the development of content-based IDB in the context of the object-relational model and the philosophy of "query by image example". Most of the available systems are, however, from academia. Some well-known examples include Candid [1], Photo book [4] and Netra [2] that all use simple color and texture characteristics to describe the image content. Most of these systems have a very similar architecture for browsing and archiving/ indexing images comprising tools for the extraction of visual features, for the storage and efficient retrieval of these features, for distance measurements or similarity calculation and a type of Graphical User Interface.

The approach includes original description of image objects shape in type of multi-metric attribute stored in image database and a technique for detecting and finding out in the database an image that is equal in shape with the target image – query. The query processing is supported by the definition of a distance that is conformable to the application specific and gives an account of the difference in shapes of two objects.

The experiments of the approach that investigates biometrically the shape of biometrical and medical image objects proves its invariance with respect of arbitrary transformations, as well as the ability of the defined distance to detect both big and small differences in image objects shapes.

II. SHAPE DESCRIPTION OF THE IMAGE OBJECTS

The information, kept in the image database is a description of the shape of the image objects. The extraction of the description includes: transformation of the input data and forming of the shape attribute in the form of a histogram. Figure 1 illustrates the process of getting the histogram description of the feature 'object shape' for the two types of objects respectively. The transformation of the input data is done by a composition of transformations, whose parameters are determined by geometric measures over the external contour of the image object and especially created for that purpose recursive selective function. The invariance of the description with respect to the transformations is achieved by a preliminary transformation with a composition, whose parameters are determined by the shape specificity of the image object.

The input data are biometrical (palm and cochleae) and medical digital images. The biometrical collection can include images of left or right hand taken from different points of view. The objects of interest are derived through segmentation by one of the existing methods from the assigned on pixel level image The image is transformed into a black-and-white image, containing only the identified object – F. For every object we get the external contour C_0, and the internal contours C_j using the common algorithm of going from up to down and from the left to the right.

From black-and-white images, containing an object F by using the coordinates of the external contour we get the parameters of the composition: the centroid of the points from the external contour – the point $O(x_0,y_0)$) and the maximal Euclidean distance from the centroid to the points of the external contour r_{0max} and the angular polar coordinate α of a point of the centroid that we call "starting" point [6]. Invariance of a shape description with respect to the angle of

[1]Mariana Tsv. Stoeva is with the Technical University of Varna, 9010 Varna, Studentska st. 1 Varna, Bulgaria, E-mail: mariana_stoeva@abv.bg

arbitrary rotation is achieved simply by determining the "starting" point from the external contour. We choose a starting point from all the points of the external contour by introduction of a criterion (applied to the geometric features of the figure), that we apply for three different measures P of the geometry of the points from the external contour [6]. After the transformation of the input data we get a normalized presentation of the contours of the object. The transformation is presented by the equations (1):

$$x'_{ji} = \frac{1}{r_{0\max}}(x_{ji} - x_0)\cos\alpha + \frac{1}{r_{0\max}}(y_{ji} - y_0)\sin\alpha \qquad (1)$$
$$y'_{ji} = -\frac{1}{r_{0\max}}(x_{ji} - x_0)\sin\alpha + \frac{1}{r_{0\max}}(y_{ji} - y_0)\cos\alpha$$

where: (xji, yji) - Carthesian coordinates of the i-th point into the j-th contour, (x_0,y_0) - coordinates of the centroid of the external contour, r_{0max} – the maximal Euclidean distance from the centroid to the points of the external contour, α - the angle between the radius-vector of the "starting" pixel of the contour and the positive direction of the X axis.

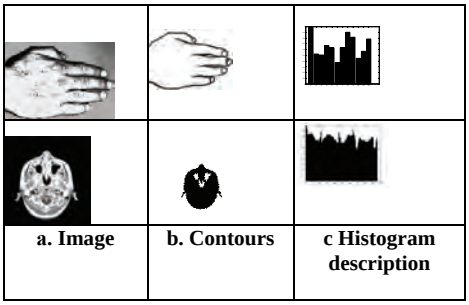


Fig.1: Illustration of obtaining an object shape description: a. image; b. contours; c. histogram description

From this way transformed contour coordinates the multidimensional index $F=((F_{\theta i}), 1\le i\le .l)$, describing the object shape in type of histograms is obtained. The value of the histogram is formed by the intersection points of the contours with the four axes, passing through the coordinate system beginning and situated on uniform steps $\Delta\theta = \pi/4$ in a circular direction. Let the line passing through the beginning of the coordinate system and subtending with the X axis angle θ_i intersects the contours of the object C_j in points with coordinates $F\cap\theta_i=((r_{j1i}, r_{j2i}), 0\le j <n)$, where: i is the consecutive number the intersecting axis $(1\le i \le l)$, l is the overall number of the intersecting axes. Shape description $F(f_1,f_2,\ldots,f_{3l})$ is given by equations (2), forming 12-dimensional attribute for shape, that is kept in the database.

$$F_{\theta i}(f_i, f_{i+1}, f_{i+2l}) = \begin{cases} f_i = \dfrac{1}{r_{0max}}\max(r_{01i}) \\ f_{i+1} = \dfrac{1}{r_{0max}}\min(r_{02i}) \\ f_{i+2l} = \dfrac{1}{r_{0max}}\sum\limits_{j=0}^{n-1}(-1)^j(r_{j1i} - r_{j2i}) \end{cases} \qquad (2)$$

III. DATABASE QUERY PROCESSING

Te technique for image retrieval as a response of a query by image example is conformed with the typical methodology for similarity query processing, including the definition of a quadratic histogram similarity distance. It measures the semantic of the histogram description and is invariant to all kinds of transformation compositions. The similarity distance shows the degree of similarity between the normalized histogram description of the query image and the image in the database. Its value is $D(Q,F) \in [0;1)$ and its value is 0 for objects with identical shape. Using the similarity distance between every image and all others a order of shape similarity can be obtained.

Let the shape query is transformed into an image histogram description $Q(q_1,q_2,\ldots,q_n)$, and the image in the database has histogram description $F(f_1,f_2,\ldots,f_n)$, where q_i, f_i are histograms. The retrieval value between Q and F for the examined retrieval model is determined by equations. (3):

$$D(Q, F) = \frac{1}{C^2 L}\min \begin{cases} \sum\limits_{i=1}^{L}(q_i - f_i)^2 \\ (q_1 - f_1)^2 + (q_{L/3+1} - f_{L/3+1})^2 + \\ (q_{2L/3+1} - f_{2L/3+1})^2 + \\ (q_i - f_{2L/3+2-i})^2 + \\ \sum\limits_{i=1}^{L/3}(q_{L/3+i} - f_{L/3+2-i})^2 + \\ (q_{2L/3+i} - f_{3L/3+2-i})^2 \end{cases} \qquad (3)$$

IV. EXPERIMENTAL RESULTS

The developed shape-research approach is experimentally investigated by use of the data of two application areas. The purpose of the first one is the identification and verification of personal data in DB containing images of unique in biometrical sense human organs as palms and cochleae. The second application uses stored numeric images for medical purposes and deformation investigations, ranges of diversions from norms, contusions, etc. Our algorithms are implemented in MatlabR12 and are evaluated by experiments on images from a test database. The images are transformed to be black-and-white and to get dimensions 64×64. All images that we use in this paper are from the experimental author's collection and they are made by the authors.

Figure 2 illustrates a test demonstration of numerous experiments whose purpose is the evaluation of the approach stability with respect of transformations both in extraction of image shape description and image retrieval from IDB. The presented experiment is with image data used for neurological investigations. A set of objects are used that are obtained from transformation variants of an original. An expert evaluation should account as equal all images in the set. The original image with target value of the parameter L=12 is used as a query and after an account of the similarity distance between

the query and each one image from the set follows an ordering of the variants by similarity. The results – the values of the similarity distance $d \in [0, 10^{-7}]$ prove the approach stability with respect of arbitrary combined compositions of transformations. The detected minimal differences in the similarity distance values in the range of 10^{-8} present the "strength" of the similarity distance, i.e. it is a low bounding sensitive object similarity distance.

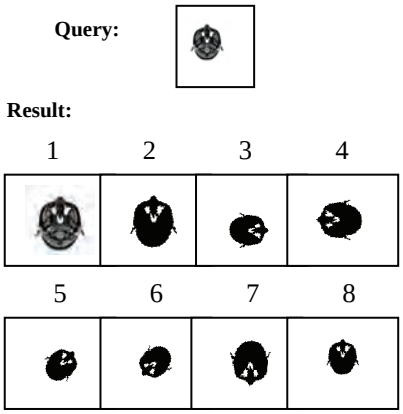


Fig.2: Similarity ordering

The next experiment presented here with Fig. 3 uses data from histological investigations and is focused on the similarity strength of the similarity distance. By this experiment we aim to ascertain the similarity distance behavior when smaller or bigger errors in the shape of the internal and external contours of the transformed images are introduced. A medical image of an object is used as a simple query model. We made some modified images of the same object transformed by translation and rotation and/or with changed internal and external contours. The replies of the k-query for k = 9 are obtained for two values of the parameter L = {12,96 }.

Result					
Similarity Distance $D \times 10^{-4}$	0	0	42	45	45
Result					
Similarity distance $D \times 10^{-4}$	145	175	388	434	

Fig. 3: Similarity ordering for parameter L=96.

The expectations for the values of the similarity distance, respectively the similarity ordering, were confirmed. Fig. 3 illustrates the data of the results obtained by this experiment as well as the ordering by shape similarity of 9 images for L=96.

Original image			
Distance - d	d=0	d=0	d=0
$d=2.\ 10^{-6}$	$d=2.\ 10^{-6}$		
Original image			
Distance - d	d=0	$d=12.\ 10^{-8}$	$d=12.\ 10^{-8}$

Fig. 4: Similarity order of different images of palms and cochleae of one person.

Distance - d	0	$2\ .\ 10^{-6}$	0.6240	0.1977	0.0510	0.1855

Fig. 5: similarity distance between 6 different images of palms

Distance - d	0	$12\ .\ 10^{-8}$	0.1670	0.0086	0.0248

Fig. 6: similarity distances between 6 different images of cochleae.

The experiments for applicability of the approach in biometrical databases includes two types of experiments. The first type of experiments has the purpose to examine the stability of the approach with respect to transformations by image retrieval. The results from one of the test groups are illustrated in Fig.4 for the two types of objects respectively. The results from the five groups show differences in the values of the similarity distance in the order of 10^{-6} and prove the stability of the approach with respect to arbitrary combined compositions of transformations and its ability to test the identity of shapes of palms and cochlea regardless of the point of view.

The second type of experiments is focused on the measuring of the similarity distance and representing the degree of shape similarity quantitatively. Fig.5 illustrates the computed similarity between 1 query image and 6 other images of palms, one of which is a transformed variant of the query. The

group includes also a second palm query image snapshot from the opposite side. Fig.6. illustrates the computed similarity between one query-image and 5 other images of cochleae, one of which is the same right cochlea image and the other one is a left cochlea image of the same person. The results confirmed our expectations for the values of similarity distance and similarity order - they show as the most similar image the image that is identical with the query, followed directly by an image that has a small change in the external contour. The represented results show the desired behavior of our similarity model with respect to the given answers and demonstrate its advantages.

V. CONCLUSION

The implemented experiments of the presented approach investigate it in details by use of various evaluations of the results and their comparison with the results from other methods. The following basic conclusions may be formulated from the obtained through the experiments results:

- The developed approach for storage and retrieval of images from IDB by similarity of the shape of the contained objects is stable with respect of arbitrary compositions of transformations;
- The approach application is efficient as it achieves completeness, correctness, and sensitiveness of the results from the similarity retrieval by shape from IDB;
- The approach achieves a very good effectiveness of information storage in IDB and a good effectiveness of query processing to IDB;
- The biometrical approach that investigates the objects shape from different types of medical images may be very useful when applied for various medical purposes. However, each concrete application has a specific that may require additional adaptation of the approach in order maximum efficient results to be obtained.

Main contributions:

- new content-based approach which provides access to the object shape of significant for various areas of biometrical and medical image objects;
- Achievement of results that are invariant with respect of arbitrary compositions of graphical transformations.

REFERENCES

[1] Kelly P. M., Cannon M. , Hush D. R., Query by image example: CANDID approach, Storage and Retrieval for Image and Video Databases III, vol. 2420 of SPIE Proc, 1995, pp. 238-248.

[2] Ma W. Y., Deng Y., Manjunath B. S., Tools for texture and texture and color-based search of images, Human Vision and Elrectronic Imaging II, vol. 3016 of SPIE Proc. San Jose, CA, 1997, pp. 496-507

[3] Muller H. Michoux N., Bandom D. Geissbuhler A. (2004) A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. *International Journal of Medical Informatics*, vol. 73, issue 1, pp.1-23

[4] Pentland a., Picard R. W., Scalaroff S., Photo-book, Toooools for content-based manipulation of image databases, International Journal of Computer Vision 18 (3) 1996 pp.233-254

[5] Stanchev P. (1999) Object-Oriented Image, *Model Technology of object-Oriented Languages and Systems TOOLS Eastern Europe'99* Blagoevgrad, June 2-4, 99-109

[6] Stoeva M., Bojikova V., Geometric Approach for Shape Description of Image Objects that is Stable with Respect of Transformations, Third International Scientific Conference "COMPUTER SCIENCE'2006",PART , pp. 95-100 , Istanbul, Turkey 12th 15nd October 2006

Digital FIR Filter with Improved Selectivity

Peter Apostolov[1]

***Abstract* – This paper describes a theory and method for designing the finite impulse response (FIR) filters. The theory is based on the approximation of ideal low pass filter transfer function. An appropriate basis function that leads to obtaining FIR filters with improved selectivity has been developed. The method has been carried out using Remez' algorithm and an optimal polynomial that approximates ideal low pass filter transfer function with high precision, has been obtained. Furthermore, an analysis of FIR filters' parameters as well as comparison with those of Parks McClellan's method has also been carried out. In addition, an example for linear phase filter design, along with calculation minimization and a result's analysis, has also been presented. With the proposed method, FIR falters with better selectivity than those of Parks McClellan's can be obtained.**

***Keywords* – FIR digital filters, Frequency response, Polynomial approximation.**

I. INTRODUCTION

The digitalization of modern electronic equipment requires the use of filters with linear phase responses. The digital FIR filters have such properties. Their synthesis is accomplished by appropriate approximation of ideal low pass filter transfer function like this.

$$H_d\left(e^{j\omega}\right)=\begin{cases}1e^{j\omega}, & 0\le\omega\le\omega_t\\ 0, & \omega_t<\omega\le\pi\end{cases} \qquad (1)$$

where $\omega_t=2\pi f_t$ is filter's normalized angular transition frequency. The most popular methods for synthesis are by using the window functions: Rectangular, Hamming, Kaiser [3], Hausdorff [1] etc. The best selective properties have Parks-McClellan's digital FIR filters [5]. They can be obtained by equiripple approximation (Fig.1) of ideal low pass filter transfer function with basis function $\cos x$. This article presents a new method for FIR filter design, which has better selectivity than Parks-McClellan's.

II. BACKGROUND

The method's theoretical base is the Alternation Theorem:

- if a function $f(x)$ is defined and continuous in closed definition area, it can be approximated by trigonometric polynomial $P_m(x)$ by power *m*, with basis function $\cos x$;
- the polynomial is the unique and best approximation, if the error function $w(x)=f(x)-P_m(x)$ has at least $m+2$ extremes in the definition area;
- all extremes are alternative and their modules are equal to positive number ε.

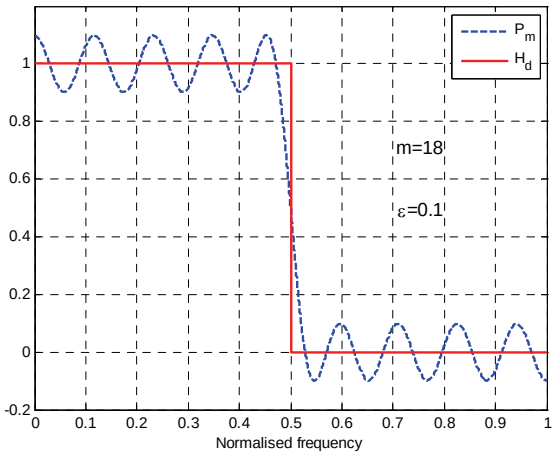


Fig.1. Parks-McClellan's approximation, *m*=18, $\varepsilon=0.1$

The method's idea is an approximation with polynomial, which posses the following properties:

- even power *m*;
- accordance with Alternation Theorem;
- polynomial's oscillations are getting dense around transition frequency f_t.

Compressing of oscillations is obtained, when the argument of basis function is "modulated" as shown

$$y(\omega)=\cos\left\{\frac{m\pi}{2}\left[\tanh\left(\beta\omega_n-\frac{\beta}{2}\right)+1\right]\right\};\ \omega_n\in[0,1], \qquad (2)$$

where β is a parameter, which changes the power of function's oscillations compressing, $\omega_n=2\omega/\omega_{sampl}$ is normalized frequency, $\omega\in\left[0,\omega_{sampl}/2\right]$ is current frequency and ω_{sampl} is sampling frequency.

On Fig. 2 the function's graph for $m=10$, for parameter's values $\beta=8$ and $\beta=15$ is shown.

The approximation's polynomial is accomplished by Remez' algorithm. It comprises iteration computing of system of $m+2$ linear equations. The obtained $m+2$ solutions are the polynomial's coefficients and the approximation error ε. The polynomial is obtained as follows

$$P_m(\omega)=\sum_{k=1}^{m+1}b_k\cos\left\{\frac{\pi(k-1)}{2}\left[\tanh\left(\beta\omega_n-\frac{\beta}{2}\right)+1\right]\right\} \qquad (3)$$

[1] Assoc. Prof. Peter S. Apostolov, PhD, Institute for Special Technical Equipment, e-mail: p_apostolov@abv.bg

On Fig.3 an approximation with the polynomial for $m = 4$, $\beta = 10$ and $\varepsilon = 0.1$ is shown

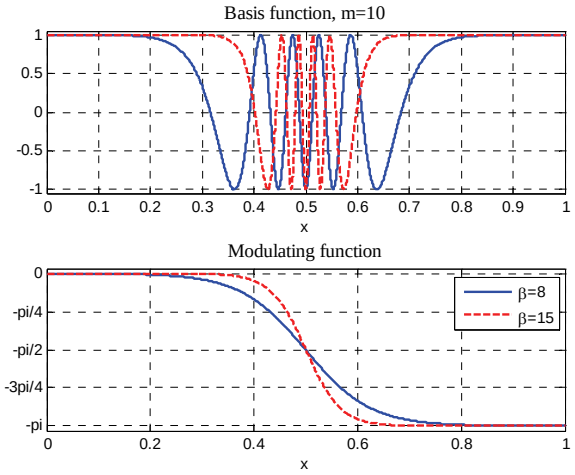


Fig.2. Basis function, with argument's modulation, $m = 10$

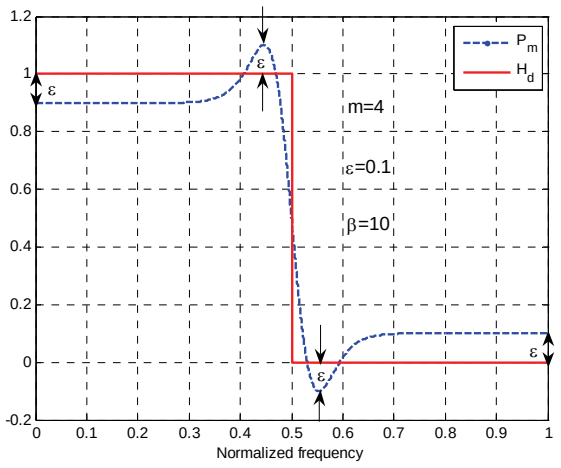


Fig.3. Approximation by offered method

On Fig.4 an approximation by offered method and Parks-McClellan's is shown. From the comparison can be seen, that with the offered method (by equal values of ε) the approximation is made by polynomial of less power.

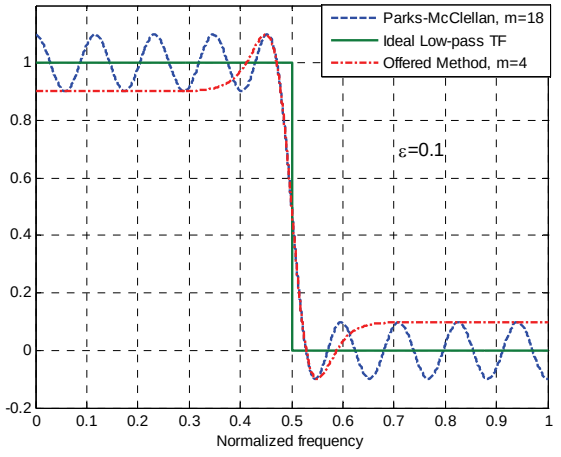


Fig.4. Approximations' comparison

The digital filter's coefficients are derived from the polynomial's coefficients in the following sequence

$$h_n = \frac{b_{m+1}}{2}, \frac{b_m}{2}, ..., \frac{b_2}{2}, b_1, \frac{b_2}{2}, ..., \frac{b_m}{2}, \frac{b_{m+1}}{2}, \quad (4)$$

where $n = 2m+1$ is the filter's length. The filter's transfer function is like this:

$$H(\omega) = \sum_{k=1}^{n} h_k \exp\left\{-j(n-k)\frac{\pi}{2}\left[\tanh\left(\beta\omega_n - \frac{\beta}{2}\right)+1\right]\right\} =$$

$$= \sum_{k=1}^{n} h_k \exp\left[-j(n-k)\varphi(\omega)\right] \quad (5)$$

III. Filter's Implementation

The implementation will be illustrated with an example for FIR filter design. Filter's specification: power of the polynomial $m = 4$ (filter length $n = 9$); transition frequency $f_t = 600$ Hz; transition band $\Delta f_t = 60$ Hz, sampling frequency $f_{sampl} = 8000$ Hz; parameter $\beta = 131$. After Remez' algorithm approximation, an optimal polynomial has been obtained with coefficients: $b_1 = 0.5$; $b_2 = 0.5655$; $b_3 = 0$; $b_4 = -0.0656$; $b_5 = 0$ and approximation error $\varepsilon = 0.0001$.

The realization is done using a method, known as "frequency sampling filter" [2], [4]. It is based on FFT in 2^N discretes, where N is an integer positive number. A value of $N = 10$ is set, i.e. 1024-FFT will be chosen. The following procedure must be carried out:

1. The minimal frequency step is calculated

$$df = \frac{f_{sampl}}{2^N} = 7.8125 \quad (6)$$

A massive, containing 512 frequencies is created

$$f_j = df : df : f_{sampl}/2; \; j = 1 \div 2^N/2. \quad (7)$$

2. A 512 values of polynomial are calculated

$$\theta_j = \sum_{k}^{m+1} b_k \cos\left\{(k-1)\frac{\pi}{2}\left[\tanh\left(\beta\left(\frac{2f_j}{f_{sampl}}\right)^X - \frac{\beta}{2}\right)+1\right]\right\}, \quad (8)$$

where

$$X = \log_2^{-1}\left(\frac{f_{sampl}}{2f_t}\right) = 0.36537. \quad (9)$$

θ_j is so called filter's "mask". It is shown on Fig.5.

3. After analog to digital conversion, the signal is applied to 2^N buffer.

4. A 1024-FFT is being performed. As a result 2^N complex frequencies F_i are obtained.

5. The first 512 frequencies are multiplied by the filter's mask θ_j, which corresponds to convolution in frequency domain. A spectrum with $2^N/2 = 512$ frequencies is obtained.

$$S_j = F_j \otimes \theta_j; \; j = 1 \div 2^N/2. \quad (10)$$

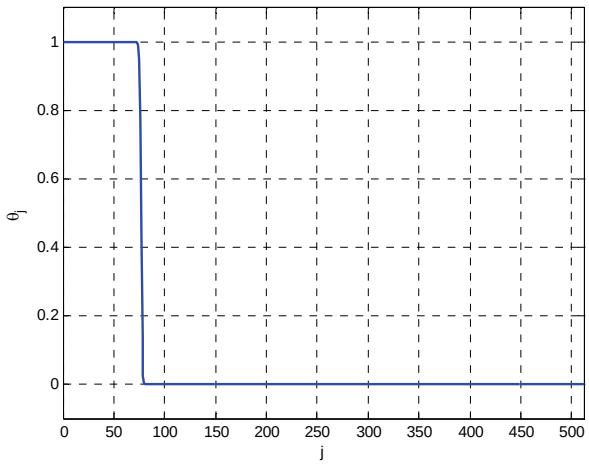


Fig.5. Filter's mask

6. The spectrum is added with $2^N/2 = 512$ zeros

$$S_i = [S_j, zeros(1,512)];\ j = 1 \div 2^N/2;\ i = 1 \div 2^N. \quad (11)$$

7. A 1024-IFFT is performed. As a result 2^N complex numbers are obtained, which real parts are the filtered signal.

8. The next 2^N discretes are taken from the input buffer and the procedure returns to p.4.

On Fig.6 the filter's structure is shown.

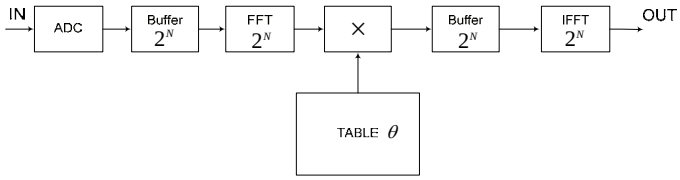


Fig.6. Filter's structure

The filter's pass-band ripple is

$$DA = 20\lg\left(1-\left|\varepsilon\right|\right) = -8.6155\text{e-}4 \text{ dB}, \quad (12)$$

and stop band attenuation

$$DS = 10\lg\left|\varepsilon\right| = -40.0356 \text{ dB}. \quad (13)$$

On Fig.7 the filter's magnitude response is shown.

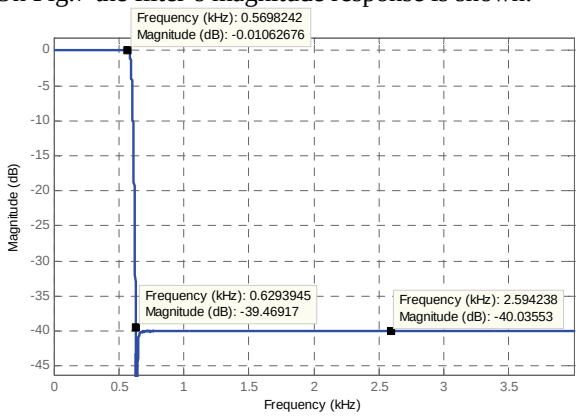


Fig.7. Magnitude response

The use of FFT leads to decreasing the filter's frequencies responses recreation accuracy. This inaccuracy comprises increasing of the pass-band ripple. The obtained value of $DA = -0.01627$ dB is reasonable from a practical point of view.

On Fig.8 the impulse response of the designed filter is shown. From the figure it can be seen, that the filter has symmetrical impulse response. This is a response of linear-phase filter.

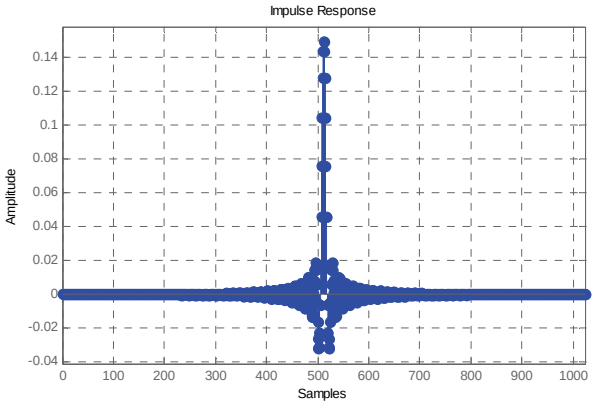


Fig.8. Impulse Response

On the next figure the phase response is shown.

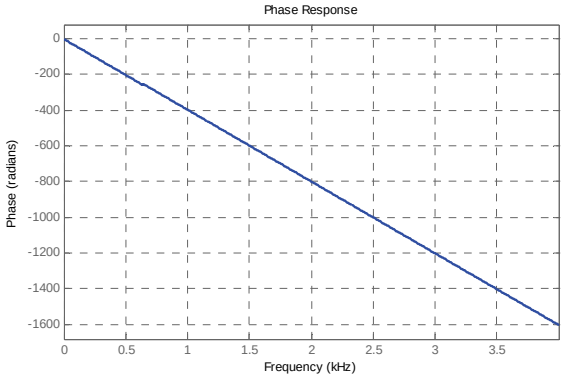


Fig.9. Phase response

The filter's group delay time (GDT) is the product of the phase response. Due to the phase response linearity the, GDT is constant – Fig.10.

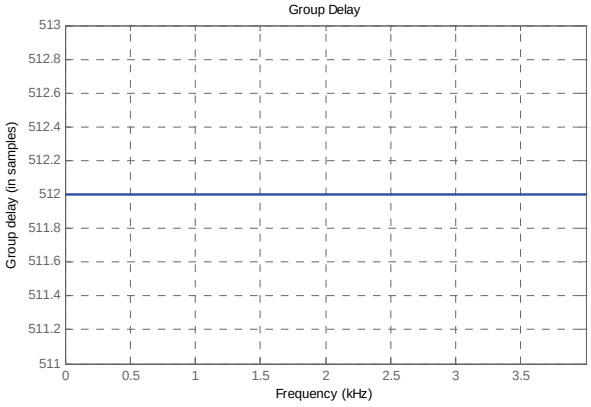


Fig.10. GDT

The GDT value corresponds to the digital processing of the input buffer numbers and can be defined by the relation

$$GDT = 2^{N-1} \text{ samples/sec}. \quad (14)$$

IV. Calculation's Minimization

From the above mentioned it can be seen, that the main calculations are the convolution (10). To get one input buffer,

the $2^N/2 = 512$ multiplications of complex with real numbers must be done, which are $2^N = 1024$ multiplications. From Fig.5 is seen, that the filter's mask contains 512 numbers, the greatest part of them, are equal either to $1-\varepsilon$ (in the pass-band), or to ε (in the stop-band). That are numbers very close to 1 or 0, because $\varepsilon = 0.0001$. The calculations can be significantly reduced if the filter's mask is divided in three zones:

1. pass-band – the numbers are set to 1;
2. transition band – the calculated mask values are accepted;
3. stop-band – the numbers are set to 0.

In our case it is appropriate for transition band that 25 numbers should be accepted, which have indexes from 66 to 90 (Fig.11)

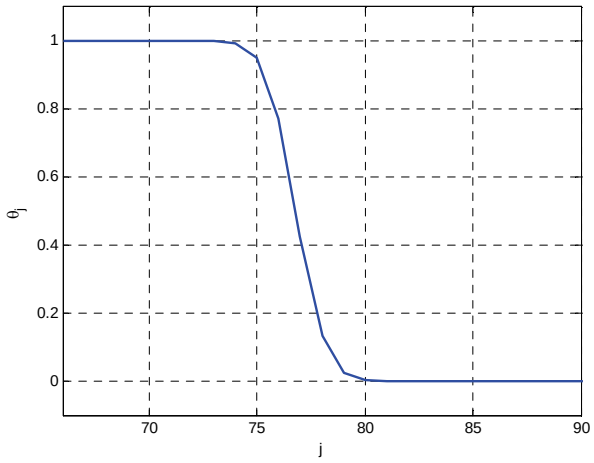


Fig.11. Filter's mask – transition band

So the filter's mask will consist of 65 ones, 25 numbers from the transition band and 422 zeroes. The first 65 complex frequencies from the input buffer go freely to the output (as if are multiplied by 1). The next 25 are multiplied with the numbers from the transition band, thereafter 422+512=934 zeroes are added without any calculation. So the calculations are reduced to 25 multiplications of complex with real number that means 50 multiplications.

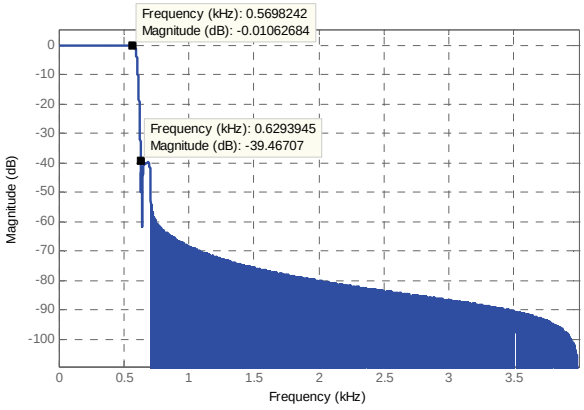


Fig.12. Magnitude response with reducing of the calculations

On Fig12 a magnitude response of a filter, realized by the offered method, is shown. It is seen, that the reducing of the calculations more than 20 times doesn't lead to Gibbs' effect and the filter remains optimal. From the comparison with Fig.7 is seen, that the magnitude response keeps its parameters in the pass-band and transition-band, the attenuation in the stop-band increases. That means the calculation's minimization can be applied successfully in practice.

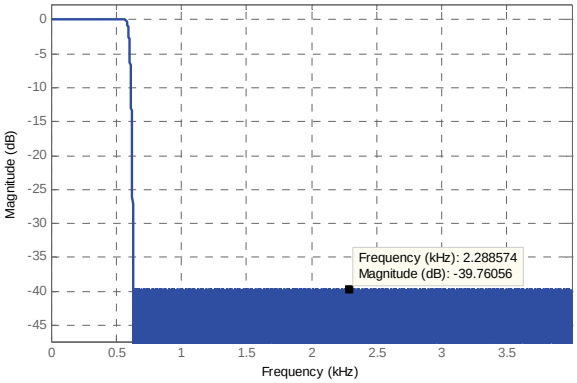


Fig.13. Parks-McClellan - Magnitude response, *n*=615

On Fig.13 a magnitude response of Parks-McClellan's filter, which has a Fig.7 fitter specification, is shown. It is seen, that the filter's length *n*=615 is over than 68 times higher.

V. Conclusion

Offered method's advantages are due to the following two circumstances:

- transfer function's coefficients are obtained from optimal approximation (Fig.3);
- every addend of transfer function (5) is multiplied with factor $\varphi(\omega) = \frac{\pi}{2}\left[\tanh\left(\beta\omega_n - \beta/2\right)+1\right]$, which is the phase response of the resonance circuit, whereas for the other FIR filters it is linear function $\varphi(\omega) = \exp(-j\omega)$.

The filters' selectivity depends on parameter β, which can grow unlimitedly. The filters don't change their length, when the ratio between transition and sampling frequency changes. They have linear phase response. They can be realized with minimal calculations.

References

[1] Apostolov, P. S. Application of Hausdorff's Window Function by FIR Filters Synthesis. *ICEST 2008*, 25-27 June, 2008, Nis, Serbia, pp. 71-74.

[2] Harris S. P. and Ifeachor E.C. (1988) Automatic design of frequency sampling filters by hybrid Genetic Algorithm Techniques. *IEEE transaction on Signal Processing*, 46(12), December, 3304-3314.

[3] Ifeachor E. C., B. W. Jervis. Digital Signal Processing. A Practical Approach. Second edition. *Pearson education limited* 2002.

[4] Lynn P. A. (1975) Frequency sampling filters with integer multipliers. In *Introduction to Digital Filtering*, Bogner R. E. and Constantenides A.G. (eds). New York: Wiley.

[5] T. W. Parks, J. H. McClellan, "A Program for the Design of Linear Phase FIR Digital Filters", *IEEE Trans. on Audio and Electroacoustics*, Vol. AU – 20, pp. 196-199, August 1972

Identification and Clusterization of Images with Neuron Networks

Vladimir A. Yanakiev[1] and Milena K. Lazarova[1]

***Abstract* – Neuron networks NN are computer models, which can "learn by their expirience". They are based on human brain on some ways. In this research is made detailed comparison between Self Organizing Map (SOM) and Feed - forward neural network (FFNN). Because of this neurons of input layer are same number with same meaning. Used input data are taken after computing from binary images of fencing weapons. Input data can be output of another application (for generation, graphical computing or even another NN). Different possibilities for outputs in Both NN are examined. FFNN without hidden layer is investigated in application.**

***Keywords* – Neural network, FFNN, SOM, backpropagation, Image identification, clusterization.**

I. INTRODUCTION

One of the first comparisons in this sphere is made by BERNARD WIDROW [3], without detailed analog. This research is based on description between different NN and accent is theoretical.

In following presentations [4-6] image identification with FFNN is study. By our knowledge in these researches investigations doesn't seems to be reported the usage of FFNN without middle layer. Reyes-Aldasor is made image identification with SOM [7]. This research is only for one class detection.

Bajwa [8] use different kind of NN for image classification, use component analysis model and compare results, but without analyze of the experiments and NN.

All of the authors use segmentation of the image and inputs of NN are the average values of these segments. In this research is used another way, in which characteristic indications like inputs are used. Characteristic indications are availability of given topology, distance between points or relations between distances. This data saved on this way is with small size.

II. EXPERIMENTAL

Short Algorithm description:

FFNN is one of the most popular NN. NN of this type have one input layer, output layer, and may have hidden layers. Neurons of every layer are connected with others by weight. Information waves are from input to output. Hidden layer is changing linear of NN. When output signals are linear, according to the input signals it is possible to avoid hidden layers, according Gochev [1] in this case they repeat inputs in some form, but it's could be applicable in many cases. The role of every neuron is to compute input signals. And output of neuron is sum of input multiplied by signals from previous layer (1).

$$N_{etj} = \Sigma W_{j,k} * Ou_{tk} \quad (1)$$

On the base of difference between target output vector and calculated output vectors are calculated errors. Errors are propagating in previous layers and change weights.

NN with hidden layer is shown on Fig.1a. NN without hidden layers is shown Fig.1b.

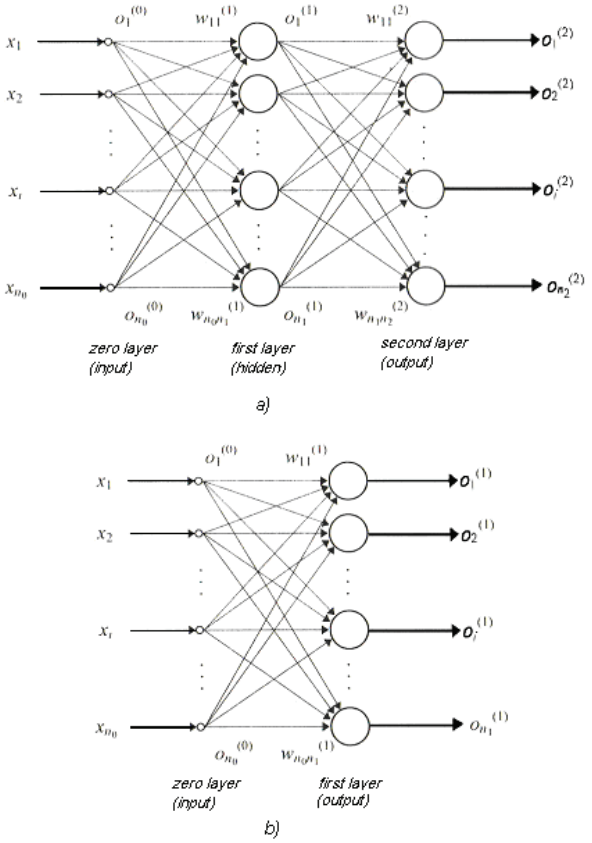


Fig. 1. FFNN with backpropagation: a) with hidden layers and b) without hidden layers.

One pass of all input vectors is calling epoch and many of epochs are enough for entire learning of NN. After this NN is

[1]Vladimir A. Yanakiev is with the Institute of Computer an Communication System, Bulgarian Academy of Science, E-mail: vladi_ian@yahoo.com

[2]Milena K. Lazarova is with the Faculty of Computer Sciences and Technology, Technical University of Sofia, Bulgaria, E-mail: milaz@tu-sofia.bg

able predictable to process data, which is not included in training vectors. This is ability to compute of unknown objects, predictable ability.

Kohonen suggest NN with algorithm without teacher, Self organizing Map (SOM) with clusterization algorithm [2]. One layer NN. Its element formed 2 Dimension matrixes. Refer to Fig. 2.

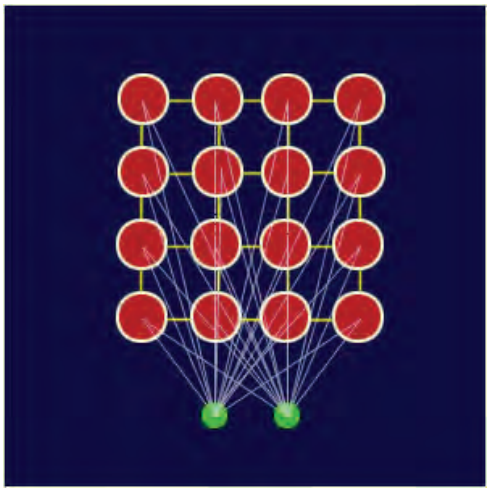


Fig. 2. Sample NN of Kohonen

The Input is usual one dimension vector with N elements. Every Input Element is connected with every output of the matrix. Input values are presented in time without target. Clusterization in NN is forming correlated input vector with weights of connections. After enough input set cluster center are formed. There are two phases: Comparing input vector and modification of weights.

Training occurs in a lot of iterations. Steps of the algorithm:

- Each node's weights are initialized.
- A vector is chosen at random from the set of training data and presented to the lattice.
- Every node is examined to calculate which one's weights are most like the input vector. The winning node is commonly known as the Best Matching Unit (BMU).
- The radius of the neighbourhood of the BMU is now calculated. This is a value that starts large, typically set to the 'radius' of the lattice, but diminishes each time-step. Any nodes found within this radius are deemed to be inside the BMU's neighbourhood.
- Each neighbouring node's (the nodes found in step 4) weights are adjusted to make them more like the input vector. The closer a node is to the BMU, the more its weights get altered.
- Repeat step 2 for N iterations

Characteristics of SOM and FFNN

Input data:

In image identification is possible to use some different kinds of inputs here are some of them. All inputs are applicable for FFNN and SOM:

- Segments of image: image is devided in segments and one every segment is found average value of colour or direction (vector) or combination. Disadvantage is necessary of similar location of image.
- Pixels of image: this is an isolated case of previous, every pixel is an input and input neurons are as many as pixels in image. This network is not applicable in big images. In NN like this even one pixel offset would make application useless. It's applicable for example for letters from same font. The advantage is the missing of graphical computing.
- Characteristic indications: In this case specific topology, distance specific points or relations between distances (for specific topology use value 1: when this one is available; 0 not available). These characteristic indications can be output from another component: with graphic computing or with another NN (recurent). In this case object can be recognized no matter of their location, rotation or scale. It is very important to correct data like input. This case is used in this research.

Output of NN:

Topology and way of working of FFNN and SOM are not allowed same output. Here are given some output case, but everyone is applicable only for one of them:

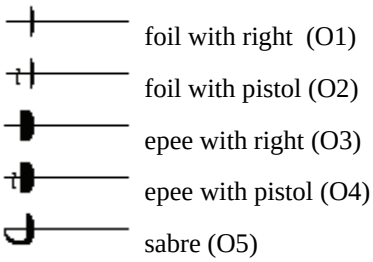
- Output Neurons number is the same as number of classes images: In this case only one output will be active (the output of the identified type) Its value will be around 1, all others should be 0. This is the most usual case in FFNN. Used in current research.
- One output neuron: In this case for every kind of image there is an interval. And depending of the interval could be decided if given image is from given class. Applicable to FFNN.
- Output neurons much more than classes images: In this case decision is taken depending on the distance to given neuron.

Topology of Hidden layers:

- FFNN: Usually used with one hidden layer, which make NN complicated and powerful. It's possible to be realized without hidden layers. This is more simple and faster. To apply this case is necessary linear dividing between input and output (every output signal is linear depending on input signals). This is applicable for used images and because of this is used in current research.
- SOM: The Topology is without hidden layer.

III. Results and Discussion

Used images:



Every epoch is from 500 images (input vector, output target vector).

The pictures which are used for this training are shown down. Characteristic indications are shown on Fig.3. The first and second are specific topology from the image (1 means that this topology is available; 0 mean not available). The third Characteristic indication is relationship between 2 distances.

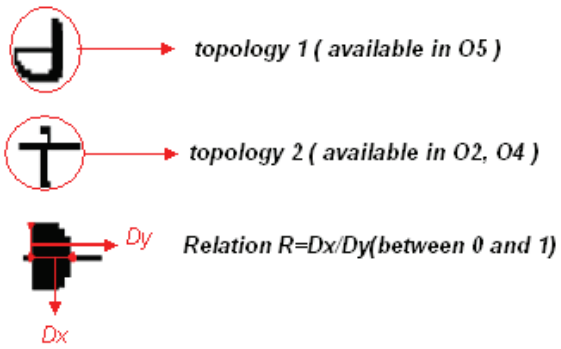


Fig. 3. Characteristic indications.

Results for FFNN:

Most researches are shown that neurons in hidden layer of FFNN should be as minimal as is possible for successful result. With a little number of neurons enough for successful training result will be much faster than with bigger number. In current research is possible without hidden layer, which reduce many calculation and remove all possibilities of local extremum. Of coarse it's not always possible to use this way, but it's possible to be checked at first without, then with hidden layers. A. Osareh [4] is found that 5 neurons are optimal for a given research tested between 1 and 30.

Following graphic on Fig. 4 is shown how the calculated outputs are moving depending from Epoch number. The output that should be active (Series1) maximum (Series2) and minimum (Series3) value of outputs that should be inactive (rest images).

After the first epoch values of output which should be active are around 0.2-0.25 the rest outputs are smaller. It is possible after 3 to 7 epochs the output which should be active to be smaller than one of the others. The reason is local extreme values. After 10 epochs the output which should be active are around 0.4-0.45the highest value. The rest are around 0.3.

Slowly output which should be active inclines to 1 all others to 0. After 50 epochs: values of outputs which should be active are around 0.9 almost 1. Rest are almost 0.

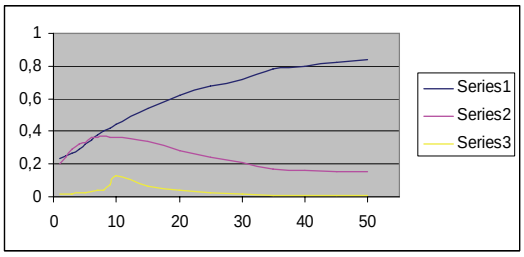


Fig. 4. Calculated output values of output that should be active (Series 1), minimum (Series 2) and maximum (Series 3) calculated outputs that should be inactive.

Results of SOM:

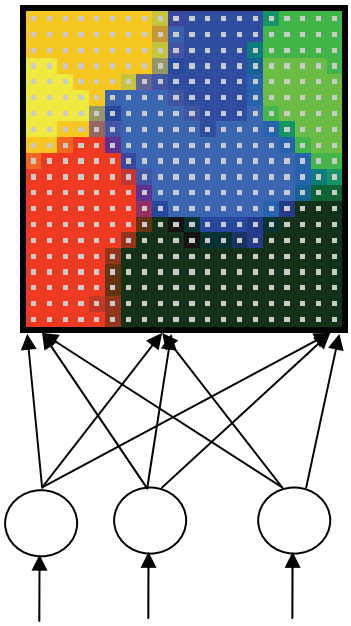


Fig.5 Scheme of using SOM and experimental result

Reyes-Aldasoro [7] uses SOM to decide if one image is from given type depending on the Euclidian distance between target and wining neuron. In given case there are 5 different types and it's not applicable. Because often the distance between winner and real target is bigger then distance between winner and fake target. This way of using will be applicable if we use five different SOMs (for all kind of weapon). This is out of the scope of this research.

After passing of one epoch all images from training set are grouped (clusterized) as can be seen in Fig 5. Let's accept that each of input parameters are passed like RGB and can be seen that all neurons are with different color. After each iteration boundaries between different groups become thinner. If iterations are not enough the images will not be clusterized

correct. One possible way to use this approach for identification is to found the boundaries and the winner will be in same boundaries with the real target, but this require new application for founding boundaries. Another using is to see how images are divided in training set to find clusters: clusterization.

IV. CONCLUSION

In current research applications with 2 kind of NN are taken into account. They are use for non graphic computing of images. Used models of NN are FFNN with backpropagation and SOM. Results show that SOM is applicable for clusterization, FFNN for identification.

REFERENCES

[1] G. Gochev “Computer sight and neural networks”, 1998

[2] http://www.ai-junkie.com/ann/som/som1.html

[3] B.WIDROW, FELLOW, IEEE, AND M. A. LEHR „30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation”, PROCEEDINGS OF THE IEEE, vol. 78, no. 9, pp. 1415-1442, 1990

[4] A. Osareh , M. Mirmehdi , B. Thomas and R. Markhamb, “Automatic Recognition of Exudative Maculopathy using Fuzzy C-means Clustering and Neural Networks”, Medical Image Understanding and Analysis, pp. 49-52, 2001

[5] M.N. Dailey, G.W. Cottrell “Organization of face and object recognition in modularneural network models”, Neural Networks, vol. 12, no.7, pp. 1053–1074, 1999

[6] S. Lawrence, C. L. Giles, A. C. Tsoi, “What Size Neural Network Gives Optimal Generalization? Convergence Properties of Backpropagation”, Technical Report UMIACS-TR-96-22 and CS-TR-3617, 1996

[7] C.C. Reyes-Aldasoro, A. L. Aldeco "Image segmentation and compression using neural networks", Advances Artificial Perception Robotics, 2000

[8] I. S. Bajwa, M. S. Naweed, M. N. Asif , S. I. Hyder “Feature Based Image Classification by using Principal Component Analysis”, ICGST - Journal of Graphics Vision, and Image processing, vol. 09, no.2, pp 11-17, 2009

Algorithm for Image Recognition on FPGA

Rosen Spirov[1]

Abstract **In this paper the author is presented the hardware application of the Haar algorithm for pattern recognition with the FPGA. The project is used the Altera DE2 board to implement a simple hardware design. For describing its behavior use the VHDL language and the Altera's Quartus tools to synthesize and to program the FPGA device**

Keywords **– Image processing, Pattern recognition, Haar wavelet, FPGA, VHDL**

I. INTRODUCTION

In this paper is presented basic hardware architecture, using wavelet Haar coefficients for face recognition.The kernel module is the parallel implementation of processing pixel data. The proposed model is implemented using VHDL and simulated and synthesized into an FPGA. To explain how different parts of the algorithm are tailored must to take advantage of FPGA architecture.

II. **U**sing Wavelet Transforms

The image recognition algorithm consist of several fairly independent stages:

- Wavelet-transform image,
- Filtering in the wavelet domain;
- Fining and Locating the area containing the image;
- Splitting the image into segments and contrasting fragments;
- Recognition of the fragment with pre-trained artificial neural network.

This document is presented basic hardware architecture, using wavelet Haar coefficients for face recognition. Kernel module is the parallel implementation of processing pixel data. The proposed model is implemented using VHDL and simulated and synthesized into an FPGA. To explain how different parts of the algorithm are tailored must to take advantage of FPGA architecture. The application of wavelet - transformation and subsequent filtration allows to recognize this key images with little contrast, such as numbers of vehicles, persons and background, distinguish underwater objects [1]. The problem with recognition of objects in background, that are two-dimensional objects, looks much more complicated than filtering one-dimensional spectra. So, first of all, it is necessary to reduce the dimension of the object, fixing the scale of the wavelet transform in the value determined by the size of the object recognition - the numbers on the plate, incidence of recurrence of black and white in this area and extent of the chosen wavelet. A good result is given symmetric wavelet, which in this case are koyflets [2, 3]. The koyflets is a special version of the fast wavelet-transformations (DWT).

Any function f in $L^2(R)$ can be expanded at some given. Level of resolution j,k in a number of species [3] is given in Eq.1:

$$f = \sum_{k} S_{jn,k}\varphi_{jn,k} + \sum_{j\geq jn,k}^{3} d_{jk}\psi_{jk} \qquad (1)$$

The wavelets $s_{j,k}$ и $d_{j,k}$ can be calculated by the Eq.2 and Eq.3:

$$s_{j,k} = \int f(x)\varphi_{j,k}(x)\cdot dx \qquad (2)$$

$$d_{j,k} = \int f(x)\psi_{j,k}(x)\cdot dx \qquad (3)$$

Once you select a particular wavelet, can be carried out wavelet transform of a signal f (x), defined as orthonormal wavelet basis. Any function $fL^2(R)$ is completely characterized by its wavelet decomposition coefficients on this basis. The function f (x) can be viewed on any n-th level resolution j_n. Then the separation between is average values at this level and fluctuations around them looks like [2, 3] in Eq.4:

$$f(x) = \sum_{k=-\infty}^{\infty} S_{jn,k}\varphi_{jn,k}(x) + \sum_{j=jn}^{\infty}\sum_{k=-\infty}^{\infty} d_{jn,k}\psi_{jn,k}(x) \qquad (4)$$

On an infinite interval, the first sum can be omitted, and the result is a "pure" wavelet decomposition. Coefficients $s_{j,k}$ and $d_{j,k}$ contains information about the composition of the signal at different scales. They can be calculated directly using formula (2), (3). However, this algorithm is not practical, since the calculation will need to spend a lot of (N^2) operations, where N denotes the number of available values of the function. The author describe a much faster algorithm. The Multistage analysis is leads naturally to a hierarchical and fast scheme for calculating the wavelet coefficients of a given function [4]. In general, the iterative formula fast wavelet transform have the form [3,5] is given in Eq.5, Eq.6, Eq.7:

$$s_{j+1,k} = \sum_{m} h_m s_{j,2k+m} \qquad (5)$$

$$d_{j+1,k} = \sum_{m} g_m s_{j,2k+m} \qquad (6)$$

$$s_{j,k} = \int f(x)\varphi(x-k)\cdot dx \qquad (7)$$

[1]Rosen P. Spirov is with the Faculty of Electronics, Technical University-Varna,
1,Studentska str . 9010 Varna , Bulgaria, E-mail: rosexel@abv.bg

These equations provide fast algorithms (the so-called pyramid algorithms) computation of wavelet coefficients, since they require now only $O(N)$ operations for its completion.The simplest solution is the direct use values of $f(k)$ of the available data set in the form of the coefficients $s_{j,k}$ and the application of fast wavelet transform using equations (6), (7). This is a safe procedure, since the pyramidal algorithm provides a complete reconstruction of the signal, and coefficients, in fact, represent the local average values, that was weighted by scaling function [6].

II. REALIZATION

The project includes the creation of parallel models respectively Matlab environment and secondly in Qartus tools. On the figure 1 is shown the Altera DE2 board.

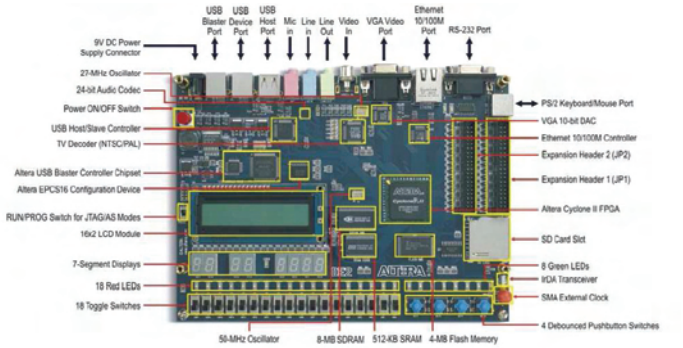


Fig 1. The Altera DE2 board

It consists of an Altera Cyclone II FPGA connected to a variety of peripherals including 512K of SRAM, 4MB of Flash, 8MB of SDRAM, VGA output Ethernet, audio input and output, and USB ports. There are three USB connectors on the top of the board. Parallel architecture of Haar Wavelet (8-inputs) are given in Fig 2:

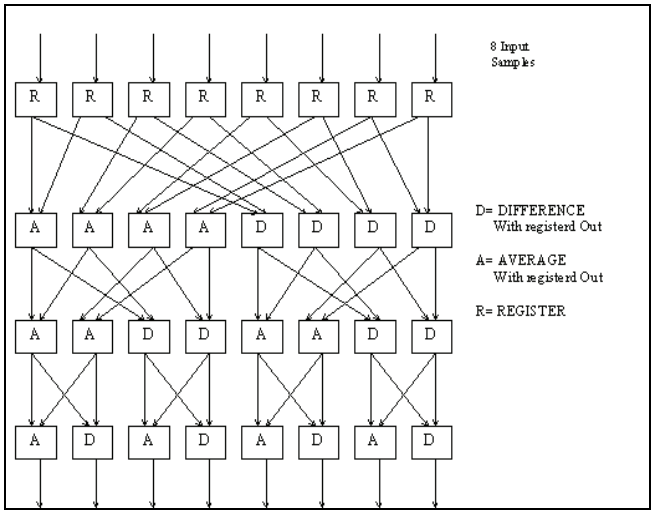


Fig 2. Parallel architecture of Haar Wavelet (8-inputs).

The Simulation Result showing data at various stages at each clock edge is shown in Fig 3.

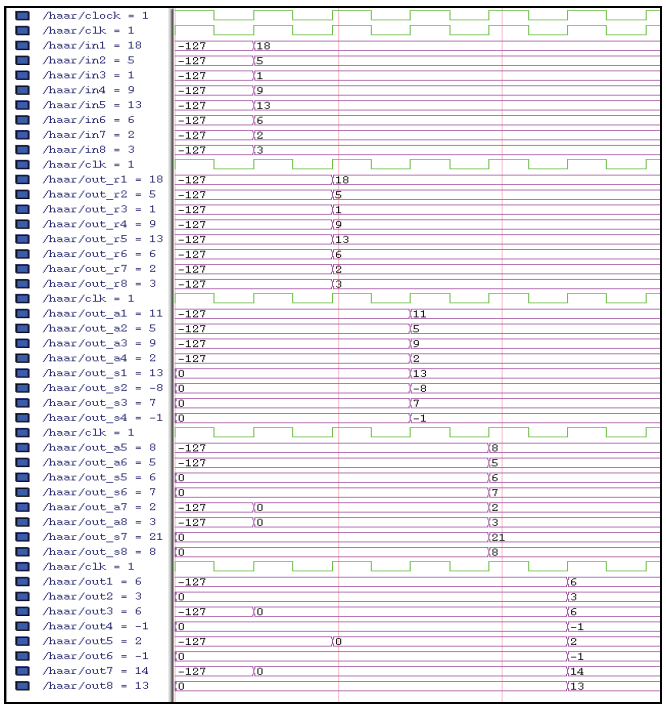


Fig. 3. Simulation Result showing data at various stages at each clock edge

The simulation result showing the parallel nature of the architecture are given in Fig 4:

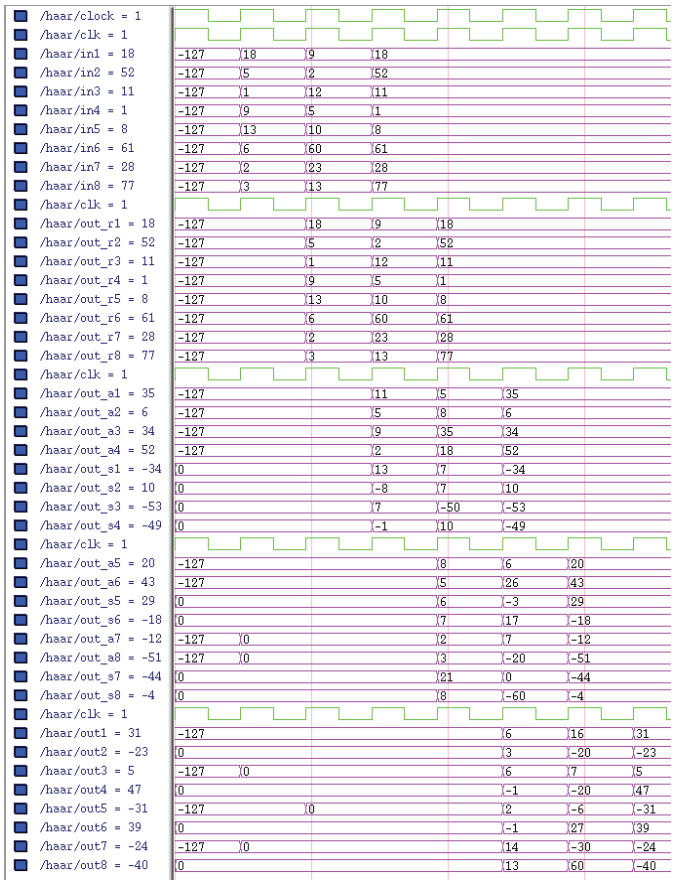


Fig 4. Simulation Result showing the parallel nature of the architecture.

This thesis was focused on design methodology for automatic learning of the optimum transformation of the input space based on Haar wavelet structure. Although the described architecture implements only the Haar wavelet, it can be used as a core subsystem in a preprocessing system which iteratively uses Haar wavelet, selects corresponding coefficients and reapply the Haar wavelet transform. Full realization of the presented concept in FPGA architecture was a subject of another thesis [5]. In abbreviated form is shown in Annex A and Annex B to realize the software Haar wavelet for image recognition and simulation in Matlab of the Quartus VHDL. Source source of Quartus is programming of FPGA on Altera kit. The resources of the FPGA used:

Number of CLBs	120/400 (30%)
Number of bonded IOBs	129/129 (100%)
Number of global buffers	1/12 (8%)
Total equivalent gate count	3948
Minimum period	25.405 ns
Maximum frequency	39.362 MHz
Maximum net delay	10.365 ns
Avrg. Con. Delay	3.494 ns
Avrg. Con. Delay on crt.nets	0.000 ns
Average Clock Skew	0.248 ns
The Max. Pin Delay	10.365 ns
Avrg. Con.Delay on the 10 Worst Nets	9.184 ns

III. SOFTWARE IN AN ABBREVIATED FORM

Annex A. Program in MATLAB: haar.m

```
%plotting the interpretation of haar transform
numCoeff=8;
% determine levels????
levls=round(log10(numCoeff)/log10(2));
x=1:numCoeff;
rows=numCoeff;
coeffMatrix=diag(ones(1,numCoeff));
coeffD=coeffMatrix;
for i=1:levls
coeffC=[];
begn=1;
while begn<numCoeff
oddr=(begn:2:begn+rows-1);
evnr=(begn+1:2:begn+rows-1);
coeffA=[];
coeffB=[];
for j=1:length(oddr)
coeffA(j,:)=( coeffMatrix(oddr(j),:) + coeffMatrix(evnr(j),:)
)/2;
coeffB(j,:)=coeffMatrix(oddr(j),:) - coeffMatrix(evnr(j),:);
end;
coeffC=[coeffC;coeffA;coeffB];
……
```

Annex B. Program in VHDL: Haar.vhd

```
LIBRARY ieee;
USE ieee.ALL;
ENTITY haar IS
PORT (
clock : IN bit;
-- Inputs
in1 : IN integer RANGE -127 TO 127;
……
in8 : IN integer RANGE -127 TO 127;
--Outputs
out1 : OUT integer RANGE -127 TO 127;
……..
out8 : OUT integer RANGE -127 TO 127
);
END haar;
ARCHITECTURE haar OF haar IS
COMPONENT bufgs
PORT (
i : IN bit;
o : OUT bit
);
END COMPONENT;
COMPONENT regPORT (
input : IN integer RANGE -127 TO 127;
clk : IN bit;
output : OUT integer RANGE -127 TO 127
);
END COMPONENT;
COMPONENT adddiv
PORT (
a : IN integer RANGE -127 TO 127;
b : IN integer RANGE -127 TO 127;
clk : IN bit;
c : OUT integer RANGE -127 TO 127
);
END COMPONENT;
COMPONENT difference
PORT (
a : IN integer RANGE -127 TO 127;
b : IN integer RANGE -127 TO 127;
clk : IN bit;
c : OUT integer RANGE -127 TO 127
);
END COMPONENT;
SIGNAL out_r1, out_r2, out_r3, out_r4, out_r5, out_r6,
out_r7,
out_r8 : integer RANGE -127 TO 127;
SIGNAL out_a1, out_a2, out_a3, out_a4, out_a5, out_a6,
out_a7,
out_a8 : integer RANGE -127 TO 127;
SIGNAL out_s1, out_s2, out_s3, out_s4, out_s5, out_s6,
out_s7,
out_s8 : integer RANGE -127 TO 127;
SIGNAL clk : bit;
BEGIN
xbufgs: bufgs
PORT MAP (
clock,
clk
);
-- Input to Register here
r1: reg
```

```
PORT MAP ( in1, clk, out_r1 );
……..
PORT MAP ( out_a7, out_a8, clk, out6 );
a12: adddiv
PORT MAP ( out_s7, out_s8, clk, out7 );
s12: difference
PORT MAP ( out_s7, out_s8, clk, out8 );
END haar;
```

The performance of this implementation can be attributed to the parallel hardware blocks used in performing the necessary calculations for the algorithm [7]. Further to this, the design can be scaled for larger databases by simply adding more processing elements in parallel The above hardware design was implemented on an Altera Quartus II board (clocked at 100 MHz) and was able to perform face recognition on a database of 10 faces in 3.88 milliseconds. A total of 7,820 logic elements were used, 2,348 of which were flip-flops. Again, performance can be attributed to the highly parallel nature of the hardware design and the composite algorithm used FPGA. Original image after wavelet-transform at Haar basis shown on Fig 5. and Fig 6.

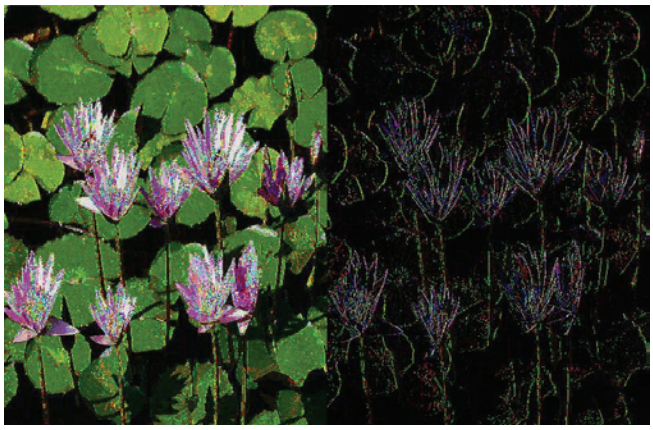


Fig 5. Original image after wavelet-transform by rows at Haar basis.

Fig 6. Original image after wavelet-transform in lines Haar basis

The experiment result shows that the whole operation time is about 60 clock cycle, which about 0.6us at 100MHz clock pulse, so the operation speed can be up to 1.5MHz. The whole design requires 2592 ALUTs and 241 registers (occupancy of resources is about 21%). The advantage of parallel processing in FPGA leads to a substantial increase in performance and accuracy in processing, extraction of the contour information than in the simulation in Matlab.

IV. Conclusion

So it provides a practical method for using FPGA to realize face recognition. The main contribution of our work is design and implementation of a physically feasible hardware system to accelerate the processing speed of the operations required for real time face recognition. The FPGA implementation and simulation results are given in this paper.. The proposed models are implemented using VHDL, and simulated and synthesized into a single FPGA. Therefore, optimization of the hardware source usage was the primary aim of this study. The research results show that the system works with high speed while consuming small amount of logic resources in comparison with original subspace feature extraction process methods. It demonstrates that this technology can produce effective and powerful applications for face recognition systems.

References

[1] P. Belhumeur, J. Hespanha, D. Kriegman. Eigenfaces vs linear projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1997, *19*(7): 711-720.

[2] H.K. Ekenel, B. Sankur. Multiresolution face recognition.*Image and Vision Computing*, 2005, *23*(5): 469-477. fisherfaces: Recognition using class specific

[3] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 1991, *3*(1).

[4] Ahn J.H., Choi S., Oh J.H. A new way of PCA: Integrated-squared-error and EM algorithms. *In: Proc. IEEE Int'l Conf. Acoustics*, *Speech* and *Signal Processing*, Montreal, Canada, 2004.M. S. Sadri and et al. An FPGA based fast face detector *Proc. of GSPX Conference*, 2004: 586-591

[5] .Pellerin D. and S. Thibault. Practical FPGA programming in C. *Prentice Hall PTR*, ISBN: 0-13-154318-0.

[6] Wang Y., Osterman J. and Zhang Y. Video Processing and Communications. C.*Prentice Hall PTR*, ISBN: 0-13-017547-1.

[7] Parhi K. VLSI Digital Signal Processing System Design and Implementation Wiley Inter Since 0471241865.

Revisiting the Fractional-Order Hold Device

Milica B. Naumović

***Abstract* – This paper is concerned with the proper setting the fractional-order hold device for signal reconstruction in digital control systems. To demonstrate the features of the considered hold algorithm, using MATLAB®Simulink software both the time and frequency domain characteristics of the hold device are analyzed.**

***Keywords* – Fractional-order hold device, signal reconstruction process, frequency response analysis, MATLAB®Simulink environment.**

I. INTRODUCTION

In computer-controlled systems, it is necessary to convert the control actions, calculated by the computer as a sequence of numbers, into a continuous-time signal that can be applied to the process. Theoretically, this kind of the reconstruction process can be done by employing an ideal low-pass filter. However, physically realizable data holds approximate in some sense an ideal low-pass filter. Higher-order hold circuits will generally reconstruct a signal more accurately than, in digital control systems commonly used, zero-order hold, but with some disadvantages related primarily to the possibility of its realization.

The paper is organized as follows. A brief review of the reconstruction process based on the fractional-order hold device is given in Section 2. Section 3 addresses to the quality analysis of the reconstruction process performed by using the frequency responses of the fractional-order hold device. Finally, in Section 4 some concluding remarks are given.

II. THE FRACTIONAL-ORDER HOLD

The sampling operation produces an amplitude-modulated pulse signal. The purpose of the hold operation is to reconstruct the original analog input signal. The hold circuit is designed to extrapolate the output signal between successive points according to some prescribed manner. So, the hold circuit that produces a staircase waveform is called a zero-order hold (ZOH). Recall that, because of its simplicity, the zero-order hold is commonly used in digital control systems. Higher-order hold circuits will generally reconstruct a signal more accurately than zero-order hold. For example, in the case of the first-order hold (FOH), the extrapolated function within a given interval is a straight line and its scope is determined by the values of the function at the sampling instants in the previous interval. The error generated in this way can be reduced by using only a fraction of the scope in the previous interval, as shown in Fig. 1b. This is achieved by using the fractional-order hold (FROH) device, given in Fig. 1a under the assumption that the value of α ranges from zero to unity [1].

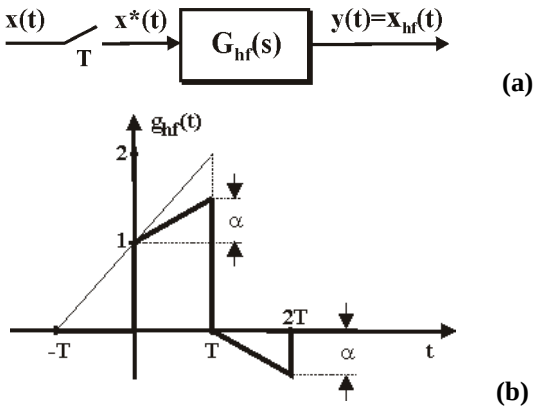


Fig. 1. Fractional-order hold device: (a) Block-diagram; (b) Impulse response.

The transfer function of the fractional-order hold can be shown to be [2]

$$G_{\mathrm{hf}}(s)=\left(1-\alpha\,\mathrm{e}^{-Ts}\right)G_{\mathrm{h0}}(s)+\frac{\alpha}{T}G_{\mathrm{h0}}^{2}(s)\,. \qquad (1)$$

After some simplifications, the above transfer function can be rewritten as follows

$$G_{\mathrm{hf}}(s)=\alpha G_{\mathrm{h1}}(s)+\left(1-\alpha\right)G_{\mathrm{h0}}(s)\ , \qquad (2)$$

where $G_{\mathrm{h0}}(s)$ and $G_{\mathrm{h1}}(s)$ are transfer functions of zero-order hold and first-order hold, respectively.

A. Discretizing the Models of the Analog Plants by using the Fractional-Order Hold

For simplicity, without loss of generality, consider the n th-order single-input single-output control object given by

$$\begin{aligned}\dot{\mathbf{x}}(t)&=\mathbf{A}_{\mathrm{c}}\mathbf{x}(t)+\mathbf{b}_{\mathrm{c}}u(t)\\ c(t)&=\mathbf{d}\mathbf{x}(t)\ ,\end{aligned} \qquad (3)$$

where $\mathbf{A}_{\mathrm{c}}\in\Re^{n\times n}$, $\mathbf{b}_{\mathrm{c}}\in\Re^{n\times 1}$, and $\mathbf{d}\in\Re^{1\times n}$.

It is convenient to introduce the realization sets as follows [3]

$$\mathrm{S}_{\mathrm{c}}\ \overset{\mathrm{def}}{=}\ \left\{\left(\mathbf{A}_{\mathrm{c}},\mathbf{b}_{\mathrm{c}},\mathbf{d}\right):G_{\mathrm{c}}(s)=\frac{N_{\mathrm{c}}(s)}{D_{\mathrm{c}}(s)}=\mathbf{d}\left(s\mathbf{I}-\mathbf{A}_{\mathrm{c}}\right)^{-1}\mathbf{b}_{\mathrm{c}}\right\}. \qquad (4)$$

Suppose a fractional-order hold is connected to the input of the object (4). Then $u(t)$ is given by

Milica B. Naumović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milica.naumovic@elfak.ni.ac.rs

$$u(t) = u(kT) + \alpha \frac{u(kT) - u\left[(k-1)T\right]}{T}(t - kT) \qquad (5)$$
$$kT \le t < (k+1)T, \quad k = 1, 2, \mathrm{K}$$

where T denotes sampling period, and α is the adjustable gain of the hold device. The resulting sampled-data system can be described by the equations

$$\begin{aligned} \mathbf{x}\left[(k+1)T\right] &= \mathbf{A}_{\mathrm{q}}\mathbf{x}(kT) + \mathbf{b}^{+}_{\mathrm{q}\alpha}u(kT) + \mathbf{b}^{-}_{\mathrm{q}\alpha}u\left[(k-1)T\right] \\ c(kT) &= \mathbf{d}\mathbf{x}(kT) \ , \end{aligned} \qquad (6)$$

where

$$\begin{aligned} \mathbf{A}_{\mathrm{q}} &= \mathrm{e}^{\mathbf{A}_{\mathrm{c}}T} \\ \mathbf{b}^{+}_{\mathrm{q}\alpha} &= \int_0^T (1 + \alpha - \alpha\frac{t}{T})\mathrm{e}^{\mathbf{A}_{\mathrm{c}}\tau}\mathbf{b}_{\mathrm{c}}\,\mathrm{d}\,\tau \\ \mathbf{b}^{-}_{\mathrm{q}\alpha} &= \int_0^T \alpha(\frac{t}{T} - 1)\mathrm{e}^{\mathbf{A}_{\mathrm{c}}t}\mathbf{b}_{\mathrm{c}}\,\mathrm{d}\,t \ . \end{aligned} \qquad (7)$$

We denote the digital model (6) by $\mathrm{S}_{\mathrm{q}\alpha}$, which can be rewritten in the form of the ordinary discrete-time state equation as

$$\begin{aligned} \begin{bmatrix} \mathbf{x}\left[(k+1)T\right] \\ u(kT) \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_{\mathrm{q}} & \mathbf{b}^{-}_{\mathrm{q}\alpha} \\ \mathbf{0} & 0 \end{bmatrix}\begin{bmatrix} \mathbf{x}(kT) \\ u\left[(k-1)T\right] \end{bmatrix} + \begin{bmatrix} \mathbf{b}^{+}_{\mathrm{q}\alpha} \\ 1 \end{bmatrix}u(kT) \\ c(kT) &= \begin{bmatrix} \mathbf{d} & 0 \end{bmatrix}\begin{bmatrix} \mathbf{x}(kT) \\ u\left[(k-1)T\right] \end{bmatrix} \ . \end{aligned} \qquad (8)$$

Thus,

$$\mathrm{S}_{\mathrm{q}\alpha} \overset{\mathrm{def}}{=} \left\{ \left(\mathbf{A}_{\mathrm{q}\alpha}, \mathbf{b}_{\mathrm{q}\alpha}, \mathbf{d}\right) : G_{\mathrm{q}\alpha}(z) = \frac{N_{\mathrm{q}\alpha}(z)}{D_{\mathrm{q}\alpha}(z)} = \mathbf{d}\left(z\mathbf{I} - \mathbf{A}_{\mathrm{q}\alpha}\right)^{-1}\mathbf{b}_{\mathrm{q}\alpha} \right\} \qquad (9)$$

where

$$G_{\mathrm{q}\alpha}(z) = \mathrm{Z}\,\mathrm{L}^{-1}\left\{G_{\mathrm{hf}}(s)\cdot G_{\mathrm{c}}(s)\right\} \ , \qquad (10)$$

represents FROH equivalent model for the continuous-time system (4) at sampling interval T. However, it is interesting to note that the model can be converted from continuous-time to discrete-time using a matrix exponential as follows [4], [5]:

$$\left[\begin{array}{c|c|c} \mathbf{A}_{\mathrm{q}} & \mathbf{b}_{\mathrm{q}0} & -\mathbf{b}_{\mathrm{q}1} \\ \hline \mathbf{0} & 1 & 1 \\ \hline \mathbf{0} & 0 & 1 \end{array}\right] = \exp\left[\begin{array}{c|c|c|c} \mathbf{A}_{\mathrm{c}}^{\mathrm{T}} & \mathbf{b}_{\mathrm{c}}^{\mathrm{T}} & 0 \\ \hline \mathbf{0} & 0 & 1 \\ \hline \mathbf{0} & 0 & 1 \end{array}\right] , \qquad (11)$$

where

$$\begin{aligned} \mathbf{b}^{-}_{\mathrm{q}\alpha} &= \alpha\,\mathbf{b}_{\mathrm{q}1} \\ \mathbf{b}^{+}_{\mathrm{q}\alpha} &= \mathbf{b}_{\mathrm{q}0} - \alpha\,\mathbf{b}_{\mathrm{q}1} \ . \end{aligned} \qquad (12)$$

The same discrete transfer function $G_{\mathrm{q}\alpha}(z)$, as that given by (10), can be also derived from (6)-(8) and represented in form of

$$\begin{aligned} G_{\mathrm{q}\alpha}(z) &= \frac{N_{\mathrm{q}\alpha}(z)}{D_{\mathrm{q}\alpha}(z)} \\ &= \frac{(z-1)\det\begin{bmatrix} z\mathbf{I} - \mathbf{A}_{\mathrm{q}} & -\mathbf{b}^{+}_{\mathrm{q}\alpha} \\ \mathbf{d} & 0 \end{bmatrix} + z\det\begin{bmatrix} z\mathbf{I} - \mathbf{A}_{\mathrm{q}} & -\mathbf{b}_{\mathrm{q}0} \\ \mathbf{d} & 0 \end{bmatrix}}{z\det\left[z\mathbf{I} - \mathbf{A}_{\mathrm{q}}\right]} \end{aligned} \qquad (13)$$

It is well-known that the zeros of a discrete-time system can be classified into two categories: those which correspond to the zeros of continuous-time system, and so called discretization zeros introduced by the chosen discretization method. As it can be found in the literature, the significant improvements in the transient performance of the closed-loop system can be obtained by using a properly adjusted FROH, instead ZOH or FOH devices. This could be achieved by using the negative values of α, because it results in a more stable discretization zeros [6].

B. One Realization of FROH Device

In general, only zero-order hold devices are implemented in hardware [7]. Higher-order hold algorithms are implemented in computer software since it is most difficult physically to implement higher-order hold devices, as fractional-order hold for example. Also, the realizations of some approximated versions of FROH can be found in the literature. It is suitable to emphasize that the lack a number of real experiments with the special hold devices is mainly due to the difficulties encountered in its implementation with the processing speed and configurability that real-time digital controllers require [8].

As a block diagram, an analog FROH device is presented in Fig. 2. This is a straight implementation of Eq. (2). The proposed realization is based on two sample and zero-order hold devices and is suitable for both the computer simulation and real experimentation.

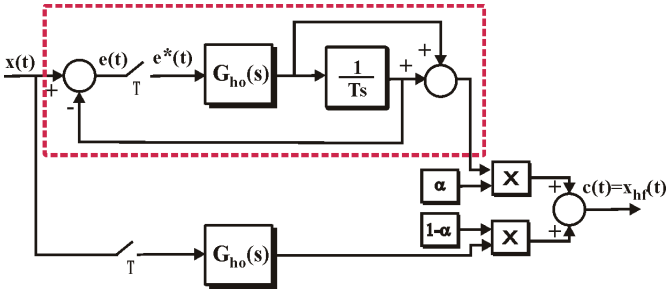


Fig. 2. Block-diagram of the fractional-order hold device

C. Rational approximation of $G_{\mathrm{hf}}(s)$

Since the transfer function (1) is not a rational function, it is suitable to find its finite-dimensional rational approximation. Among the many methods, PADÉ approximations are widely used to approximate by a rational function the dead-time given in continuous-time control systems with transfer

function $e^{-s\tau}$. Instead of the standard approximation, the PADÉ approximation, where the numerator degree is one less than the denominator degree, is recommended. Recall that the PADÉ approximation of order (m,n) is defined to be a rational function $R_{m,n}(s)$ expressed in a fractional form [9]:

$$e^{-\tau s} \approx R_{m,n}(\tau s) = \frac{N_m(\tau s)}{D_n(\tau s)} \qquad (14)$$

where

$$N_m(\tau s) = \sum_{k=0}^{m} (-1)^k \frac{(m+n-k)!m!}{(m+n)!k!(m-k)!} \tau^k s^k \qquad (15)$$

$$D_n(\tau s) = \sum_{k=0}^{n} \frac{(m+n-k)!n!}{(m+n)!k!(n-k)!} \tau^k s^k . \qquad (16)$$

Note that the high-order PADÉ approximations produce transfer functions with clustered poles. Because such pole configurations tend to be very sensitive to perturbations, PADÉ approximations with order $n > 10$ should be avoided.

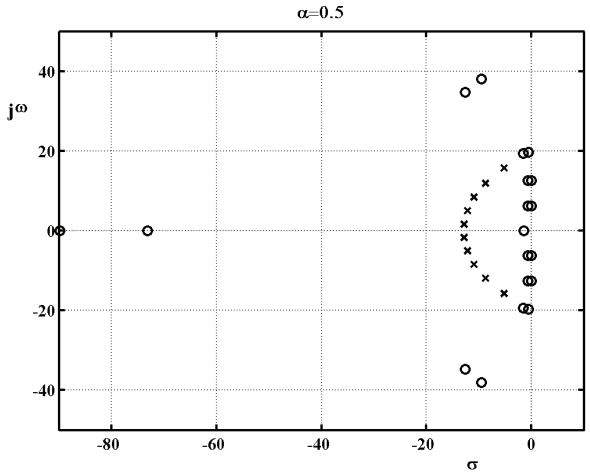


Fig. 3. Pole-zero configuration of the $G_{f\,9,10}(s)$ approximation of transfer function $G_{hf}(s)$ for $\alpha = 0.5$

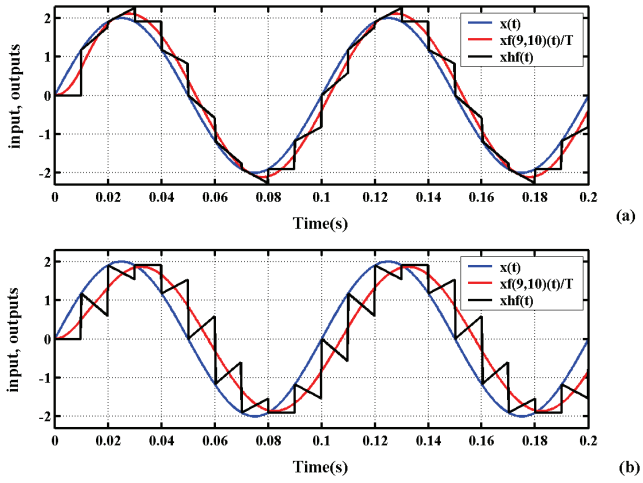


Fig. 4. Responses of FROH to sine wave input for different values of parameter α : (a) $\alpha = 0.5$; (b) $\alpha = -0.5$.

In this paper, the approximation $R_{9,10}(Ts)$ for the transfer function e^{-Ts} is adopted. Fig. 3 shows the pole-zero configuration of the $(9,10)-$order PADÉ approximation for the transfer function $G_{hf}(s)$ and $\alpha = 0.5$ in (1). Traces in Figs. 4 correspond to two different values of parameter α, and visualize the outputs $x_{hf}(t)$ of the sampler and fractional-order hold given in Fig. 2, as well as the responses of the rational function $G_{f\,9,10}(s)$, corrected by T according to sampling theory. Note that the input $x(t)$ is a sinusoidal signal, as well as the sampling frequency is relatively low of 10 samples per cycle. It can be seen that in the case of negative value of parameter α the fractional-order hold introduced a much larger time lag to the input signal.

III. FREQUENCY RESPONSES OF FRACTIONAL-ORDER HOLD DEVICE

The frequency response data can be visualized in two different ways: via the NYQUIST diagram or via the BODE plots. Note that the BODE diagrams of the hold devices can be found mostly in the usual textbooks in the field of digital control systems.

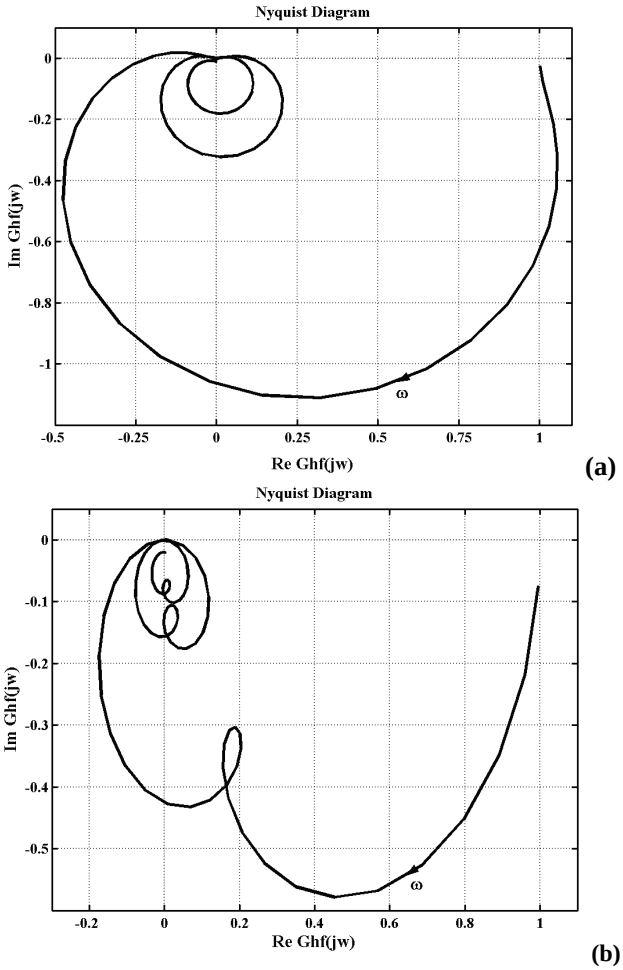
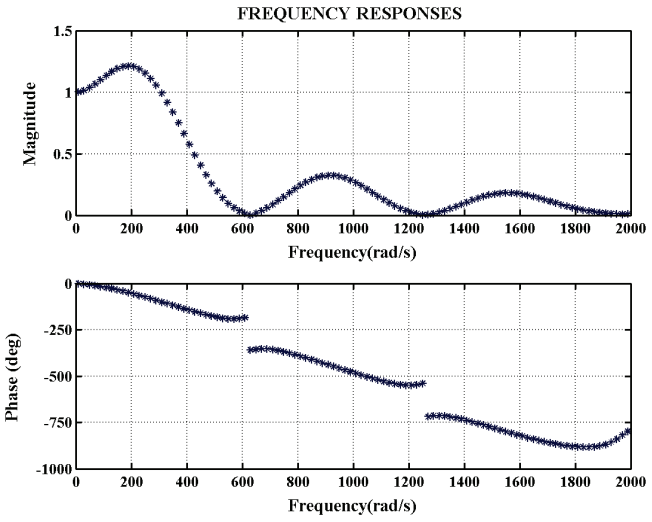
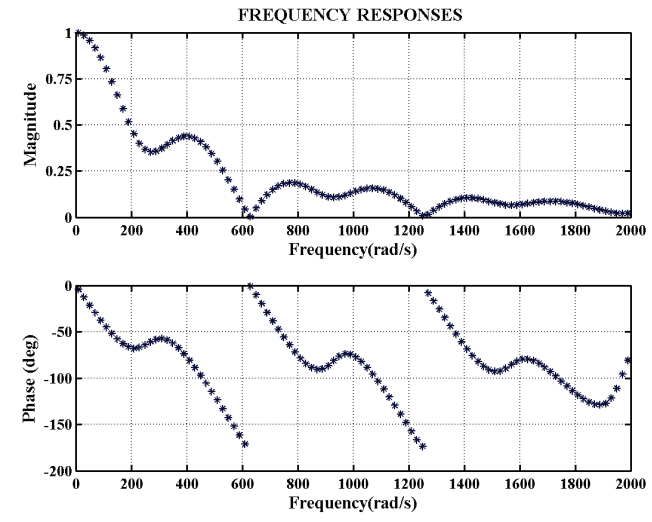


Fig. 5. The NYQUIST plots of the fractional order hold device for different values of parameter α : (a) $\alpha = 0.5$; (b) $\alpha = -0.5$.

In Figs. 5a and 5b it is possible to recognize the typical amplitude-phase frequency (NYQUIST) plots for systems of infinite dimension. Figs. 6a and 6b show plots of the magnitude and phase characteristics of the fractional-order hold device for the sampling period $T = 0.01\,\mathrm{s}$ and two values of the adjustable parameter $\alpha = \pm 0.5$. It can be seen that the fractional-order hold for both values of α does not have the ideal low-pass filter characteristics. Namely, the FROH device allows significant transmission above NYQUIST frequency $\omega_S/2 = \pi/T$. Thus, the signal must be low-pass-filtered before the sampling operation, so that the frequency components above the NYQUIST frequency are negligible. Moreover, in the case of negative value of parameter α, both amplitude and phase characteristics are significantly distored, even for frequencies which belong to the NYQUIST area of frequencies.



(a)



(b)

Fig. 6. Frequency responses of fractional order hold device for different values of parameter α : (a) $\alpha = 0.5$; (b) $\alpha = -0.5$.

Note that the time and frequency domain characteristics given in Figs. 4 and 6 for different values of parameter α fully correspond to each other.

IV. CONCLUSION

The reconstruction using the fractional-order hold is not very common in practice because its dynamics is relatively complicated to implement. Both the time and frequency domain characteristics of the considered hold device are analyzed for different values of the adjustable parameter. To generate the frequency response plots the well-known MATLAB® procedures are used. Bearing in mind that these procedures have several limitations specially in the case of nonrational transfer function, it would be suitable to consider the possibility of the development of some new procedures to generate the BODE and NYQUIST frequency response plots for nonrational transfer functions.

ACKNOWLEDGEMENT

This work was supported in part by Ministry of Science and Technological Development of the Republic of Serbia under the Grant ET 11029/2008-2010.

REFERENCES

[1] C. L. Phillips, and H. T. Nagle, Digital Control System Analysis and Design, Englewood Cliffs, N.J.: Prentice-Hall, 1995.

[2] M. B. Naumović, *Problem Solver in Digital Control Systems, Part I: Discrete Signals*, Faculty of Electronic Engineering, Niš, 1997 (in Serbian).

[3] M.B. Naumović, Z- *or Delta Transform?*, Faculty of Electronic Engineering, Niš, 2002 (Monograph in Serbian).

[4] M. Ishitobi, "Properties of Zeros of a Discrete-Time System with Fractional Order Hold", in Proc. 35th Conference on Decision and Control, Kobe, Japan, pp. 4339-4344,1996.

[5] M.B. Naumović: Some Discretizing Problems in Control Theory, ICEST 2007, Conference Proceedings, Vol. 2, pp. 573-576, Ohrid, Macedonia, 2007.

[6] R. Bárcena, M. De la Sen, I. Sagastabeitia, and J.M. Collantes, Discrete control for a computer hard disc by using a fractional order hold device, IEE Proc. Control Theory Appl., vol. 148, no. 3, pp. 117-124, 2001.

[7] M.B. Naumović, and B. Petrović, "An Experimental Setup for Studying Sampling and Signal Reconstruction Process", ICEST 2008, Conference Proceedings, Vol.1, pp. 211-214, Niš, Serbia, 2008.

[8] K. Basterretxea, R. Bárcena, and U. Ugalde, Design and Synthesis of a Configurable Fractional Order Hold Device for Sampled-data Control Systems, WSEAS TRANSACTIONS on CIRCUITS and SYSTEMS, vol. 7, no. 8, pp. 869-878, 2008.

[9] G. H. Golub, and C. F. Van Loan, *Matrix Computations*, Baltimore: Johns Hopkins University Press, 1983.

POSTER SESSION PO V

PO V – Robotics

Transfer Line for Manufacturing of Armatures for Electric Hand-Held Tools

Tchakarsky D., T. Vakarelska P. Tomov, R. Dimitrova,, I. Yanakiev[1]

Abstract: **Subject of the presented research is an armature set used in manual electric tools. A methodology for assembly automatic lines design is developed and tested in regard of the mentioned subject. The methodology is presented in a block format and all interactions between the single stages are shown. On hand of a preceding optimal route trajectory developed by the authors 3 variants of automatic lines were generated, analyzed, evaluated and the optimal one was selected.**

Keywords: **automation, automatic line, armature, variability, optimal variant, 3D model.**

I. INTRODUCTION

Assembly automation is a complex, versatile and very labor-consuming process. Hard work is needed to feed the assembly positions and to perform assembly operations. To overcome this situation all major and backup assembly operations must be automated.

Some essential problems must be considered, among them: choice of appropriate automation subjects, optimization of the technological process for automated manufacturing, generating of structural variants of automated Manufacturing systems, analysis and evaluation of automatic lines (AL) variants and choice of the optimal one, determination of basic features of the automated complexes.

In most cases automation is done on hand of existing machines. As a rule the latter must have an automated work cycle and, if possible, a programmable control unit.

II. PROBLEM BACKGROUND

Aim of this research work is to design an AL based on the newly developed technology for automated production of armature Ø53. Following typical actions were undertaken: developing of AL oriented methodology; testing the methodology and generating of AL variants; design of the first work position (machine for gouge isolation); design of aggregate for coupling the second work position (winding machine) with the AL; design of the forth work position (wedging machine); design of the third work position (machine for welding the coil-ends with the commutator). For every single work positions 3 D models were prepared and based upon them typical aggregates underwent engineering analysis.

The article features 3D models of just one work position (the fourth one – wedging machine) due to page number restrictions.

III. AL VARIANTS– ANALYSIS, EVALUATION AND CHOICE OF THE OPTIMAL ONE

Automated Manufacturing of high quality armature is based on a preliminary developed technological process realized on an automated complex of the type automatic line (AL). Such complex is very appropriate due to the differentiation of technological operations.

Prior to the AL design a matching methodology oriented at the automated production of the chosen subject needs to be developed. In the given case the object is **Armature Ø53** (Fig.1).

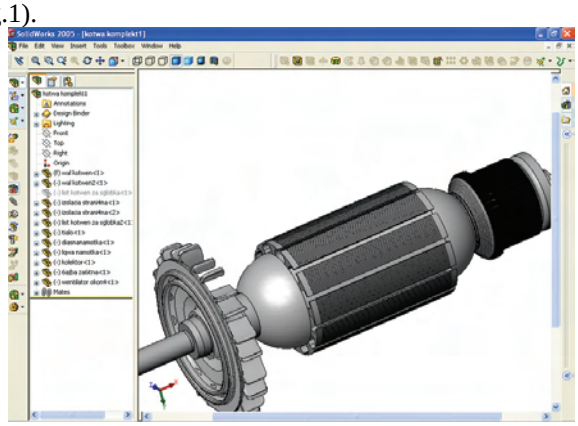


Fig.1 3D model of Armature Ø53

The methodology for AL design comprises:

- Synchronization and optimization of the technological process;
- Conceiving of AL variants;
- Defining the work positions number;
- Design of the single work positions;
- Design of the automated parts flow;
- Developing of the transportation and manipulation systems;
- Preparing of a preliminary specification for purchasable components of the intended AL ;
- Acquirement of the purchasable elements in line with the specification;
- Project conception development;
- Deliberation of variants, analysis, assessment and choice of the optimal variant;
- Simulation of the AL optimal variant functioning;
- Preparation of design documentation on hand of the ideational project;

[1] **Information about authors:**
Republic of Bulgaria, Technical University Sofia, 8-Kliment-Ohridski Blv., Department "ADP"
Phone/Fax: +359 2 965 36 85, e-mail: adp@tu-sofia.bg

- Development of routine technology for production of original aggregates and assembly groups;
- Preparation of specification for production of original aggregates and assembly groups;
- Conceiving of schedule for producing and testing of aggregates and assembly groups;
- Development of methods for control tests of each aggregate and assembly group;
- Initiating a Journal for precise and detailed recording of surveillance results and corrections ;
- Exercise of active authorial control during the parts production and assembly in line with the time-table;
- Performance of functionality tests and verification of aggregates and groups according to test methodologies;
- Undertaking of fundamental repair operations on existing production machinery intended as parts of the AL;
- Conceiving and coordinating of time schedules regarding assembly, tests and industrial implementation;
- AL assembly at the plant in cooperation with plant's representatives under active authorial control;
- AL programming and tuning;
- Arranging of functional tests and pivotal production;
- AL tests and exploitation start at the plant;
- Conceiving and applying of Instructions for safe exploitation of AL;
- Training of AL operators selected among the plant's staff;
- Assessment of the commercial efficiency of the AL implementation;
- Warranty maintenance and optimization of the AL;
- Conclusions from the AL implementation and conceptions for further researches.

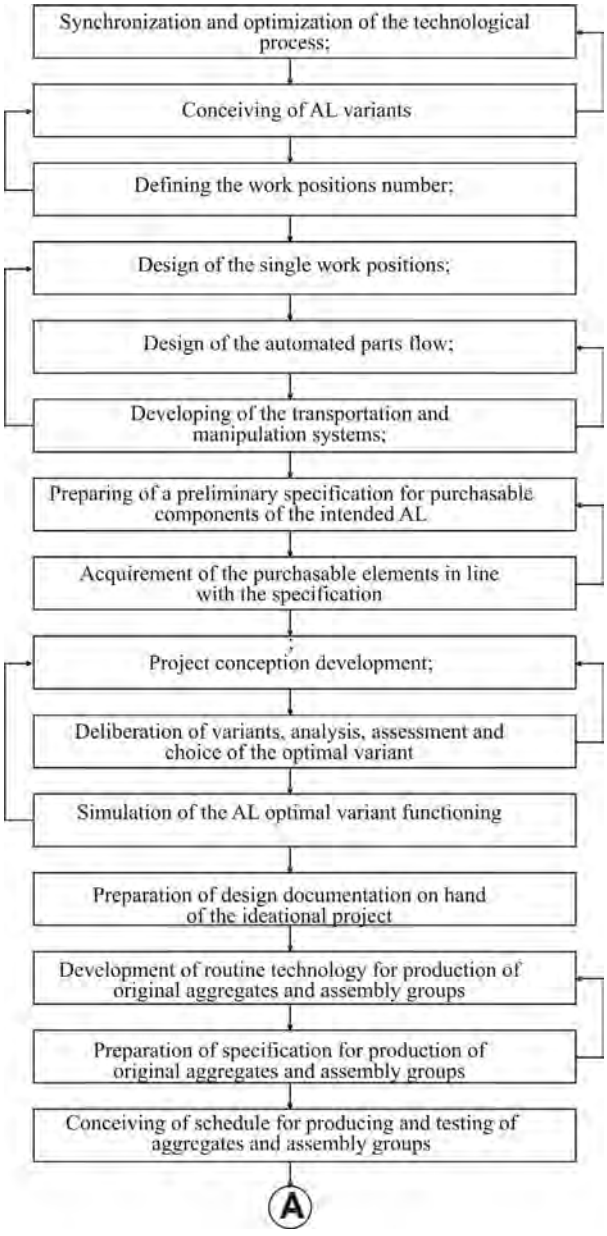


Fig.2a Methodology for design and implementation of Al for armature manufacturing

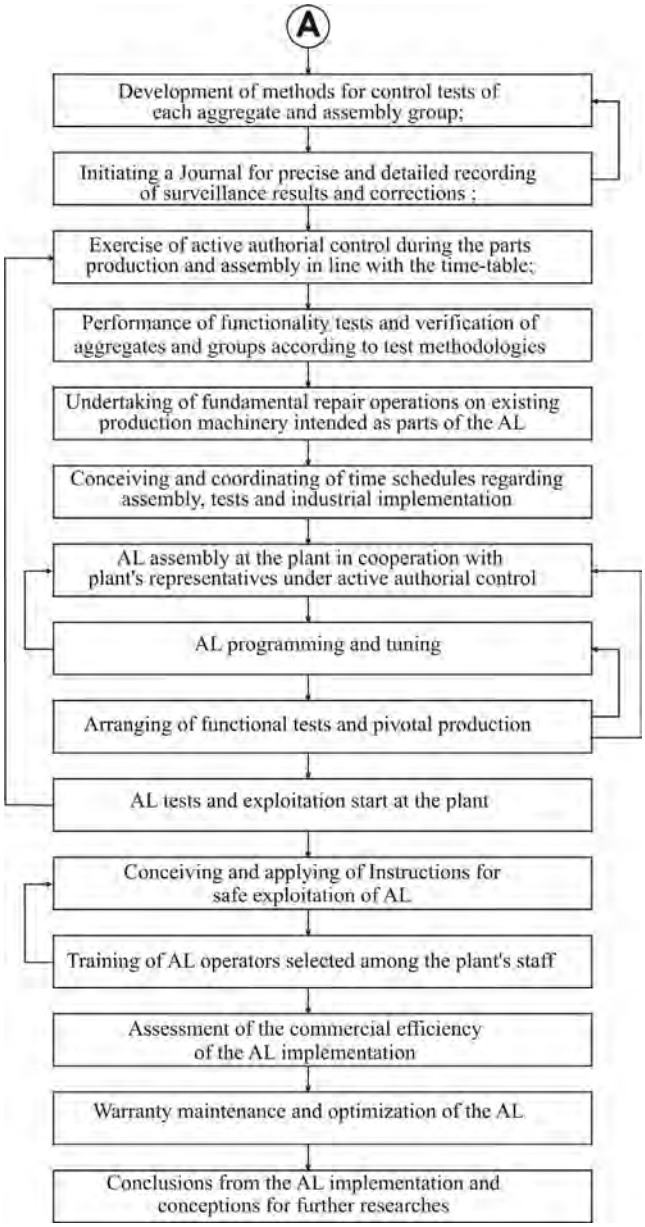


Fig.2b Methodology for design and implementation of Al for armature manufacturing

The above listed main stages and their interactions are visualized on Fig. 2a and Fig.2b.

The automation of the part flow must meet a series of requirements:

✓ The feeding with details automation must be "compatible" with the production machinery in a way that allows for repairing, tuning and operating activities;

✓ The feeding with details automation must provide for repair works and exploitation that are not depend on the machinery type;

✓ The feeding with details automation must lead to minimizing of machine delays in the Al;

✓ The feeding with details automation must guarantee a minimal change in the production machinery;

✓ The feeding with details automation must proceed with technical devices of a sophistication level similar or lesser than that of the existing machines (mechanical part, electrical control, pneumatic drive);

✓ The feeding with details automation must be least time and money consuming;

✓ The feeding with details automation must proceed with an optimal operations synchronization in order to minimize delays on the single work positions;

✓ The feeding with details must provide for a flexible interaction between work positions so that failure in one position does not harm the functioning of the others.

The assessment results for different variants of the technological process regarding the Manufacturing of high quality armature shows that the second one is optimal for a trajectory technology featuring a maximal summarized coefficient **K_{O2} = 1,182 (Table 1)**.

When generating AL variants for the production of high quality armature the bellow listed variables are put in use:

- Structural units type;
- Structural units model;
- Structural units drives;
- Structural units control;
- Transportation and manipulation modes
- Mutual positioning of structural units
- Types of positioning and fixing of processed parts
- Types of changing the automation objects;
- etc.

The basic AL variant includes a number of structural units:

- Work position for fixing of isolation sideline plates and the commutator.
- Work position for fixing full-scale groove isolation.
- Work position for winding the armature coils – 2 pieces.
- Work position for welding the coil-ends to the commutator channels.
- Work position for wedging the armature channels through isolation plates.
- Work position for testing the armature electrical parameters.
- Work position for armature impregnation with polyester pitch.
- Work position for fixing of disk and ventilator.
- Work position for armature balancing.
- Work position for end-control the armature electrical parameters.
- Stepper transporter.
- AL automatic control system.
- Work position for bearing fixing.

Table 1 Optimal trajectory technological process

№	Technological operations	Machine times [s]
1.	**Operation 1** - fixing of isolation sideline plates on the commutator – left and write and **Operation 2** – fixing the commutator	30
2.	**Operation 3** –fixing the full-scale groove isolation	12
3а.	**Operation 4** – winding of armature coils	48
3б.	**Operation 4** – winding the armature coils	48
4.	**Operation 5** – welding the coil-ends to the commutator channels	48
5.	**Operation 6** – wedging the armature channels through isolation plates	12
6.	**Operation 7** – testing the armature electrical parameters.	12
7.	**Operation 8** – armature impregnation with polyester pitch.	20
8.	**Operation 9** – fixing of disk and **Operation 10** – fixing of ventilator	30
9.	**Operation 11** – armature balancing	44
10.	**Operation 12** – end-control the armature electrical parameters	10
11.	**Operation 13** – fixing of a bearing	5

On Fig.3 the principal scheme of the basic AL variant is displayed.

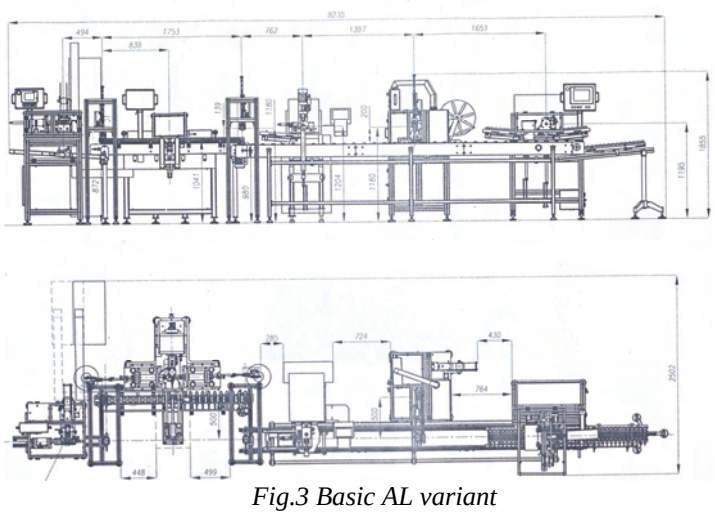


Fig.3 Basic AL variant

The basic variant poses as Variant 1:

- Work position for fixing of isolation sideline plates on the commutator.
- Work position for winding the armature coils – 2 simultaneously functioning pieces.

- Work position for wedging the armature channels through isolation plates.
- Work position (test stand) for end-control the armature electrical parameters.

The established work positions are modernized as follows:

- Work position for fixing of isolation sideline plates and the commutator.
- Work position for welding the coil-ends to the commutator channels.
- Work position for armature impregnation with polyester pitch.
- Work position for fixing of disk and ventilator.
- Work position for armature balancing.

The conception intends a manipulation and feeding system on each working station. The systems consist of industrial robots plus diverse automation devices executing according functions.

Two more variants with bellow listed features were generated:

- According to variant two all structural units must be repeatedly designed excluding the old work positions.
- Variant three is similar to the basic one except for the bearing assembly position..

Electronic control system variants meant for a concurrent, double-sided, co-axis processing of openings in closed constructions are assessed upon criteria including:

- Productivity
- Reliability
- Automation level
- Flexibility level
- Economic indicators

For designing of automation devices the most frequently used is the ***cycle productivity*** $\mathbf{Q_ц}$.

$$Q_Ц = \frac{1}{t_Ц} = \frac{1}{t_р + t_{сп.н}}$$

where:

$t_ц$ cycle time;

$t_р$ – machining time;

$t_{сп.н}$ – supporting non-overlapping time

An automation of existing units aims a maximal reduction of $t_{сн}$ in order to accomplish the highest possible productivity level.

With the development of modern automation devices the accomplishment of minimal cycle times is intended, i. e. both $t_р$ and $t_{сн}$ to be of minimal values.

To reach **maximum productivity** levels the tact **τ** must be minimized.

$$\tau = \max\{t_{цi}, i = 1 \div m\} \rightarrow [\min]$$

Productivity is defined by the equation:

$$Q = 3600 / \tau \ [pc/h]$$

The cycle duration of the technological process of cleaning metallic surfaces before painting them depends on:

$$t_ц = \sum_{i=1}^{m} t_{цi} \rightarrow [\min]$$

where:

i – number of sequence of the technological operation;

$t_ц$ – total cycle time [s];

$t_{цi}$ cycle time of the i – numbered operation;

m – number of operations.

In many cases the main indicator for the productivity level is the rise in productivity λ, which considers the latter as a result of the automation compared with the status quo, i. e.:

$$\lambda = \frac{Q_a}{Q_0}$$

where:

Q_a – productivity level in the variant with automation

Q_0 - in the existing situation

For choosing an effective AL variant for Manufacturing of high quality armature relative (non-scaled) coefficients are applied and the summarized coefficient $\mathbf{K_{oi}}$ for each of the discussed variants i (i = 1÷ m; m – number of variants) is calculated.

$$K_{oi} = \prod_{j=1}^{n} (K_{ij}), \ i = 1 \div m$$

where:

n – number of non-scaled coefficients

As optimal is regarded the variant, which features a maximal coefficient $\mathbf{K_{oi}}$, i. e.:

$$\mathbf{max\ \{\ K_{oi}\ ,\ i\ =\ 1\ \div\ m\ \}}$$

Following criteria (non-scaled coefficients) are applied:

$$\begin{cases} K_1 = \lambda \\ K_2 = K_Г \\ K_3 = K_A \\ K_4 = (1 - K_G) \\ K_5 = 1/n \end{cases}$$

The described approach is more objective since it eliminates the factor of subjectivity and provides for the choice of the optimal solution.

Table 2 features the quantity values of non-scaled coefficients as well as the summarized coefficient for the separate variants.

Table 2 Quantity values of non-scaled coefficients

K_i / V_i	K_1	K_2	K_3	K_4	K_5	K_{Oi}
V_1	2,1	0,9	0,9	0,8	0,35	0,476
V_2	2,2	0,92	0,9	0,8	0,25	0,364
V_3	2	0,93	0,85	0,75	0,30	0,356

max { K_{oi} , i = 1 ÷ m } = max {0,476; 0,364; 0,356} = 0,476

The assessment results for all variants prove that the first variant of AL for Manufacturing of high quality armature Ø53 is the optimal one hence it features a summarized coefficient $\mathbf{K_{O1} = 0,476}$.

IV. Creating 3D models of a work position for automated wedging of gouge isolation into the gouge

Fig.4 shows the 3D model of a work position for automated wedging of gouge isolation material into the gouge. Backgrounds for creating the 3D model of the position are the basic armature parameters. The 3D models were created with the help of the CAD System Solid Works.

The work position for automated wedging of gouge isolation into the gouge comprises the following basic aggregates:

- 3D model of the base - Fig. 5
- 3D model of the carriage - Fig.6
- 3D model of the fixing mechanism - Fig.7
- 3D model of a mechanism for cutting and bending - Fig.8
- 3D model of a mechanism for submitting the wedge - Fig.9
- 3D model a mechanism for cutting the wedge - Fig.10
- 3D model of a mechanism for submitting the wedge material - Fig.11
- 3D model of a manipulator - Fig.12
- 3D model of a control desk - Fig.13

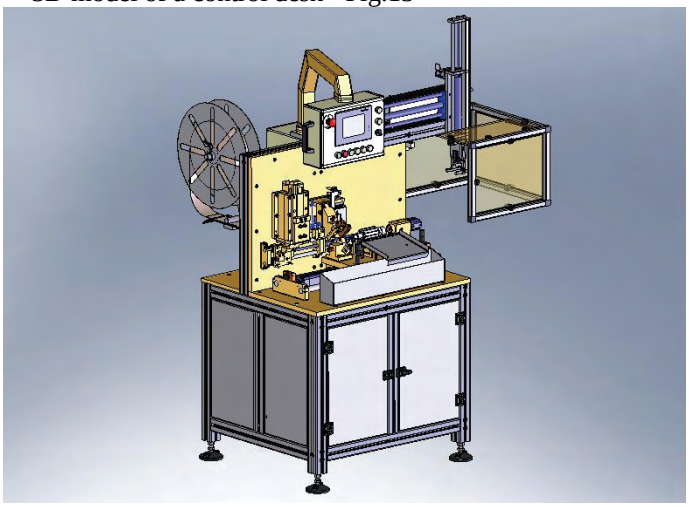


Fig. 4 3D model of a wedging machine

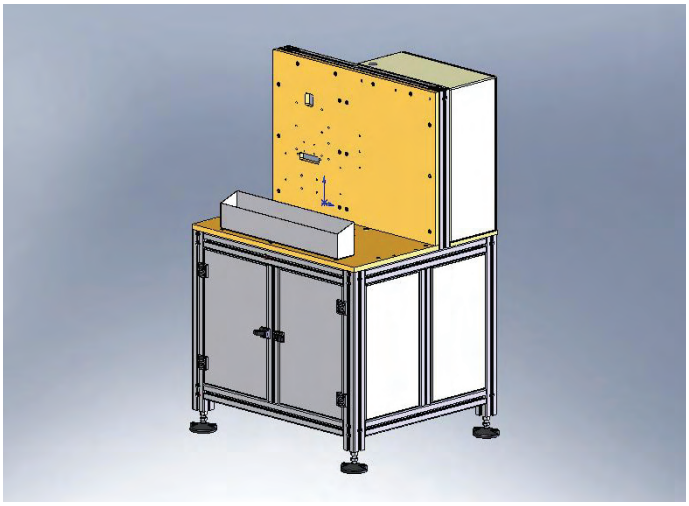


Fig.5 3D model of a base

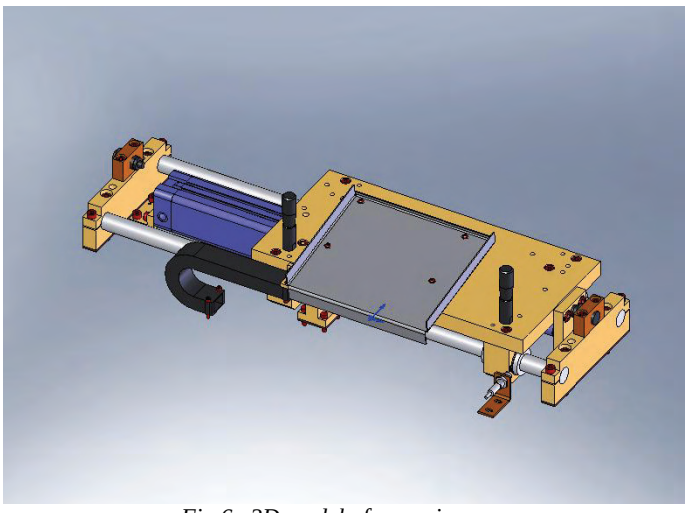


Fig.6 3D model of a carriage

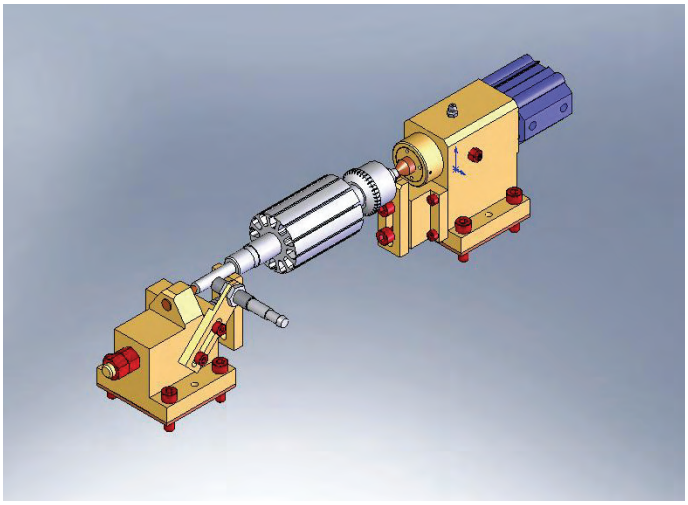


Fig.7 3D model of a fixing mechanism

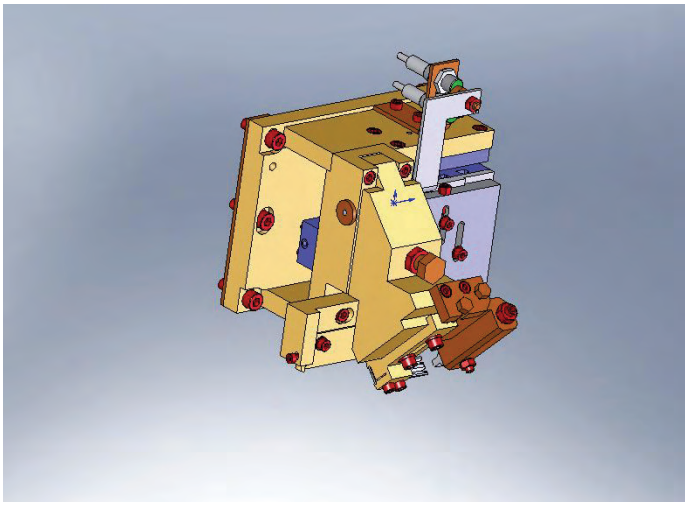


Fig.8 3D model of a mechanism for cutting and bending

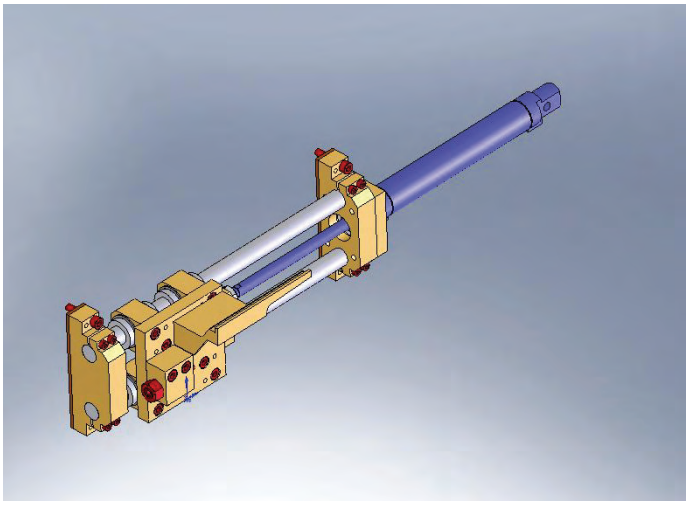


Fig.9 3D model of a mechanism for submitting the wedge

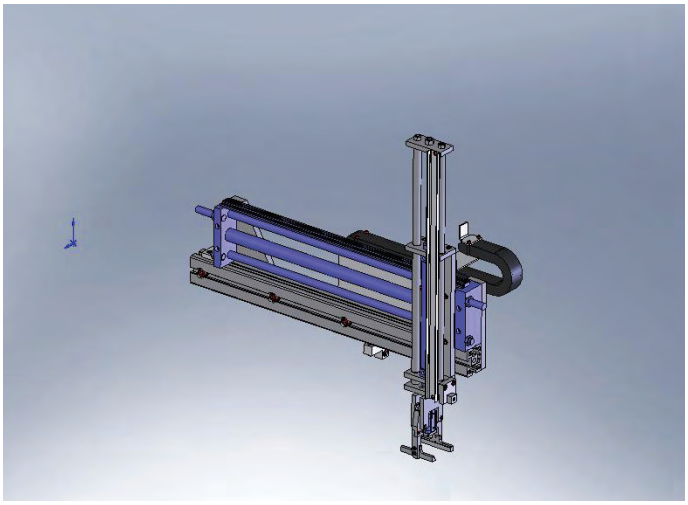


Fig.12 3D model of a manipulator

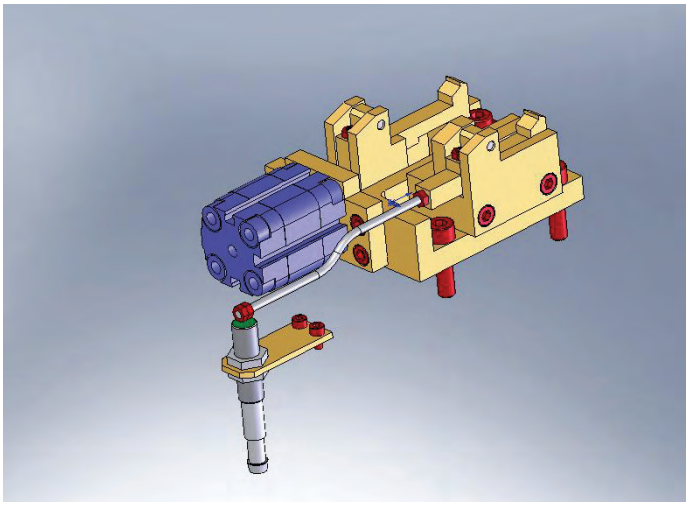


Fig.10 3D model a mechanism for cutting the wedge

Fig.13 Prototype of automated lain

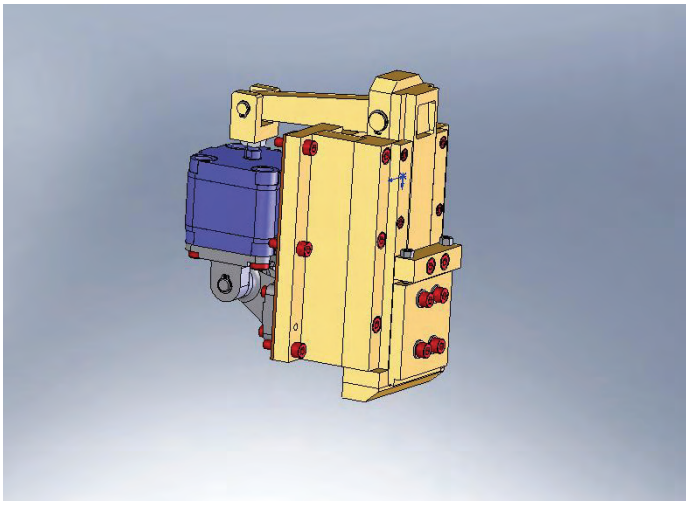


Fig.11 3D model of a mechanism for submitting the wedge material

Following the 3D models creation static and dynamic analysis were undertaken and the design documentation for an AL for armature Manufacturing were completed.

V. CONCLUSIONS

- Generated is an optimal technological process for automated assembly of an electrical engineering armature.
- Developed is a methodology aimed at designing automatic assembly lines. The methodology is tested in the armature assembly.
- Generated are variants of AL for armature manufacturing. The same are analyzed and evaluated and the optimal one is selected on the ground of the non-scaled-coefficient method.
- Created are 3D models of a work position for automated wedging of gouge isolation into the gouge.

REFERENCES

1. Чакърски Д. и к-в. Комплексна автоматизация на дискретното производство. С., ИК на ТУ, 2005.
2. Николов С. Ръководство за лабораторни упражнения по CAD/CAM системи в машиностроенето и уредостроенето. Пловдив, 2006.
3. Banerjee P.,Zetu D. Virtual Manufacturing. John Wiley & Sons Inc. №1, 2002

Object-Oriented Engineering Test-Stand for a Double-Sided Co-Axial Processing of Openings in Products of a "Column" Type

Vakarelska T., D. Tchakarsky, P Tomov, R. Dimitrova, I. Yanakiev[1]

Abstract: **The report deals with methodology for design of a specialized stand. The problem for achieving the required accuracy is solved when manufacturing concentric holes in large welded parts. The suggested concept assures simultaneous manufacturing of the holes using special equipment eliminating in this way the mistakes from multiple positioning.**

Key words: **automation, object-oriented engineering test stand, "column" type products, co-axial processing, 3D models**

I. INTRODUCTION

The processing of co-axial openings with classical methods often leads to inaccuracies negatively affecting the product qualities. This is the case especially with welded objects where lingering tensions can be observed. Objects of the presented research are column-type products that are distinguished through their large sizes and significant weights.

II. PROBLEM STATING

Aim of this research is the building of an engineering test stand (ETS) for automated processing of openings in welded "column" type constructions and to identify their disadvantages. The following action types were performed: defining of technical parameters for the engineering test stand; defining of requirements towards the test stand (technical, economical etc.); development and testing of a methodology; conceiving of structural variants; variant analysis and choice of the optimal one; completing of technological documentation for original parts Manufacturing; completing the Manufacturing schemes and drawings; conceiving of typical units and aggregates.

III. DEFINITION OF TECHNICAL PARAMETERS AND REQUIREMENTS FOR THE TEST STAND

No preliminary inquiries suggested the presence of any technical solutions allowing a simultaneous double-sided coaxial processing of welded products.

The test stand must consists of:

- **Body** providing possibilities for fixing and functionally combining transporting and power components. The body has to be composed on the base of special corpus elements;
- **Transportation-fixing elements** allowing to pass the welded construction between work positions, to allocate and precisely fix it according to a determined in advance position and to guarantee a resistance to cutting forces. Those elements include:
  - **One horizontal linear table** for transportation (passing) alongside the X- axis;
  - **Two vertical tables** – for transportation (passing) of power elements alongside axis У and $У_1$ via a movable carriage on roller guideways; the motion is created by an AC motor with break gears and gear reducer with a ball screw; the change of position is accounted for by a increment line;
  - **Two movable carriages** driven by an AC motor with break gears and gear reducer with a ball screw; The exact positioning is guaranteed by hydraulic break gears.
  - **Two turntables** – for securing the moving and passing of tools along axis Z and W. They are driven by AC motors and the position changes are accounted for through PES.
- **Power components** for providing the basic rotational cutting motion; the cutting is performed by two power heads driven by AC motors; the change of tools follows over a hydraulic collet; the spindle balancing is mechanically done;

The design and building of ETS requires:

- To conceive an one-position test stand of the "combine" type permitting the execution of various technological processes and operations like: boring, chafing, jig-boring, drilling through interpolation (spiral-like), screw threading.
- To guarantee technical potential allowing the performing of different operations on the test stand:
  - Boring of holes with diameters in the range from 3 to 50 [mm] and standard tolerance level IT12.
  - Chafing – openings with diameters in the range from 3 to 50 [mm] and standard tolerance level IT6 and rancidity Ra from 1.25 [μm] to 0,32 [μm].

[1] **Information about authors:**
Republic of Bulgaria, Technical University Sofia, 8-Kliment-Ohridski Blv., Department "ADP"
Phone/Fax: +359 2 965 36 85, e-mail: adp@tu-sofia.bg

Jig-boring – openings with diameters up to 150 [mm], maximal tolerance level IT6, level 6 in geometrical accuracy and rancidity R_a=1,6 [μm]

- Milling – surfaces with rancidity R_a=5 [μm]
- Cutting of screw treads in the range M4 to M32 and level 6 in the geometrical
- Positioning accuracy – ±0.005[mm]
- Repeatability of positioning –±0.005[mm]

The ETS design seeks to introduce a planned innovative technology and to satisfy several economic requirements and indicators:

- Rise in productivity - double;
- Rise in quality and reduction of droppings – 20%;
- Investment reimbursement – 3 years;
- Exploitation duration – 15 years.

IV. CONCEIVING OF STRUCTURAL AND COMPOSITION VARIANTS – METHODOLOGY, TESTING, ANALYSIS AND CHOICE OF AN OPTIMAL VARIANT.

Prior to the AL design a matching methodology oriented at the automated production of the chosen object needs to be developed. In the given case the object is a product of the "column" type.

The methodology for design of an engineering test stand for (ETS) for automated processing of openings in welded "column" type constructions comprises stages listed below:

- Synchronization and optimization of the technological process;
- Generating of engineering ideas;
- Conceiving of variants;
- Definitions of parameters for the working zone;
- Definition of kinetic parameters;
- Ideation project generating;
- Variant analysis and evaluation;
- Choice of the optimal variant;
- 3D model of the chosen optimal EST variant;
- Simulation of the EST functioning;
- Definition of main EST indicators;
- Design of basic EST units;
- Design of fixing and positioning devices;
- Design of EST equipment;
- Optimization of the work cycle;
- Preparing of a preliminary specification for purchasable components of the intended EST;
- Design of CNC system for EST;
- Acquirement of the purchasable elements in line with the specification;
- Preparation of design documentation on hand of the ideational project;
- Development of routine technology for production of original aggregates and assembly groups;
- Preparation of specification for production of original aggregates and assembly groups;
- Conceiving of schedule for producing and testing of aggregates and assembly groups;
- Development of methods for control tests of each aggregate and assembly group;
- Initiating a Journal for precise and detailed recording of surveillance results and corrections ;
- Exercise of active authorial control during the parts production and assembly in line with the time-table;
- Performance of functionality tests and verification of aggregates and groups according to test methodologies;
- Undertaking of fundamental repair operations on existing production machinery intended as parts of the EST;
- Conceiving and coordinating of time schedules regarding assembly, tests and industrial implementation;
- EST assembly at the plant in cooperation with plant's representatives under active authorial control;
- EST programming and tuning;
- Arranging of functional tests and pivotal production;
- EST tests and exploitation start at the plant;
- Conceiving and applying of Instructions for safe exploitation of EST;
- Training of EST operators selected among the plant's staff;
- Assessment of the commercial efficiency of the EST implementation;
- Warranty maintenance and optimization of the EST;
- Conclusions from the EST implementation and conceptions for further researches.

The automation of the part flow must meet a series of requirements:

- ✓ The feeding with details automation must be "compatible" with the production machinery in a way that allows for repairing, tuning and operating activities;
- ✓ The feeding with details automation must provide for repair works and exploitation that are not depend on the machinery type;
- ✓ The feeding with details automation must lead to minimizing of machine delays in the POAWS;
- ✓ The feeding with details automation must guarantee a minimal change in the production machinery;
- ✓ The feeding with details automation must proceed with technical devices of a sophistication level similar or lesser than that of the existing machines (mechanical part, electrical control, pneumatic drive);
- ✓ The feeding with details automation must be least time and money consuming;
- ✓ The feeding with details automation must proceed with an optimal operations synchronization in order to minimize delays on the single work positions;
- ✓ The feeding with details must provide for a flexible interaction between work positions so that failure in one position does not harm the functioning of the others.

For the products considered in this article the basic technological route sets Varian 1. In Variant 2 the technological operations are performed consecutively and in any given moment a toll positioned in one of the spindles is active. In Variant 3 small openings are processed simultaneously multi-spindle devices.

The assessment of results for different variants of the technological process shows that the first one is optimal for a

trajectory technology featuring a maximal summarized coefficient $\mathbf{K_{O1} = 0,396.}$

Table1 shows the application of typical technological route for double sided coaxial processing of the product type "Column 6040818".

When generating variants of an ETS for double sided coaxial processing the bellow listed variables are implemented:

- Structural units type;
- Structural units model;
- Structural units drives;
- Structural units control;
- Transportation and manipulation modes
- Mutual positioning of structural units
- Types of positioning and fixing of processed parts
- Types of changing the automation objects;
- Etc

Fig 1 displays principle schemes of the single variants. The variant shown in Fig. 1 is considered as a basic one.

➢ The basic variant (Variant I) comprises following components:

- Base;
- Longside table with transmission;
- Cross-table with 2 transmissions – left-handed and right-handed;
- Columns with spindle holding transmissions - left-handed and right-handed;

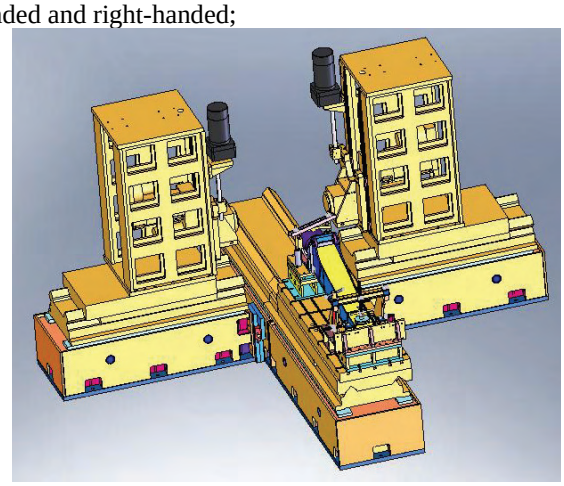


Fig. 1. 3D Model of the basic variant of an ETS for double sided coaxial processing of openings in welded constructions

- Spindles with drives - left-handed and right-handed;
- Positioning and fixing device for processing the product type "Column";
- CNC system.

Table 1 Typical technological route for double sided coaxial processing of the product type "Column 6040818"

№	Operation title	Tool		Cutting mode		
				V[m/min]	**S**[mm/r]	**n**[min^{-1}]
0	Fixing the part in the device in bases A and B			-	-	-
20	Part centring			-	-	-
30	Jig-boring of opening A	ТИ 002-01-2004	Г	60	0,1	300
			Ч	80	0,07	390
40	Jig-boring of opening B	ТИ 002-01-2004	Г	60	0,1	390
			Ч	80	0,07	510
50	Jig-boring of opening C"	ТИ 002-01-2004	Г	60	0,1	325
			Ч	80	0,07	425
60	Boring of 2 openings X	Borer	Ц	30	0,05	910
			П	20	0,07	750
70	Boring of 2 openings Y	Borer		-	-	-
80	Drilling of surface A"	Front milling machine		50	0,2	320
90	Boring of 3 openings on "Surface A"	Borer	Ц	30	0,05	765
			Г	20	0,07	620
100	Manual - chamfer 1x30° for opening "B"	Manual shutter		-	-	-
110	Cleaning of all sharp edges			-	-	-
120	Self regulation			-	-	-
130	Filling in of individual control card			-	-	-
140	Part unloading			-	-	-
150	Cleaning of the device and machine table			-	-	-
160	Cleaning of particles from the rough processing of openings A and B, change of tools			-	-	-
Legend: Г – rough processing, Ч – clean processing, Ц – central boring, П passage boring Г – blind boring						

Two more variants with typical features were generated:

Variant II – Used are the same structural elements as in Variant 1.This variant intends 2 devices for positioning and fixing of the processed part and a table with extended length. When a part is processed on the second ETS device it is possible to remove the finished product and to proceed with the next one's positioning and fixing. Thus the supporting non-overlapping times needed to change objects can be reduced.

Variant III uses the same structural units. Two spindles process the large openings while two more do the same with the smaller ones. Thus the technological cycle time can be diminished, yet on the other hand the equipment's design becomes more complex and complicated. The complications with the design are due to the fact that for proper functioning the spindles must be granted independence from each other.

The variants of an ETS for double sided coaxial processing are assessed upon criteria including:

- Productivity
- Reliability
- Automation level
- Flexibility level
- Economic indicators

For choosing an effective ETS variant for double sided coaxial processing relative (non-scaled) coefficients are applied and the summarized coefficient $\mathbf{K_{oi}}$ for each of the discussed variants i (i = 1÷ m; m – number of variants) is calculated.

$$K_{oi} = \prod_{j=1}^{n} \left(K_{ij}\right) \text{ , i = 1÷ m}$$

where:

n – number of non-scaled coefficients

As optimal is regarded the variant, which features a maximal coefficient $\mathbf{K_{oi}}$, i. e.:

$$\mathbf{max \{ K_{oi} , i = 1 \div m \}}$$

Following criteria (non-scaled coefficients) are applied:

$$\left|\begin{array}{l} K_1 = \lambda \\ K_2 = K_\Gamma \\ K_3 = K_A \\ K_4 = \left(1 - K_G\right) \\ K_5 = 1/n \end{array}\right.$$

The described approach is more objective since it eliminates the factor of subjectivity and provides for the choice of the optimal solution.

In our case the method of non-scaled coefficients will be applied in order to perform a qualitative analysis prior to choosing the optimal variant. Table 2 features the quantity values of non-scaled coefficients as well as the summarized coefficient for the separate variants.

Assessed are the next three variants:

- **Variant 1** – Basic variant featuring one device for fixing the processed part and two swindles for parallel processing of the openings.
- **Variant 2** – features 2 sections for fixing the processed parts.
- **Variant 3** – characterized by 4 swindles: 2 for the large and 2 – for the small openings.

Table 2 Quantity values of non-scaled coefficients

$\mathbf{K_i}$ / $\mathbf{V_i}$	$\mathbf{K_1}$	$\mathbf{K_2}$	$\mathbf{K_3}$	$\mathbf{K_4}$	$\mathbf{K_5}$	$\mathbf{K_{Oi}}$
$\mathbf{V_1}$	**2,1**	**0,9**	**0,7**	**0,7**	**0,3**	**0,278**
$\mathbf{V_2}$	**2,3**	**0,85**	**0,8**	**0,7**	**0,2**	**0,219**
$\mathbf{V_3}$	**2,5**	**0,75**	**0,7**	**0,7**	**0,15**	**0,138**

$$\mathbf{max \{K_{oi}, i =1 \div m \} = max\{0,278; 0,219; 0,138\} = 0,278}$$

The assessment results for all variants prove that the first ETS variant for double sided coaxial processing of is the optimal one due to the summarized coefficient value $\mathbf{K_{O1}}$ **= 0,278.**

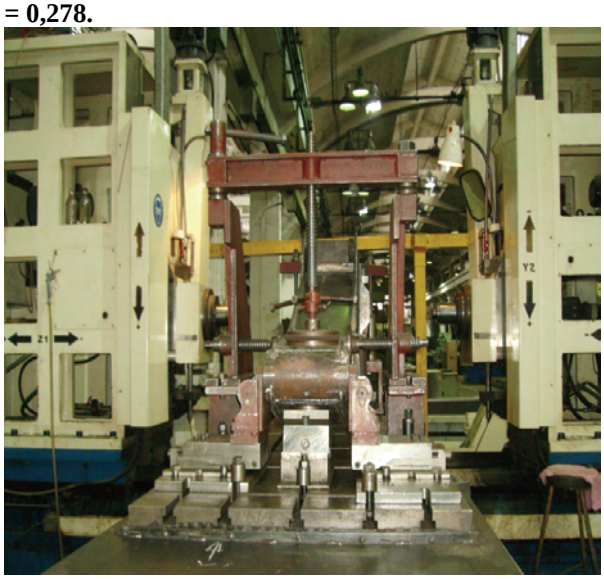


Fig. 2. Prototype of automated machine

V. PLANNED PARAMETERS FOR THE PLANNED ETS

The design and building of the test stand must reach following technical characteristics:

- Controllable axis - 5;
- swindles 2;
- simultaneously controlled axis - 3;
- maximal range of the tools:
  - along axis X-2700 [mm]
  - along axis Y_1 and Y2 -720 [mm]
  - along axis Z и W -500 [mm]
- functional passage along the axss:
  - по ос X -1 ÷2000 [mm/min]
  - по ос Y_1 и Y_2 - 1÷6000 [mm/min]
  - по ос Z и W -1 ÷ 4000 [mm/min]
- скорост на бързите премествания по осите
  - along X-up to 2 [m/min]
  - along Y_1 and Y_2- 6÷9 [m/min]
  - along Z and W -4÷5 [m/min]
- swindle spins -10÷2000 [min^1]

CONCLUSIONS

- Developed is a methodology aimed at designing an engineering test stand (ETS) for automated double sided coaxial processing of openings in welded "column" type constructions.
- Generated are variants of an engineering test stand (ETS) for automated double sided coaxial processing of openings in welded "column" type constructions. The same are analyzed and evaluated and the optimal one is selected on the ground of the non-scaled-coefficient method.
- Created is 3D model of the basic variant of an engineering test stand (ETS) for automated double sided coaxial processing of openings in welded "column" type constructions. The model is used for performing of engineering analysis, generating of control software and completing the Manufacturing documentation.

Gesture-based Human-Robot Interaction and Control

Saso Koceski[1], Natasa Koceska[1]and Predrag Koceski[1]

***Abstract* – Aiming at the use of hand gestures for human-computer interaction, this paper presents a novel approach for hand gesture-based control of mobile robot's freight ramp. The research was mainly focused on solving some of the most important problems that current HRI (Human-Robot Interaction) systems fight with. Presenting a simple approach to recognizing gestures through image processing techniques and a single web camera, we address the problem of hand gestures recognition using motion detection and algorithm based on histograms, which makes it efficient in unconstrained environments, easy to implement and fast enough.**

***Keywords* – Human-Robot Interaction (HRI), gesture-based control.**

I. INTRODUCTION

Advances in computer technology, artificial intelligence, speech simulation and understanding, and remote controls have led to breakthroughs in robotic technology that offer significant implications for the human computer interaction community.

Human-Robot Interaction (HRI) can be considered as one of the most important Computer Vision domains. In HRI based systems, the communication between human operators and robotic systems should be done in the most natural way. Typically, communication is done through hands/head postures and gestures. This type of communication provides an expressive, natural and intuitive way for humans to control robotic systems. One benefit of such a system is that it is a natural way to send geometrical information to the robot, such as: up, down, etc. Gestures may represent a single command, a sequence of commands, a single word, or a phrase and may be static or dynamic. Such a system should be accurate enough to provide the correct classification of hand gestures in a reasonable time. Human-robot interaction using hand gestures provides a formidable challenge. This is because the environment contains a complex background, dynamic lighting conditions, a deformable hand shape, and real-time execution requirement. There has recently been a growing interest in gesture recognition systems. For example, Stergiopoulou and Papamarkos [1] proposed YCbCr segmentation. In [2] a real-time hand posture recognition using 3D range data analysis is proposed. Ribeiro and Gonzaga [3] proposed different approaches of real time GMM (Gaussian Mixture Method) background subtraction algorithm using video sequences for image segmentation. Mariappan [4] uses motion detection algorithm for gesture recognition. Popa et al. [5], proposed trajectory based hand gesture recognition using kernel density estimation and the related mean shift algorithm. Bugeau and Pérez [6] proposed a method for detecting and segmenting foreground moving objects in complex scenes using clusters.

The main objective of this work was the developing of a control system for a robot freight ramp, based on gesture recognition. We aim to develop a cheap and robust control system without using data gloves or colored gloves, or other devices. With that purpose, we decided to use a generic webcam for the image acquisition process, and we have defined a gesture vocabulary for the telerobotic control, using motion detection and gesture recognition algorithm based on histograms, which makes it suitable for real time control, easy to implement and efficient in unconstrained environments.

II. FREIGHT RAMP WORKING PRINCIPLE

Since, this work is focused on the HRI and the controller development, the details about the conducted research on the ramp mechanics, will not be discussed here. Instead of that, the main working principle of the mobile robot's ramp will be described. As shown on Fig. 1. the robot ramp has two degrees of freedom:

-rotation around the Y axis which can be obtained by one stepper motor

-translation along the Z axis which can be obtained by stepper-based linear actuators.

In classic mechanical engineering, linear systems are typically designed using conventional mechanical components to convert rotary into linear motion. Converting rotary to linear motion can be accomplished by several mechanical means using a rotary motor, rack and pinion, belt and pulley, and other mechanical linkages, which require many components to couple and align. Although these methods can be effective, they each carry certain limitations. Conversely, stepper motor-based linear actuators address all these factors and have fewer issues associated with their use. The reason is that rotary-to-linear motion is accomplished in the motor itself, which translates to fewer components, high force output, and increased accuracy. The ramp has a warning light

[1]Saso Koceski, Natasa Koceska and Predrag Koceski are with the Faculty of Computer Sciences, University "Goce Delcev" – Stip, Krste Misirkov bb, 2000 Stip, Macedonia, E-mail: saso.koceski@ugd.edu.mk, natasa.koceska@ugd.edu.mk, pkoceski@gmail.com

which is activated whenever the ramp performs a movement and is deactivated when the ramp has finished the last movement. This light can be used as a visual feedback for the operator. The ramp is equipped with a sensor, which measures the angle between 90° ... -10°. In open position the angle is 5°. In closed position the angle is 90°. Another position sensor is used to calculate the vertical position (pos) as a value between 0 ... 100%. If the lift is up the position is 100%. If the lift is down the position is 0%. It is only possible to operate the lift upwards / downwards if the door is fully opened. If the door is open the tilt can be adjusted between +5° and -10° with the resolution of 3°. Adjusting the tilt is only possible if the door is fully open.

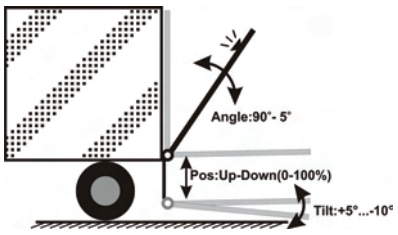


Fig. 1. Freight ramp working principle

III. Hand Gesture Language

Application specific vocabulary of 12 gesture poses (Fig. 2) was designed for robot freight ramp control tasks. The first two gestures control the ramp opening and closing. Next two gestures move the ramp up and down. Stop hand gesture stops any action the ramp performs moving the system into the Stopped position. StopTilt gesture stops any tilt movement the ramp performs and takes the ramp into the Stopped position. All other hand gestures control the tilt angle of the ramp.

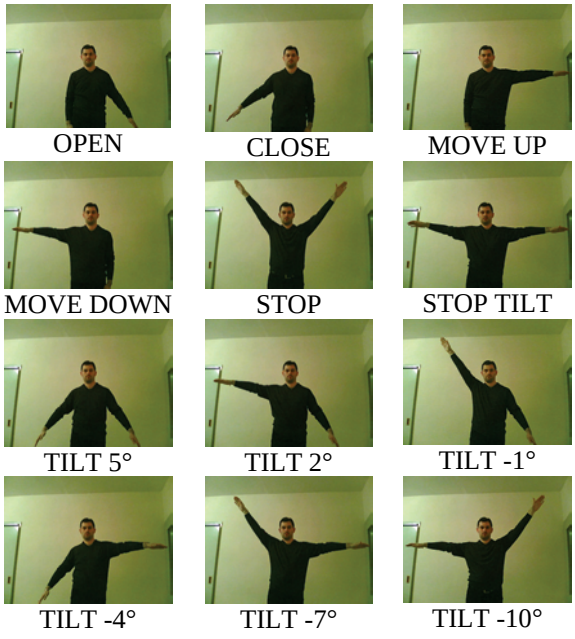


Fig. 2. Hand gesture vocabulary

IV. Control System Architecture

The system architecture is based on client-server model Fig. 3. To control a robot's ramp movement, the operator evokes a gesture from gesture vocabulary. The user performs his hand gestures in front of a web cam, which captures a video stream with a frame rate of 30fps. The motion detection is performed and hand gesture is classified by the client control software application, which is running on a client console.

Each recognized gesture is converted into a command, and then it's sent through TCP/IP protocol to a distant robot PC server for execution by the robot Fig. 4.

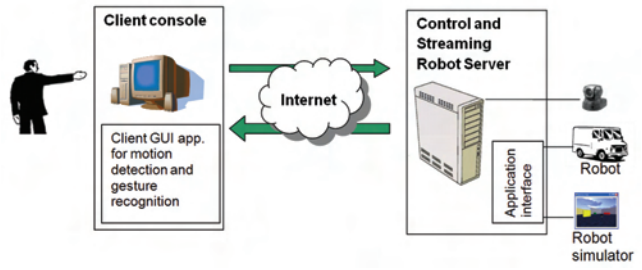


Fig. 3. System architecture

In the case of no match, the system outputs "No gesture detected" message, with suggestions for the possible operator actions, which correspond to commands applicable in the current robot state. The application context restricts the possible visual interpretations and the user exploits the immediate feedback to adjust their gestures.

Robot server can be connected through the unified application interface both to the real robot and the robot simulator. A web cam with a streaming server provides a real time view (feedback) of the robot movements. Captured video is sent to the client using the FTP protocol. Besides the video feedback, a graphic and audio message for the current robot action/state, received by the robot sensors, is communicated via TCP/IP to the client application, and is presented to the user. When the robot completes an action, it is also reported to the operator. General control algorithm consists of three phases: a) Image acquisition; b) Image segmentation; c) Feature extraction and gesture recognition.

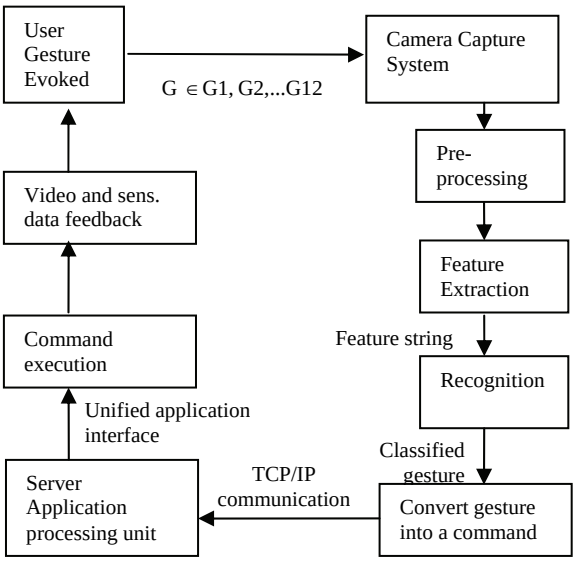


Fig. 4. Gesture recognition control flow diagram

V. MOTION DETECTION AND OBJECT EXTRACTION

In this research motion detection was used for locating a potential region of interest (ROI). Motion detection approach applied in our algorithm is not bound to any specific video stream format/protocol. Instead of this, it just analyze consequent video frames, which makes them free from any video processing routines and makes them applicable to any video stream format.

This approach, based on finding inter-frame differences, tries to use simple techniques of modeling scene's background and updating it through time to get into account scene's changes. To take these changes into account, we are going to apply adaptive background modeling approach which tends to decrease the differences between two given images. The reference image is constantly updated with newly arriving image using the Eq.1:

$$\forall x, \forall y \; R_t(x,y) = \frac{N-1}{N} \cdot R_{t-1}(x,y) + \frac{1}{N} I(x,y) \tag{1}$$

With R standing for the reference image and I for the newly arrived frame. The formula calculates a running average over all frames, with a weighting that decreases exponentially over time. For setups in front of a wall or white board, we found that the user practically never rests with his hand in the same position for more than 10 seconds, which implies a value of about 500 for N.

The main problem with this type of reference image updating is that dark, non-moving objects, such as the body are added to the reference image. If the user moves his/her hand into regions with dark objects in the reference image, there is not sufficient contrast between foreground and background to detect a difference. For this reason, we can update dark regions slowly, but light regions instantly. The value N in Eq. 1 is calculated as in Eq. 2.

$$\forall x, \forall y \; N(x,y) = \begin{cases} 1 \; for \; I_{t-1}(x,y) - I_t(x,y) \le 0 \\ \approx 500 \; for \; I_{t-1}(x,y) - I_t(x,y) > 0 \end{cases} \tag{2}$$

With this modification, the algorithm provides the maximum contrast between foreground and background at any time.

This kind of modeling approach gives the ability of more precise highlighting of motion regions and makes the

Fig. 5. Image differencing with reference image. From left to right: reference image, input image, output image

algorithm adaptive to unconstrained dynamic scenes.

By applying a differentiation operation on the current and reference image, it is possible to see absolute difference between the two images (Fig.5).

VI. HANDS GESTURE RECOGNITION

The algorithm for hand gesture recognition takes into consideration human's body proportions. It is based on histogram analysis, and presents two advantages: implementation simplicity and execution quickness. The core idea is based on analyzing two kinds of object's histograms: horizontal and vertical histograms (Fig. 6).

Fig. 6. Horizontal and vertical histogram of the detected human's body

Analyzing the horizontal histogram, one can observe that the hands' areas have relatively small values on the histogram, and the torso area is represented by a peak of high values. Considering relative proportions of humans' body and the horizontal histogram, we may say that human hand's thickness can never exceed 30% percent of human's body height. So, hands' areas and torso area can be easily classified, their length can be determined. Both experimentally and analytically, it can be proven that in the case of not raised hand, in most cases, the width on horizontal histogram will not exceed 30%. Otherwise it can be considered that the hand is raised somehow. When the hand is raised it can be raised in three different positions: raised diagonally down, raised straight or raised diagonally up. In order to determine the exact hand's position when it is raised, the vertical histogram of the hands objects only (Fig. 7), will be analyzed.

Fig. 7. Vertical filtered histograms of the raised hand

In some cases the histograms may be cluttered with noise, which may be caused by light conditions and shadows. Therefore two additional preprocessing steps of filtering will be performed, on the vertical histogram:

- removing the low values from the histogram, which are lower than some threshold percentage value of maximum histogram's value. This will remove some artifacts caused by the shadows.
- removing all peaks, which are not the highest peak. This will remove the strong shadows and other artifacts caused by the environment changing conditions. After these preprocessing steps exact hand position will be determined, taking into consideration one more assumption about body

proportions, length of the hand is much bigger than its width. Therefore, one may observe that in the case of straight raised hand its histogram should have quite high, but thin peak, for the diagonally up hand its shifted to the top of vertical histogram, but the peak of the diagonally down hand is shifted more to the center (Fig. 7).

VII. Experimental results

In order to evaluate the performances of the developed control algorithm, physics-based robot simulator and server applications are deployed on PC with AMD Athlon 64 Processor on 2.4GHz with 1GB of RAM, NVIDIA GeForce FX5500 with 256MB memory and Windows XP operating system. Gesture recognition client application is implemented in C# language and is deployed on HP Pavilion dv62000 with Intel Core 2 Duo Processor on 2.2GHz with 2GB RAM, it has built in WebCam with 1.3 Megapixels resolution. Both computers are connected on (8 Mbps) Internet connection.

The performances of the developed control method were tested in two different conditions:

1. with sample images grabbed from static scenes in which the lighting and the distances from the camera were approximately constant.
2. in unconstrained scenes with changes of lighting, moving objects and different distances from the camera.

The accuracy was measured for 18 untrained users (9 males, 9 females with different body sizes) in both cases. Each user tried all command gestures 95 times in front of the camera watching the monitor. Table I shows the rates of successfully performed actions by the robot simulator. The results are average of all users.

TABLE I
HAND GESTURE DETECTION RATES

Gesture	Static scene detection rate	Unconstrained scene detection rate
Open	93.2%	89.7%
Close	91.5%	90.4%
Move up	94.9%	91.8%
Move down	95.1%	92.2%
Tilt -4°	94.6%	91.3%

VIII. Conclusion

This paper presents a fast, robust and accurate method for hand gestures recognition under unconstrained scenes. Experimental results show satisfactory recognition percentage of the gestures. The failure of the system to recognize the gesture is mainly due to the very changeable lighting conditions and moving objects (persons) entering the scene, operator's failure to move the hand to the proper posture. It must be emphasize that after a short experience operators get used to the system.

References

[1] E. Stergiopoulou and N. Papamarkos, "A new technique for hand gesture recognition", Proceedings of IEEE International conference on Image Processing, Atlanta, pp. 2657-2660, Oct. 8-11, 2006

[2] Malassiotis, S. and Strintzis, M. G., "Real-time hand posture recognition using range data," Image Vision Comput. 26(7), 1027-1037 (2008).

[3] Hebert Luchetti Ribeiro, Adilson Gonzaga, "Hand Image Segmentation in Video Sequence by GMM: a comparative analysis", Symposium on Computer Graphics and Image Processing (SIBGRAPI'06), 2006

[4] R. Mariappan, "Video Gesture Recognition System," iccima, vol. 3, pp.519-521, International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), 2007.

[5] Popa, D., Simion, G., Gui, V., and Otesteanu, M. 2008. Real time trajectory based hand gesture recognition. WSEAS Trans. Info. Sci. and App. 5, 4 (Apr. 2008), 532-546..

[6] Bugeau, A. and Pérez, P. 2009. Detection and segmentation of moving objects in complex scenes. Comput. Vis. Image Underst. 113, 4 (Apr. 2009), 459-476.

[7] Nielsen, J. "Enhancing the explanatory power of usability heuristics." In Proc. of the CHI 94 Conference on Human Factors in Computing Systems, Boston, MA, ACM.

Exploration of Algorithms to Control a Maze Robot

N. Nikolov[1], S. Ivanov[2], S. Mustafa[3], M. Todorova[4]

Abstract **- The present paper explores different techniques to control a robot designed on the basis of an embedded microcontroller. A number of robot control algorithms have been applied to the hardware. These robots are used for teaching purposes in the hardware division of "Computer Systems and Technologies". A mobile object would always heighten the students' interest and improve the quality of the teaching process.**

Key words **- sensor, embedded microcontroller, robot, control system**

I. INTRODUCTION

The studies of subjects such as Embedded Microcontrollers, Design of Microprocessor Systems (MPS), System Analysis etc. being a part of the Bachelor's and Master's Degree in "Computer Systems and Technologies" require specific teaching facilities. The teaching process is provided by scale models but the objects to be studied are static. Technical universities similar to ours employ MPS to control mobile objects (robots). Their application in the teaching process enhances the degree of visualization and understanding and enables the students to learn the subject matter being taught faster and more effectively.

II. DESCRIPTION OF HARDWARE

The task pursued by our team is to develop a MPS which should control a mobile platform by means of two reversible DC motors and their reduction gear. Depending on the input data, various tasks for robot control can be defined. On the basis of the information coming from distance sensors, reflective sensors etc., the computer program should control the robot movement and make it move in the following ways:

- on a line route (line following robot);
- in a maze (maze robot);
- in a maze trying to detect an area having the highest temperature etc.

MPS is so versatile that the tasks to be solved can be expanded. An important advantage is the presence of different variants for communication with other objects using standard wired and wireless interfaces.

The block diagram of the system is shown in fig.1:

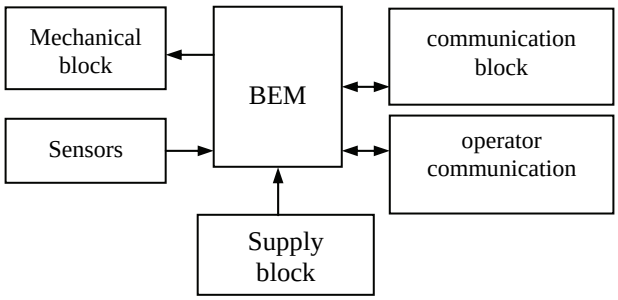


Fig.1. Block diagram

The system consists of the following blocks:

- mechanical block – it consists of a platform, two motors, two driving wheels. The mobile car has a platform moved by two wheels. For stability purposes there is a third support point which is a foot with a ball end. The axis of each wheel is driven by a reduction gear which is driven by a reversible DC motor. (fig.2)

Fig.2. Two DC motor with reduction gear

- sensor block – sensors of both analogue and digital (ON/ OFF) type can be connected to the system. These sensors give information about the distance from the object to a wall, the colour (dark or light) of the surface on which it is moving, the ambient temperature etc. Fig.3 shows a distance sensor and a surface colour sensor (reflective type).

[1]Nedyalko N. Nikolov, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: ned.nikolov@tu-varna.bg.

[2]Sava I. Ivanov, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: ivanovsi@abv.bg.

[3]Suzan Ch. Mustafa, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: suzan.chetinova@abv.bg,

[4]Maya P. Todorova, Technical University – Varna, Computer Sciences Departament, 9010 Varna, Bulgaria, E-mail: mayasvilen@abv.bg

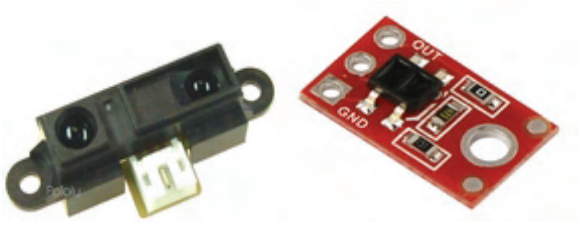


Fig 3. Distance and surface colour sensors

- block for operator communication – it consists of a LCD alphanumeric display to show parameters as well as system status and a matrix keyboard with a 3 by 4-button structure to input parameters and to start the system on different modes. The buttons are tested dynamically by activating one column through ports RB1 to 3 and the button status is read through inputs RB4 to 7. The change in the status of each port can cause "set interrupt flag" and "interrupt request".

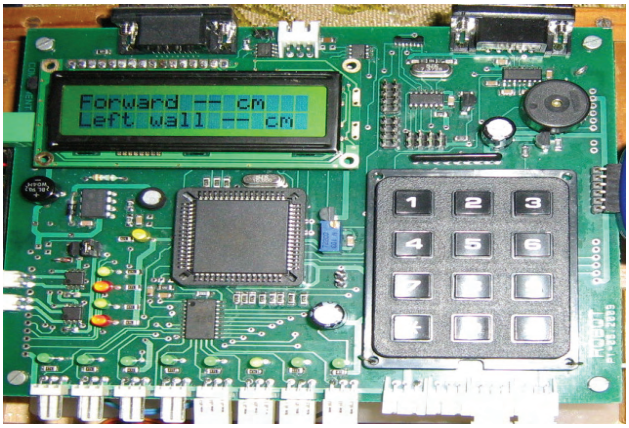
The LCD display has a structure of 2 by 16 characters and is controlled by Port D. A half-byte mode of exchange has been selected in order to economize on ports of the embedded microprocessor. It is a matter of programming what alphanumerical information will be displayed.

- communication block– it enables the system to establish communication with other systems by one of the following standard interfaces: RS 232, IrDA, CANbus, LINbus and radio interface.

- block "embedded microcontroller" (BEM) – it serves to process the information coming from the sensors, to compute the control signals to the motors according to the application program loaded in the microcontroller, to monitor the status of the keyboard.

When deciding which one to choose, we had in mind the following considerations:

1. It should enable direct control of all blocks included in the system;
2. It should be of a series familiar to the students; in this way, familiar tools such as MPLAB and C18 can be employed to develop, load and adjust the application programs;
3. The modules built-in in the embedded microcontroller (timers, PWM, ADC, UART, SSP, CAN etc) can facilitate the development of application programs;
4. A PLCC should be installed in a PLCC socket which enables replacement of the embedded microcontroller in case of damage.



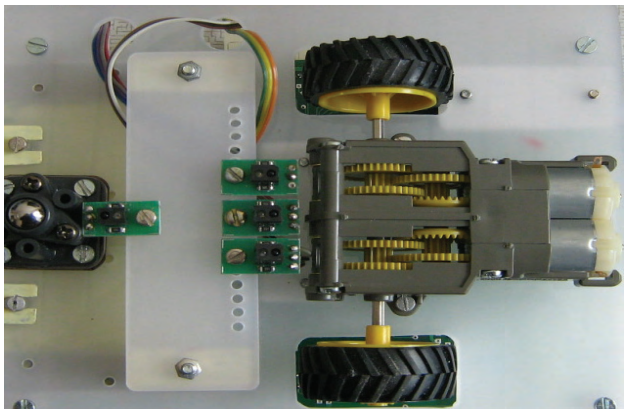


Fig. 4. General view of the robot system

- Motor control block

These two are absolutely one and the same designed and constructed by means of a specialized chip BA6287F i.e. driver for control of DC motor. The chip enables the motor to operate in one of the following modes: forward movement, reverse, stop, and stand by. By means of two pins the embedded microcontroller (EMC) controls the direction and the speed of the motor. The speed change is achieved by PWM signal generated by the PWM1 module or PWM2 of the embedded microcontroller. The duty cycle is programmable and it is proportional to the speed of the motor.

- Block for processing analogue sensor information

It processes information coming from a sensor of the GP2Y0A21KOF type by means of which distances from 10 to 80 cm can be measured. There are 4 analogue inputs (RA0,1,2,3) from which the information goes into the analogue multiplexer and 10-bit analogue digital converter (ADC) built-in in the embedded microcontroller. The sensor power supply is +5V. Other types of sensors can be connected as well provided that their output analogue signal changes from 0 to +5V. Also, there is SMT160 sensor in the system which measures the ambient temperature. Its output is a PWM signal whose duty cycle is a function of the temperature measured. This signal is transmitted to the input RA4 and the value of the duty cycle can be calculated with the help of a program.

- Block for processing digital sensor information

Optical reflective sensors of the RPR220 type can be used in the system and they register the light reflected from the surface at a distance of 8mm. With their help black and white straps can be registered. The condition of each input signal from the sensor is indicated by a LED, it is buffered and transmitted to the embedded microcontroller input. There are 8 inputs of this kind. Other sensors with a power

supply of +5V and generating ON/OFF output signal can be connected to them as well.

These four blocks are sufficient in order that programs for robot control can be developed, loaded and adjusted. The remaining blocks serve to expand the possibilities of the system by providing a connection with the operator and other micro-processing systems by a standard interface.

Programming and program testing can be accomplished with the help of a programmer i.e. debugger MPLAB IDC2 Lite.

III. ALGORITHMS FOR CONTROL OF A ROBOT MOVING IN A MAZE

The analogue and digital inputs in the system interface present a wide variety of tasks to be solved involving movement and control of the robot. The present paper offers a number of algorithms to solve the following task: using one or more analogue distance sensors, both an algorithm and a program to be developed in order to control a robot moving in a maze.

In order to perform the task two analogue sensors for distance are used – they are mounted on the mechanical part and the board with the embedded microcontroller. The distance range of the sensors used is [10, 60] cm; This means that their location on the construction should always be at a distance not less than 10 cm as compared to the distance to be measured. As a result of that the sensors are mounted approximately in the middle of the platform near to the axis connecting both motors.

Fig.5 shows a diagram of the maze and a photo of the robot in the maze:

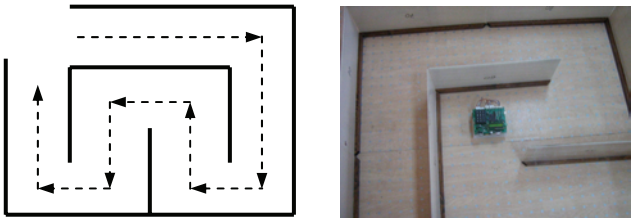


Fig. 5. A maze and an ideal trajectory

The walls have a width of 40 cm, the overall dimensions of the robot are 20 x 12 cm.

The algorithm for control of a robot in a maze having parallel and perpendicular walls is relatively simple. It consists of measuring the distance from the robot to an opposite wall. If this distance is bigger than 20 cm, the robot keeps moving forward. Otherwise, the distances to the right and the left walls are being measured (90 degrees on the left and 90 degrees on the right of the trajectory of the movement). A decision is taken to turn left or to turn right depending on the fact which distance is bigger.

The trials under real conditions have shown some disturbances. The most common of them are as follows:

- The distance sensors require a graphic representation of the equation $S = f(V)$, where S is the distance, and V is voltage [0,5] to be presented in a tabular form.. A testing program has been developed in this system in order to improve the degree of accuracy in the table containing the distance values as a function of the voltage. The testing program visualizes the sensor readings for the distance to a wall when the motor speed is zero. By means of a digital voltmeter, the voltage at the sensor outputs is measured. The distances are fixed with an increment of 5 cm i.e. (5, 10, 15, etc.) the values of the distances in the table and the fixed ones are compared. If necessary, the voltage values for the fixed distances are corrected.

- Although theoretically both motors and their reduction gear are equal, in real conditions they differ which results in the speed of each wheel which can also differ;
- The smoothness and friction coefficient of the surface on which each wheel is moving are not always one and the same;
- The 90-degree left and right turns and the U-turn of 180 degrees are not performed precisely;

These disturbances cause deviations from the ideal trajectory in the straight sections as well as in the turns This necessitates corrections to be made in the control algorithm in order to position the robot as close to the ideal trajectory as possible. When the robot does not deviate from the ideal trajectory, the above rules are sufficient for performing the task.

Before performing the real task i.e. a robot moving in a maze, tests should be made for movement in a straight line and parameter adjustment for taking the three types of turns (left, right and U-turn).

The tests for straight line movement require a smooth track to be selected. Different speed values are set for movement of the robot in forward and reverse direction. If there is no significant deviation (more than 2-3%) from the straight line over the distance of 200 cm, then no corrections are required. Otherwise, the speed of one of the motors should be increased or decreased. The tests under real conditions have shown that the speed settings for both wheels in forward and reverse direction are different. After completion of the tests for straight line movement, the parameters for turn taking should be adjusted. Each turn can be performed in one of the following ways (e.g. for a right turn):

- The speed of the right motor is zero, and the speed of the left motor remains the same in value and direction. In this case, the turn radius equals the distance between both wheels, and the right wheel is the circle centre of the turn;
- The speed of the left motor is zero, and the speed of the right one is reversed. In this case, the turn radius equals the distance between both wheels, and the left wheel is the circle centre of the turn;
- The speed of the right motor is reversed, and the speed

of the left one remains the same in value and direction. In this case the turn radius equals half the distance between both wheels, and the middle of the axis between both wheels is the circle centre of the turn.

A testing program has been developed for adjustment of the parameters when the robot takes 90-degree left or right turns. The program makes the robot move in a square i.e. a left or right turn is made 4 times. If the square is repeated several times (2 or 3) without any substantial deviations, then the parameters should be stored.

After making an analysis of the causes leading to deviations from the ideal trajectory, the following conclusion has been drawn: Some additions to the basic algorithm for movement in a maze should be made so that a left or right wall could be followed. If such a result could be reached, it is not of any importance whether the geometrical shape of the maze is rectangular or not.

Fig.6 shows the two variants for a deviation from the ideal trajectory:

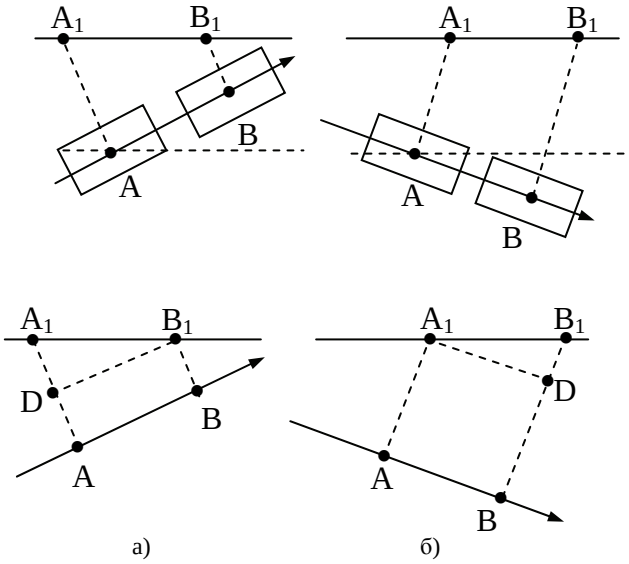


Fig 6. Types of deviations

Let us assume that Δ is the relative deviation resulting from two consecutive measurements of the distance to the left wall:

$$\Delta = \frac{A_1D}{AA_1} = \frac{AA_1 - BB_1}{AA_1} \quad \text{or} \quad \Delta = \frac{B_1D}{BB_1} = \frac{BB_1 - AA_1}{BB_1} \qquad (1)$$

When Δ exceeds the pre-set value (0.2-0.3%), commands “easy to left” and ‘easy to right” should be included in the algorithm. These commands can be executed by increasing or decreasing the speed of one of the two motors. For example, in order to correct the movement from fig. 6[а] and make it parallel to the wall, the speed of the right motor should be decreased for a short period of time, after that the original value of the speed should be restored. Likewise, these steps can be taken for fig. 6[б] but they relate to the speed of the left motor.

If the distance BB_1 becomes less than the pre-set minimum distance to the left wall, then a manoeuvre for clearing the wall should be made. Fig.7 shows a variant for making such a manoeuvre.

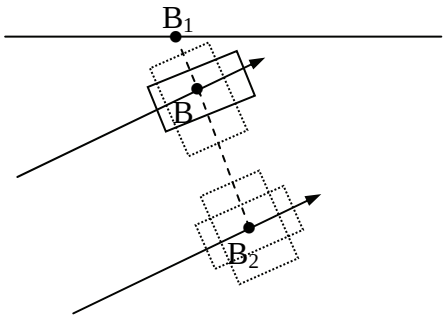


Fig.7. Manoeuvre for clearing the left wall

Clearing is executed by a 90-degree right turn, 5-cm movement in a straight line, a 90-degree left turn and then the original direction is resumed. Likewise, a manouevre for approaching the left wall is made, when the clearing distance exceeds the maximum pre-set value. In this case, the sequence of manoeuvres is as follows: a 90-degree left turn, movement forward and a 90-degree right turn.

These corrections in the algorithms result in achieving the near ideal trajectory and when the robot is heading for an opposite wall, the position of the robot is almost perpendicular.

IV. Conclusions

The manoeuvres “easy to left” and “easy to right” which have been executed as an addition to the basic algorithm have resulted in a considerable improvement of the quality of the robot control.

The robot is being used for performance of some other tasks as well. It has found an application in the laboratory practice of several subjects taught in the university.

References

[1]. R. Dofr, R. Bishop, Modern Control System, Peason Education, 2008.

[2]. Н. Кенаров *PIC микроконтролери* част 2- Варна 2006.

[3]. www.microchip.com – PIC18F6585 Data Sheet – 68 pin High performance 64 KB Enhanced Flash Microcontroller with ECAN module – 2004

[4]. www.pololu.com

Path Control with Input Saturation for Reversing a Car-Like Mobile Robot

Plamen P. Petrov[1] and Lubomir M. Dimitrov[2]

***Abstract* – This paper proposes a path control with input saturation for backward driving of a car-like mobile robot. The control law is constructed using high-gain design techniques and involves error coordinates and invariant properties with respect to the vehicle speed. The stability of the closed-loop system is analyzed using Lyapunov stability theory. Simulation results illustrate the effectiveness of the proposed controller.**

***Keywords* – Car-like mobile robot, Path following, Backward driving, High-gain control, Input saturation**

I. INTRODUCTION

In this paper, we consider the path following problem for backward driving of a car-like mobile robot with input saturation. In recent years, significant advances have been made in designing feedback controllers for nonholonomic mobile robots. While there has been significant amount of work on controlling the motion of mobile robots without bound of the control inputs [1, 2, 3], there has been much less work on mobile robot motion control with input saturations [4].

In this paper, we propose a path following controller with input saturation for a car-like mobile robot during backward driving. The paper is organised as follows: In Section II, the kinematic model of the car-like mobile robot is presented and the control problem is formulated. In Section III, the control law is designed. Simulation results are presented in Section IV. Section V concludes the paper.

II. PROBLEM FORMULATION

A. Mobile Robot Kinematic Model

A plan view of the mobile robot is given in Fig. 1. The mobile robot has three non-deformable wheels. The wheels are assumed to roll on a horizontal plane without slipping. The longitudinal base *PS* of the vehicle is denoted by *l*. To describe the position and orientation of the robot in the plane, we assign the following coordinate frames: Px_Py_P located at the center of the rear wheel axle and stationary with respect to the robot body where the x_P axis is along the longitudinal base of the robot, and an inertial coordinate frame Fxy in the plane of motion. The coordinates of a reference point *P* placed at the center of the rear robot axle, with respect to *Fxy*, are denoted by (x_P, y_P). The angle θ is the orientation angle of the robot with respect to the frame *Fxy*. The angle α, is the front wheel steering angle. The steering angle is measured with respect to the robot body.

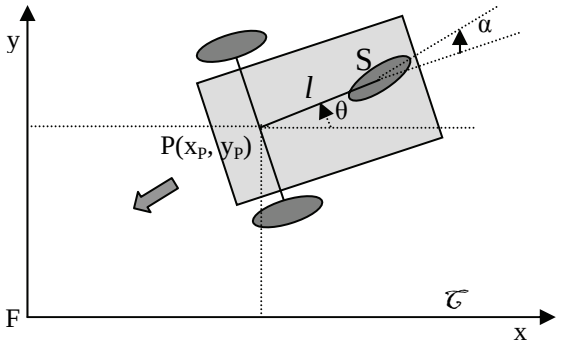


Fig. 1. A plan view of the car-like mobile robot

If the rotation of the wheels with respect to their proper axes is ignored, the mobile robot configuration can be described by four generalized coordinates $q = [x_P, y_P, \theta, \alpha]^T \in \Re^4$. The kinematic model of the robot in the plane can be described by the following system of nonlinear differential equations

$$\begin{aligned} \dot{x}_P &= v_P \cos\theta \\ \dot{y}_P &= v_P \sin\theta \\ \dot{\theta} &= \frac{v_P}{l}\tan\alpha \end{aligned} \tag{1}$$

where v_p is the velocity of point *P*. The front-wheel steering angle α is the control input of the system and is a subject of physical limitations $|\alpha| \le \alpha_{\max}$.

B. Control Problem

The path following geometry used in this paper is represented in Fig. 1. Consider a car-like mobile robot moving backward on a flat surface. We assume that the path $\mathcal{C}$ is a straight line which for simplicity coincides with the *Fx* axis of the inertial frame *Fxy*. We assume that v_P together with its derivative are bounded and also, that the following inequalities hold: $0 < v_{Pmin} \le |v_P(t)| \le v_{Pmax}$. In this case, using the parameterization (y_P, θ) and given a path $\mathcal{C}$, the path

[1]Plamen P. Petrov is with the Faculty of Mechanical Engineering, Technical University of Sofia, 8, Kl. Ohridski str., 1797 Sofia, Bulgaria, E-mail: ppetrov@tu-sofia.bg

[2]Lubomir M. Dimitrov is with the Faculty of Mechanical Engineering, Technical University of Sofia, 8, Kl. Ohridski str., 1797 Sofia, Bulgaria, E-mail: lubomir_dimitrov@tu-sofia.bg

following problem consists of finding a feedback control law for the system which consists of the second and third equation of (1) system with control input α subject to physical limitations $|\alpha| \le \alpha_{\max}$, such that the state vector $[y_P, \theta]^T$ tends to $[0, 0]^T$, as $t \to \infty$.

III. FEEDBACK CONTROL DESIGN

In this Section, we present a path following controller for backward driving of the robot given by the second and third equations of (1) with input saturation for the steering angle α. The control objective is to regulate the state vector $[y, \theta]^T$ to zero. Since $v_P(t)$ is assumed to be strictly negative ($v_P(t) = -|v_P(t)| < 0$), to obtain a time-invariant system, the differentiation with respect to time is replaced by differentiation with respect to s ($ds = v_P dt$), where s is the real path length drown by the reference point P. In that way, we express the vehicle's equations of motion in terms of s and denote the derivation with respect to s by " ′ ". Next, we proceed with the following change of input

$$u = \frac{\tan\alpha}{l} \tag{2}$$

In that way, the system can be written in the form

$$\begin{aligned} y_P^{'} &= -\sin\theta \\ \theta^{'} &= -u \end{aligned}. \tag{3}$$

For the first equation of (3), the control

$$\xi(y_P) = \theta = y_P \tag{4}$$

achieves local exponential stability of $y_P = 0$. For the (y_P, θ) system (3), we propose the following feedback control

$$u = Lsat\left(\frac{\lambda}{L}\right) \tag{5}$$

where

$$\lambda = ka(\theta - \xi), \quad k = cte > 0, \quad a = cte > 0 \tag{6}$$

and $sat(.)$ is a saturation function

$$|u(t)| \le L, \qquad L = cte > 0$$

$$sat\left(\frac{\lambda}{L}\right) = \begin{cases} -1 & for \quad \frac{\lambda}{L} < -1 \\ \frac{\lambda}{L} & for \quad \left|\frac{\lambda}{L}\right| \le 1 \\ 1 & for \quad \frac{\lambda}{L} > 1 \end{cases}. \tag{7}$$

The control (6) achieves semi-global stabilization of (3). The semi-global stabilization [5] means that for any compact neighborhood Γ of (y_P, θ) = (0, 0), there exists k^* such that for all $k \ge k^*$, the region of attraction contains Γ. Indeed, we introduce the following change of coordinates

$$\eta = \theta - \xi \ . \tag{8}$$

The system (3) can be rewritten in the new coordinates as

$$\begin{aligned} y_P^{'} &= -\sin(y_P + \eta) \\ \eta^{'} &= -ka\eta - Lsat\left(\frac{\lambda}{L}\right) + \lambda + \sin(\eta + y_P) \end{aligned}. \tag{9}$$

For the system (9) we consider the following Lyapunov function

$$V = \frac{1}{2}y_P^2 + \frac{1}{2}\eta^2. \tag{10}$$

Using (9), for the derivative of (10), we obtain

$$\begin{aligned} V' = &-y_P^2\frac{\sin(\eta + y_P)}{\eta + y_P} - ka\eta^2 - y_P\eta\frac{\sin(\eta + y_P)}{\eta + y_P} \\ &+ \eta\left[\lambda - Lsat\left(\frac{\lambda}{L}\right) + \sin(\eta + y_P)\right] \end{aligned}. \tag{11}$$

Denoting

$$\sigma = \lambda - Lsat\left(\frac{\lambda}{L}\right), \tag{12}$$

and using the fact that

$$\forall \ |\lambda| \le L(1+\delta) \ \rightarrow \ |\sigma(\lambda)| \le \frac{\delta}{1+\delta}|\lambda| \tag{13}$$

we have

$$V' \le -y_P^2\frac{\sin(\eta + y_P)}{\eta + y_P} - \eta^2\left[ka - \frac{\delta}{1+\delta}ka - \frac{\sin(\eta + y_P)}{\eta + y_P}\right]. \tag{14}$$

Since

$$0 < \frac{\sin(\eta + y_P)}{\eta + y_P} \le 1 \tag{15}$$

by choosing

$$k \ge k^* = \frac{1+\delta}{a} \tag{16}$$

we have $V' < 0$.

IV. SIMULATION RESULTS

Simulation results using MATLAB were performed to illustrate the effectiveness of the proposed controller. A straight line reference path which coincides with the Fx axis of an inertial frame was chosen for the simulations. The longitudinal base of the robot was *1m*, and velocity of the mobile robot was chosen to be *1m/s*. The bound for the front wheel steering angle was $|\alpha| \leq \alpha_{max} = 0.785rad$ (*45 deg*). The controller gains were $k = 1$, $a = 1$. Initial conditions was chosen to be $y_P(0) = 1.5m$; $\theta(0) = -0.5rad$, $\alpha(0) = 0rad$. Evolution of the state coordinates $[y_P, \theta]^T$ with respect to the variable $s = \int_0^t |v_P| \, d\tau$ is depicted in Fig. 2.

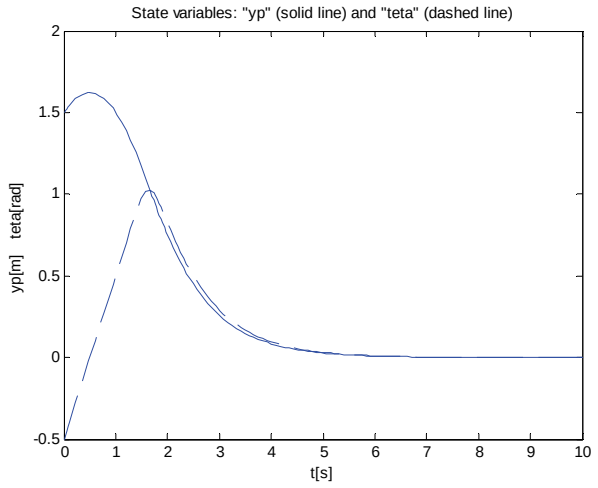


Fig. 2. Evolution in time of the state coordinates (y_P, θ)

Evolution of the front wheel steering angle α is depicted in Fig. 3.

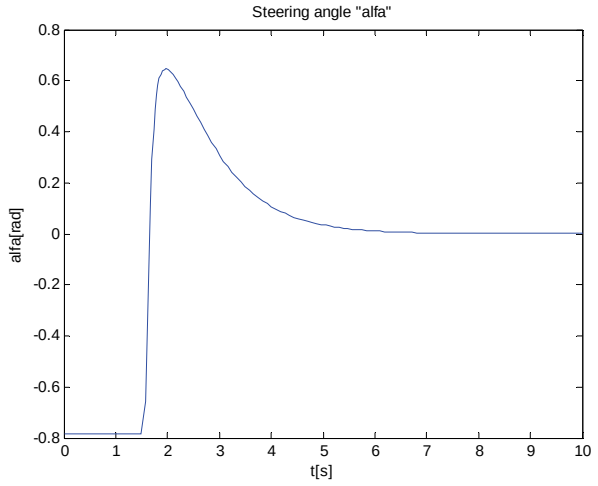


Fig. 3. Evolution in time of the front wheel steering angle α

The simulation results confirm the effectiveness of the proposed controller.

V. CONCLUSION

In this paper, a path control with input saturation for backward driving of a car-like mobile robot is proposed. The control law is constructed using high-gain design techniques and involves error coordinates and invariant properties with respect to the vehicle speed. The stability of the closed-loop system is analyzed using Lyapunov stability theory. Simulation results illustrate the effectiveness of the proposed controller. Our future work will consist in dynamic extension of the proposed controller in the presence of uncertainties in the model of the mobile robot.

ACKNOWLEDGEMENT

This work was supported by National Ministry of Science and Education of Bulgaria under contract BY-I-302/2007: "Audio-video information and communication system for active surveillance cooperating with a mobile security robot".

REFERENCES

[1] F. Diaz del Rio, G. Jimenez, J. Sevillano, S. Vicente, A. Balcells, A path following control for unicycle robots, in *Journal of Robotic Systems*, 18 (7), pp. 325-342, 2001.

[2] P. Petrov, M. Tetreault, J. de Lafontaine, Path control during backward driving of a tractor-trailer with off-axle hitching, in *Proceedings of the International Conference on Advanced Robotics*, pp. 571- 576, 1999.

[3] P. Petrov, L. Dimitrov, Nonlinear path control for differential drive mobile robot, *Recent Journal*, vol. 11, no. 1, pp. 41-45, 2010.

[4] Z.-P. Jiang, E. Lefeber, H. Nijmeijer, Saturated stabilization and tracking of a nonholonomic mobile robot, *Systems&Control Letters*, 42, pp.327-332, 2001.

[5] R. Sepulchre, M. Jankovic, P. Kokotovic, *Constructive nonlinear control*, Springer-Verlag London, 1997.

This page intentionally left blank.

POSTER SESSION PO VI

PO VI - Power Transmission and Distribution Systems

Simulation Models of the Automatic Transfer Switch in the Electrical Power Distribution Network

Mediha Mehmed-Hamza

Abstract – **The article presents simulations of distribution electrical power network, relay protection and automatic transfer switch, used in electrical networks 20 kV. For the development of simulation models standard libraries in software are used as well the own models. The developed simulation model allows to: visualize the research network's parameters for different modes of operation; to study of the automatic transfer switch; increase the quality learning by the trainees and enhance their self-employment.**

Keywords **– automatic transfer switch, power distribution network**

I. SIMULATION MODEL DESCRIPTION

The simulation model of electrical network 20 kV with 100% back- up is under consideration. The model scheme is shown in Fig. 1. The blocks of the scheme are: power systems 110 kV (S and S1), a main power transformer 110/20 kV (PT1), back-up power transformer 110/20 kV (PT2), RT1 breakers (Q1 and Q3), breakers of RT2 (Q2 and Q4), bus 20 kV (L1, L2, L3), model power lines (W1 and W2), loads (B1 and B2), block modeling short circuit (Fault), block relay protection line W1 (RP), breaker of W1 (Q5), unit for automatic transfer switch (AKQ) and measurement units (U/I, U).[2,3]

The blocks AKQ and RP includes models developed by the author for automatic transfer switch and relay protection. For the other blocks of model scheme, presented in Figure 1, are used standard software libraries.

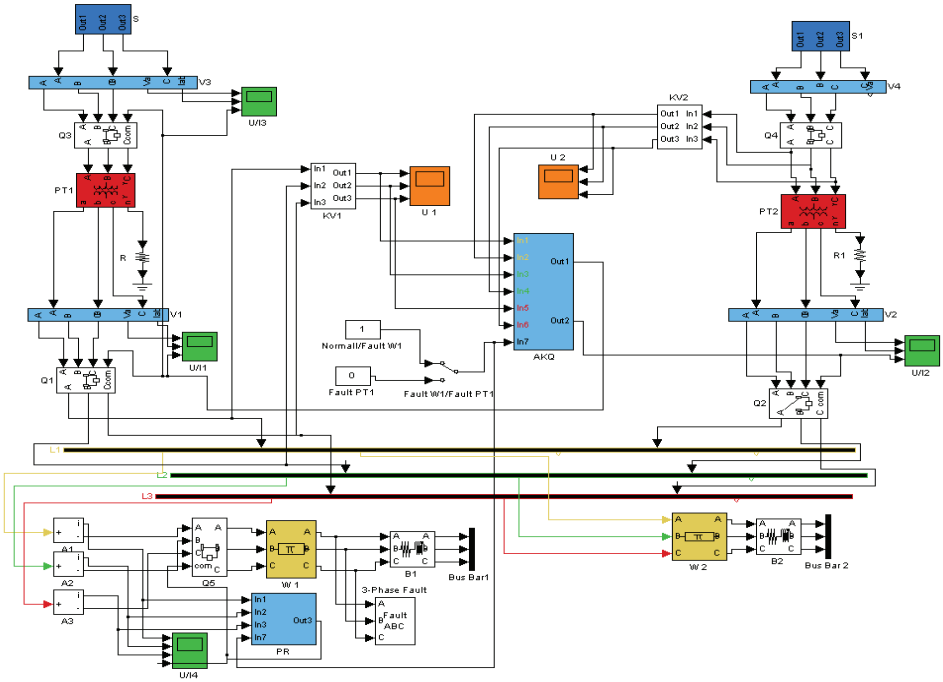


Fig.1. Simulation model scheme

[1]Assist Prof. PhD. Mediha Mehmed-Hamza – Department of "Electrical Engineering and Electrical Technologies", TU- Varna, e-mail: mediha.hamza@mail.bg

The blocks AKQ and RP includes models developed by the author for automatic transfer switch and relay protection. For the other blocks of model scheme, presented in Figure 1, are used standard software libraries.

Block AKQ description

The block AKQ performs all logic functions, which are put in back- up automation. The input variables for the block are:

- Three phase voltages (KV1), measured on the bus (L), which is reserved;
- Three phase voltages (KV2), measured on feeders of the back-up power transformer;
- A signal that the scheme operates in normal mode (supplied by PT1, but PT2 is idling) or external short circuit (on the power line W2) and accordingly a signal at the failure in PT1(Fault W1/Fault PT1).

Thus the functions of the undervoltage relay (KV1) and overcurrent relay (KV2) in automation are realized by the variables and the logic in block AKQ. With the same block the time delay of the automaton is realized, too (time relay).[1]

The output values from the block AKQ are:

- Out1 - signal for switching off Q1 and Q3 of the main power transformer PT1;
- Out2 - signal for switching on Q2 of the back-up power transformer PT2.

The input parameters for the blocks that are assigned are: primary tripping voltage of the starting relay undervoltage (KV1) in [V], primary tripping voltage of the overvoltage relay (KV2) in [V] and tripping time of the automation in [s]. Set tripping time of the automation is aggregated over tripping time of the terminal protection.

Figure 2 shows the menu for setting- up of the Block AKQ.

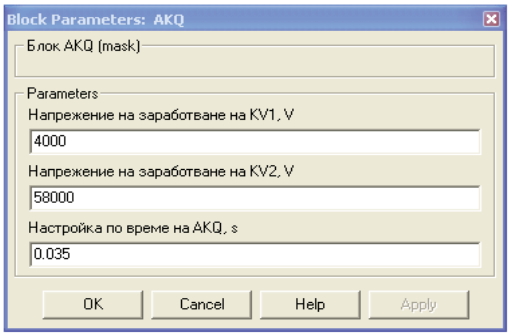


Fig.2. AKQ Menu for the Block AKQ set- up

Block RP description

Relay protection block perform the functions of overcurrent protection (OP), which is the most - often used protection for medium voltage networks as a protection against phase to phase fault.

Input parameters for the block that are assigned are: primary tripping current in [A] and the tripping time in [s]. At start-up the protection switches off the breaker Q5 of the power line W1.

Figure 3 shows the menu for setting- up of the Block RP

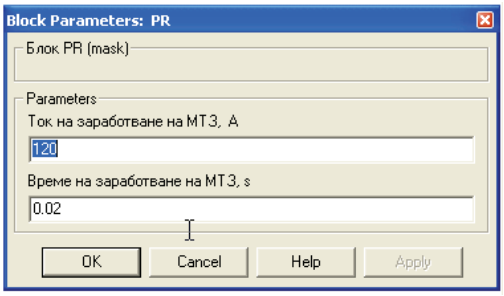


Fig.3. Menu for the Block PR set- up

The key "Fault W1/Fault PT1" chooses if the scheme operates in normal power supply of loads or under short circuit of the power line W1 (external short circuit) and hence the scheme works at failure over the main power source.

In blocks the phase voltages' and currents' alterations (U/I) and respectively phase voltages (U) are displayed.

II. THE PROCESSES' VISUALIZATION

The processes' visualization under three phase s.c. at the power line W1

Fig. 4 presents variations of phase voltages and currents under three phase s.c. at the end of the power line. The loads are supplied by the main power transformer of power line W1. In time t = 0.02 s a three-phase s.c. occurs. Breakers Q1, Q3, Q4 and Q5 are switched on and Q2 is off. Time delay of the protection of power line W1 is set up 0.02 s. and after this time the s.c. is switched off. Breaker Q5 switches to be opened. The automation of the back- up supply dos not trip, i.e. the supply of the loads remains from PT1.

The last graph, presented in Fig. 4, is a control signal of breaker Q1 (SQ1). It is 1, which means switched breaker. Time tripping set-up of the protection is less than those used in practice because the necessary computing time and the need visualization process.

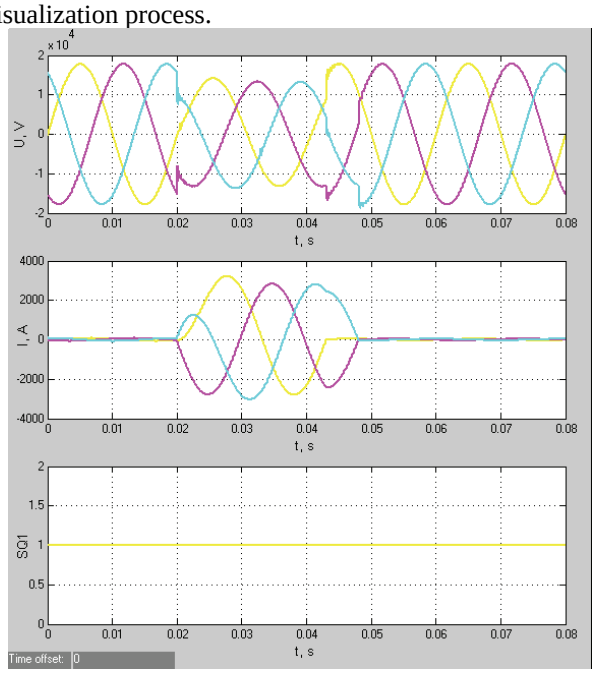


Fig. 4. Variations of phase voltages and currents under three phase s.c. at the end of the power line

The processes' visualization as a result of lowering the voltage due to fault in the main power transformer

Fig. 5 presents variations of phase voltages and currents at fault in the main power transformer. Values were measured before the breaker Q1. At the lowering of voltage of PT1 AKQ trips with set time 0.055 s and switches off breakers Q1 and Q3. The signal presented in the last graph of Fig. 5 is the control signal to the breaker Q1. It becomes from 1 to 0 which means switched off breaker Q1.

At the same time the signal 1 is given to switch on Q2 (Fig. 6). Consumers are supplied by PT2.

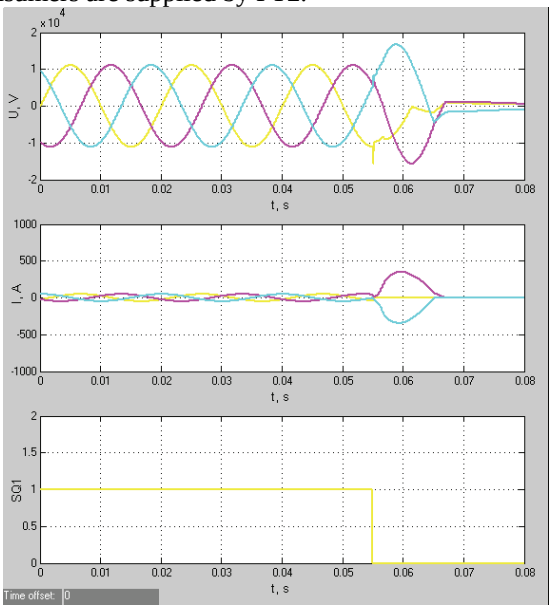


Fig. 5. Variations of phase voltages and currents, measured before Q1 at fault in PT1

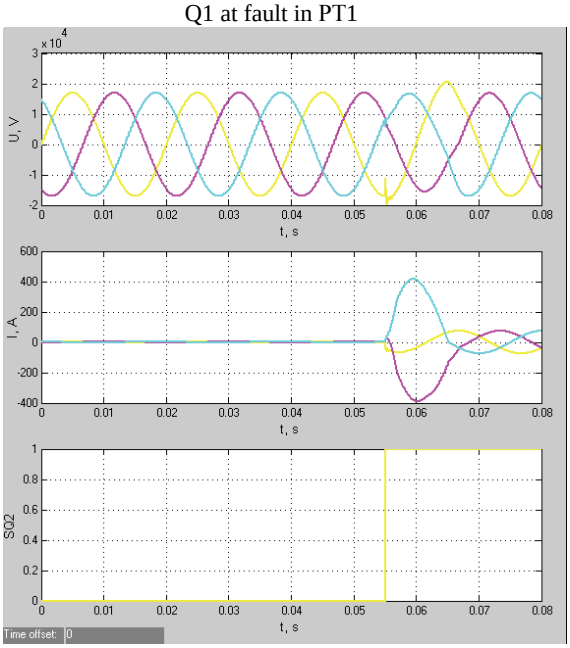


Fig. 6. Variations of phase voltages and currents, measured before Q1 at fault in PT1

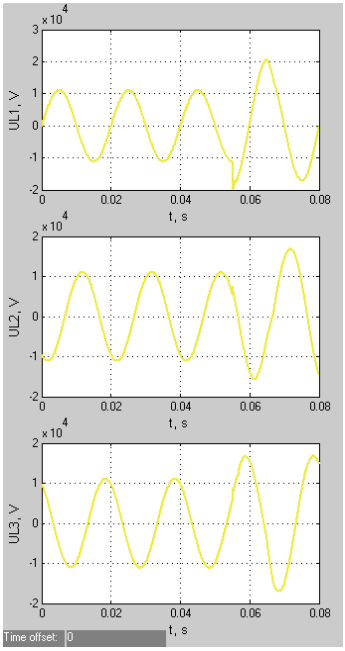


Fig. 6. Variations of phase voltages and currents, measured on the medium voltage bus at fault in PT1 and action of AKQ

After 0.055 s AKQ switches off the main power transformer and switches on breaker Q2 of the backup power transformer, which restores the normal power supply of 20 kV bus.

Conclusions:

Blocks "Automatic transfer switch " and "Relay Protection" have been developed., They model the operation of 100 percent back- up and overcurrent protection used in medium voltage electrical networks.

Students are free to model the above situations, to change relay protection settings and the back- up automat settings, to check whether at the specified settings the protection and AKQ will work, and to receive graphic data.

REFERENCES

[1] Andreev, St. Basic of automation of electrical power systems.

[2] Mediha Mehmed-Hamza, Margreta Vasileva. "Using of Matlab Simulink for Education in higt voltage technics and relay protection". ICEST 2009, Veliko Tarnovo, 2009.

[3] Mediha Hamza, Margreta Vasileva, Marinela Yordanova. "Co-ordination of the operation of the relay protection and surge protective devices in electrical power networks medium voltage 20kV".Journal of Electrical engineering, vol. 60,№ 3, 2009, 170-172.

This page intentionally left blank.

Assessment of the Safety Conditions from High Touch and Step Voltages in the Grounding Systems of Distribution Network

Nikolce Acevski[1], Kire Mijoski, Tomce Mijoski

Abstract: **In this paper are described and elaborated mathematic models which are applied for solution of problem with transfer of potential in the distribution systems. Appropriate algorithm is made as programmed package which is used for analysis or better says, for anticipation of the conduct of all grounding system (GS) in the area of the TS 110/20 kV "Gostivar" when faults to ground appeared in the system 110 kV**

Keywords: **groundings, transferred potentials, safety conditions, touch and step voltages**

I. INTRODUCTION

The distribution system (DS) in Gostivar is radial distribution network which is supplied by TS 110/20 kV/kV Gostivar. Most of the medium voltage (MV) from this distribution network presents cable network with isolated metallic shield (XHP, XHE) and a small part, with transmission lines without protective rope. So all objects in part of DS (TS 110/20 kV/kV, TS 20/0,4 kV/kV, metallic shield of MV cables, copper rope and FeZn tapes) are mutual galvanic harnessed and together with their groundings, they formed GS of these distribution network. Transmission lines without protective rope do not take part from the GS of this distribution network. At appearance of fault to ground at station 110/20 kV/kV of the DS or the network 110 kV in nearness of the DS, that power is distributed at all GS. The problem with transfer of potentials is expressed with the metallic shield of cable network, as and substation TS MV/LV by 20 kV network in DS, where groundings of that TS are sources of current field and dangerous voltage of touch and step.

II. GROUNDING SYSTEM - MODEL

Modelling of Cables 20 kV

Cables largely affect on features of GS in distribution network. They create connections between groundings of substations (with isolated metallic shield). Cables with conductive to ground metallic shield are behavior as groundings in GS. As a part of GS cables can be:

1. cables without metallic shield or cables with isolated metallic shield,
2. cables with conductive to ground metallic shield
3. cables placed with copper rope or Fe Zn tape.

Cable with length l placed between two substations 1 and 2 can be presented by distributed parameters as a chain of π four-extremities. Each of the four-extremities consists of a per unit length impedance $\underline{z} = r + jx$ calculated according to the Carson's relations [1], and two admittance's to ground per unit length $\underline{y}$. These admittances are nearly equal to the conductivity of a horizontal ground grid in shape of a tape with a equivalent diameter of the cable's shield d placed on a depth H with length l and they are calculated according to:

$$R_Z = \frac{\rho}{\pi \cdot l} \cdot \ln \frac{l}{\sqrt{h \cdot d_k}} (\Omega), \tag{1}$$

$$\frac{Y}{l} = \frac{1}{R_Z \cdot l} = \frac{\pi}{\rho} \cdot \frac{l}{\ln\left(l / \sqrt{h \cdot d_k}\right)}, \tag{2}$$

$$\underline{y} = \left(\frac{Y}{l}\right)_{l=1000m} = (g + jb), \tag{3}$$

$$b = 0; \quad g = \frac{\pi}{\rho} \cdot \frac{1000}{\ln\left(l / \sqrt{h \cdot d_k}\right)}, \left(\frac{S}{km}\right) \tag{4}$$

By applying the equation for lines with distributed parameters

$$\begin{aligned} \underline{U}_1 &= \underline{U}_2 \cdot ch\underline{\gamma}l + \underline{Z}_C \cdot \underline{I}_2 \cdot sh\underline{\gamma}l, \\ \underline{I}_1 &= \frac{\underline{U}_2}{\underline{Z}_C} \cdot sh\underline{\gamma}l + \underline{I}_2 \cdot ch\underline{\gamma}l. \end{aligned} \tag{5}$$

where the distribution constant $\underline{y}$ and the characteristic impedance of the shield of cable $\underline{Z}_C$ are calculated from:

$$\underline{\gamma} = \sqrt{\underline{z} \cdot \underline{y}} = \sqrt{(r + jx) \cdot g} = (\alpha + j\beta) \tag{6}$$

$$\underline{Z}_C = \sqrt{\frac{\underline{Z}}{\underline{Y}}} = \sqrt{\frac{\underline{z}}{\underline{y}}} = \sqrt{\frac{r + jx}{g}} \tag{7}$$

The cable can be examined as π four-extremities, Figure 1.

$$\underline{Z}_P = \underline{Z}_C \cdot sh\underline{\gamma}l \tag{8}$$

$$\underline{Y}_P = \frac{ch\underline{\gamma}l - 1}{\underline{Z}_C \cdot sh\underline{\gamma}l} \tag{9}$$

[1] Nikolce Acevski is with the Faculty of Technical Sciences, 7000 Bitola, R. Macedonia, E_mail: nikola.acevski@uklo.edu.mk
Kire Mijoski master student, Faculty of Technical Sciences, 7000 Bitola, R. Macedonia
Tomce Mijoski master student, Faculty of Technical Sciences, 7000 Bitola, R. Macedonia

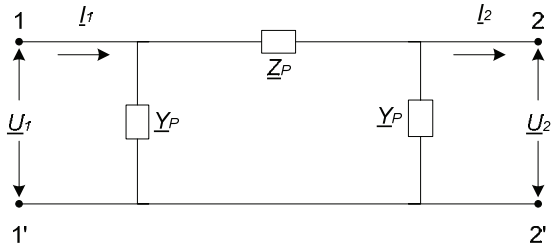


Figure 1. π four-extremities

According to [1] cables with isolated metallic shield placed without copper rope, can be presented with I-removal scheme, i.e. with one ordinal impedance $\underline{Z}_P = \underline{z} \cdot l$, $\underline{Y}_P = 0$.

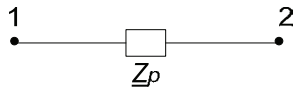


Figure 2. I-removal scheme

If cables are placed with copper rope can be examined as π four-extremities with parallel impedances ($\underline{Z}_{P1}, \underline{Z}_{P2}$) and two placed admittances to ground parameters ($\underline{Y}_{P2}$):

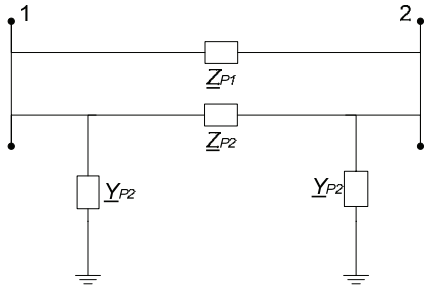


Figure 3. π four-extremities when cables are placed with copper rope or Fe Zn tape

$$\underline{Z}_P = \underline{Z}_{P1} \Big| \Big| \underline{Z}_{P2} \tag{10}$$

$$\underline{Y}_P = \underline{Y}_{P2}, \underline{Y}_{P1} = 0 \tag{11}$$

Modelling of Groundings on TS 20/0,4 kV/kV

Each TS 20/0,4 kV/kV which has own grounding with known grounding resistance *R*, in removal scheme of GS will be node, so-called "grounding place", and will be modelled with cross located active resistance *R*.

Groundings on TS 20/0,4 kV/kV in DN Gostivar are 7 types.

Type 1: Grounding on build TS 20/0,4 kV/kV, 500x500 cm.

Type 2: Grounding on mounted concrete TS 20/0,4 kV/kV for one transformer (400 (600) kVA), 452x284 cm.

Type 3: Grounding on mounted concrete TS 20/0,4 kV/kV for one transformer (630 (1000) kVA), 484x322 cm.

Type 4: Grounding on mounted concrete TS 20/0,4 kV/kV for two transformers (2x630 (1000) kVA), 584x484 cm.

Type 5: Grounding on tin TS 20/0,4 kV/kV 1 Tr 180x270 cm

Type 6: Grounding on pillar TS 20/0,4 kV/kV (grid, concrete, wooden), dimensions 180x180 cm.

Type 7: Grounding on adapted TS 20/0,4 kV/kV

III. Estimation-Currents and Potentials

When one-phase fault to ground appear at station 110/20 kV/kV "Gostivar" of the DS or the network 110 kV in nearness of the DS, comes to flux of currents per grounding of station, and the potential of grounding, V_g. The analyses with MATLAB show that most unfavourable case, from aspect of the quantity of potential of grounding, is for one-phase fault to ground produced in the station. The whole current of fault doesn't go into the earth, but just one part of it, which will be indicated with I_{gf}. This current in different ways, flows through earth to source of current.

In the paper it is supposed that current on fault to ground is known, produced in TS 110/20 "Gostivar". Data for value of the current of one-phase fault to ground in power system of R.M. are received from AD Mepso. The current I_{gf} =8100 A that is injected in GS in DS is calculated like in [7]. After that, by known (calculate or measured) value of fault current we can determine potentials for all groundings in the region of the DS. In the algorithm according to [2] is used matrix approach for solution of the problem of distributing of currents and transfer of potentials in GS in the DS. Solution of mentioned problem is done with help of matrix of impedances, [*Z*], of GS. For this purpose, primary, according to known plot of feeder 20kV and their known parameters, is generated so-called matrix of admittances [*Y*] of GS.

$$[Z] = [Y]^{-1} \tag{12}$$

Then, potentials $\underline{V}(j=1,\ldots n)$, of the separate groundings are:

$$\underline{V}_j = \underline{Z}_{ji} \cdot \underline{I}_i \tag{13}$$

IV. Results of the Estimation

Different types of groundings of TS MV/LV which are represented in the distribution network, in model network are represented as a concentrated active resistance Table 1.

Details are developed and established appropriate models for power cables with isolated metallic shield type XHP 48-A and XHE 49-A. Here we consider two cases:

1. When cables with isolated metallic shield are placed with copper rope, the results are in the Table 2
2. Cables with isolated metallic shield are placed without copper rope, the results are represented in the Table 3

Table 1 Characteristics of the protective groundings on TS 20/0.4 kV, connected on the TS 110/20 kV Gostivar

Num.	Type	Rz[Ω]	Ed[%]	Ec[%]
1	Type 1	5,04	27	15
2	Type 2	3,812	18	14
3	Type 3	4,19	17	13
4	Type 4	3,756	18	13
5	Type 5	4,65	15	13
6	Type 6	6,22	16	15
7	Type 7	2,76	16	12
8	TS 110/20	0,7	12	7

Table 2 Voltage and current values in GS, case 1

TS	Name	U(V)	Ed(V)	Ec(V)	Iz(A)
1	TC 110/20kV Gostivar	258	30,96	18,06	368
2	Magacini	118	21,24	16,52	31
3	Industriska Zona	109	29,43	16,35	22
4	Mlekara	106	19,08	14,84	28
5	Zac	0	0,00	0,00	0
6	Gudalat	0	0,00	0,00	0
7	Pitanica	48	8,16	6,24	12
8	Pitanica 1-Vukica	33	5,61	4,29	8
9	Pitanica 2-Kula	24	6,48	3,60	5
10	Javor	17	3,06	2,38	4
11	Stojce	14	2,10	1,82	3
12	Sv. Petka	14	3,78	2,10	3
13	Staklara 1	33	5,61	4,29	8
14	Staklara 2	32	5,44	4,16	8
15	Klanica G. Polog	24	6,48	3,60	5
16	Ciglana Nova	12	2,04	1,56	3
17	Ciglana	11	1,87	1,43	3
18	Mlin	146	24,82	18,98	35
19	Star Dekon	24	4,08	3,12	6
20	Goteks 1	20	3,60	2,60	5
21	Goteks 2	19	3,42	2,66	5
22	Crkva	17	4,59	2,55	3
23	D. Vlahov	14	2,38	1,82	3
24	Mal Park	11	1,87	1,43	3
25	I. L. Ribar	9	1,53	1,17	2
26	Sreten	5	0,90	0,70	1
27	Golem Pazar 1	3	0,51	0,39	1
28	Golem Pazar 2	3	0,51	0,39	1
29	Dr. Ante	2	0,34	0,26	0
30	Vojne	1	0,17	0,13	0
31	Banjesnica	1	0,17	0,13	0
32	Medat	120	20,40	15,60	29
33	M. Stopanstvo	64	10,88	8,32	15
34	Vase	71	12,07	9,23	17
35	Pozaren Dom	44	7,48	5,72	10
36	Podrumska (Stojan)	47	7,52	5,64	17
37	Getro Zeleznicka	34	6,12	4,42	9
38	Z. Stanica	31	5,58	4,34	8
39	Det. Gradinka 2	23	3,91	2,99	5
40	Avtobuska	22	3,74	2,86	5
41	Det. Gradinka 1	8	1,20	1,04	2
42	Balindolska	4	0,72	0,56	1
43	Dobre	3	0,54	0,42	1
44	Bejta	3	0,48	0,45	0
45	7-ma Zona	214	36,38	27,82	51
46	Teh. Uciliste	107	18,19	13,91	26
47	Boro	65	11,05	8,45	16
48	G. Delcev Osmoletka	56	9,52	7,28	13
49	Cveta	66	11,88	9,24	17
50	B. Kidric	29	4,93	3,77	7
51	Meljami	29	7,83	4,35	6
52	Policija	28	4,76	3,64	7
53	Garazi	32	5,44	4,16	8
54	Posta	16	2,72	2,08	4
55	Gostirvancanka	15	2,70	2,10	4
56	JNA	11	2,97	1,65	2
57	Getro S. Banka	8	1,44	1,12	2
58	7-ma Zona 1	118	20,06	15,34	28
59	Kula	32	5,44	4,16	8
60	Sumsko	25	4,25	3,25	6
61	K. Jovanoska	21	3,57	2,73	5
62	Vodna 2	16	4,32	2,40	3
63	Mesut	15	2,70	2,10	4
64	Kulturen Dom	21	3,57	2,73	5
65	G. Delcev	26	4,42	3,38	6
66	Parapunov	28	7,56	4,20	6
67	Livadi	99	16,83	12,87	24
68	Grobista	28	4,76	3,64	7
69	Mak. Pat	25	4,50	3,50	7
70	Mlaki	23	3,91	2,99	5
71	Korab Mermer	22	5,94	3,30	4
72	6 Zona 1	18	3,24	2,52	5
73	6 Zona 2	13	2,34	1,82	3
74	6 Zona 3	12	2,04	1,56	3
75	Novo Malo	10	2,70	1,50	2
76	Kasarni Tanja	7	1,26	0,91	2
77	Bolnica	6	1,08	0,84	2
78	Gradski Park 2	6	1,02	0,78	1
79	Gradski Park 1	6	1,08	0,78	1
80	DTV Partizan	5	0,85	0,65	1
81	Sud	2	0,32	0,24	1
82	Rodna B. Trud	1	0,18	0,14	0
83	Leska Nova	1	0,17	0,13	0
84	Rizvance	55	9,35	7,15	13
85	Petko	31	5,27	4,03	7
86	Posta Semafori	25	4,25	3,25	6
87	Sumatoska	17	3,06	2,38	4
88	Sijam	9	1,53	1,17	2
89	Benz. Pumpa	6	1,02	0,78	1
90	Vodna	5	0,85	0,65	1
91	Detska Gradinka	3	0,54	0,42	1
92	Satki	3	0,54	0,42	1
93	Stadion	1	0,27	0,15	0
94	Simun-Dutlok	1	0,18	0,14	0
95	Lak Vrbjanci	1	0,17	0,13	0
96	Gimnazija	1	0,17	0,13	0
97	Dutlok 3 (Mitre)	1	0,15	0,13	0
98	Abduraim	1	0,18	0,14	0
99	Stara Bolnica	1	0,17	0,13	0
100	Gorceva Vodenica	1	0,18	0,14	0
101	Leska	1	0,27	0,15	0

Table 3 Voltage and current values in GS, case 2

TS	Name	U(V)	Ed(V)	Ec(V)	Iz(A)
1	TC 110/20 kV	1000	120,00	70,00	1429
2	Magacini	697	125,46	97,58	183
3	Industriska Zona	669	180,63	100,35	133
4	Mlekara	655	117,90	91,70	172
5	Zac	1	0,17	0,13	0
6	Gudalat	1	0,17	0,13	0
7	Pitanica	471	80,07	61,23	112
8	Pitanica 1-Vukica	409	69,53	53,17	98
9	Pitanica 2-Kula	369	99,63	55,35	73
10	Javor	339	61,02	47,46	89
11	Stojce	250	37,50	32,50	54
12	Sv. Petka	245	66,15	36,75	49
13	Staklara 1	592	100,64	76,96	141
14	Staklara 2	578	98,26	75,14	138
15	Klanica G. Polog	401	108,27	60,15	80
16	Ciglana Nova	322	54,74	41,86	77
17	Ciglana	316	53,72	41,08	75
18	Mlin	784	133,28	101,92	187
19	Star Dekon	342	58,14	44,46	82
20	Goteks 1	351	63,18	45,63	94
21	Goteks 2	341	61,38	47,74	91
22	Crkva	276	74,52	41,40	55
23	D. Vlahov	247	41,99	32,11	59
24	Mal Park	222	37,74	28,86	53
25	I. L. Ribar	208	35,36	27,04	50
26	Sreten	158	28,44	22,12	41
27	Golem Pazar 1	129	21,93	16,77	31
28	Golem Pazar 2	126	21,42	16,38	30
29	Dr. Ante	106	18,02	13,78	25
30	Vojne	99	16,83	12,87	24
31	Banjesnica	98	16,66	12,74	23
32	Medat	665	113,05	86,45	159
33	M. Stopanstvo	484	82,28	62,92	115
34	Vase	518	88,06	67,34	124
35	Pozaren Dom	411	69,87	53,43	98
36	Podrumska Stojan	411	65,76	49,32	149
37	Getro Zeleznicka	351	63,18	45,63	93
38	Z. Stanica	335	60,30	46,90	88
39	Det. Gradinka 2	296	50,32	38,48	71
40	Avtobuska	286	48,62	37,18	68
41	Det. Gradinka 1	197	29,55	25,61	42
42	Balindolska	144	25,92	20,16	38
43	Dobre	137	24,66	19,18	36
44	Bejta	136	21,76	20,40	22
45	7-ma Zona	926	157,42	120,38	221
46	Teh. Uciliste	680	115,60	88,40	162
47	Boro	542	92,14	70,46	129
48	G. Delcev Osmolet	473	80,41	61,49	113
49	Cveta	667	120,06	93,38	175
50	B. Kidric	344	58,48	44,72	82
51	Meljami	341	92,07	51,15	68
52	Policija	336	57,12	43,68	80
53	Garazi	363	61,71	47,19	87
54	Posta	263	44,71	34,19	63
55	Gostirvancanka	257	46,26	35,98	67
56	JNA	229	61,83	34,35	45
57	Getro S. Banka	200	36,00	28,00	53
58	7-ma Zona 1	782	132,94	101,66	187
59	Kula	393	66,81	51,09	94
60	Sumsko	366	62,22	47,58	87
61	K. Jovanoska	346	58,82	44,98	83
62	Vodna 2	320	86,40	48,00	63
63	Mesut	310	55,80	43,40	81
64	Kulturen Dom	345	58,65	44,85	82
65	G. Delcev	369	62,73	47,97	88
66	Parapunov	377	101,79	56,55	75
67	Livadi	677	115,09	88,01	162
68	Grobista	380	64,60	49,40	91
69	Mak. Pat	370	66,60	51,80	97
70	Mlaki	357	60,69	46,41	85
71	Korab Mermer	351	94,77	52,65	70
72	6 Zona 1	322	57,96	45,08	85
73	6 Zona 2	281	50,58	39,34	74
74	6 Zona 3	277	47,09	36,01	66
75	Novo Malo	253	68,31	37,95	50
76	Kasarni Tanja	215	38,70	27,95	57
77	Bolnica	195	35,10	27,30	51
78	Gradski Park 2	181	30,77	23,53	43
79	Gradski Park 1	174	31,32	22,62	46
80	DTV Partizan	159	27,03	20,67	38
81	Sud	114	18,24	13,68	41
82	Rodna B. Trud	102	18,36	14,28	27
83	Leska Nova	85	14,45	11,05	20
84	Rizvance	493	83,81	64,09	118
85	Petko	392	66,64	50,96	94
86	Posta Semafori	357	60,69	46,41	85
87	Sumatoska	299	53,82	41,86	79
88	Sijam	235	39,95	30,55	56
89	Benz. Pumpa	203	34,51	26,39	48
90	Vodna	187	31,79	24,31	45
91	Detska Gradinka	165	29,70	23,10	43
92	Satki	159	28,62	22,26	42
93	Stadion	83	22,41	12,45	16
94	Simun-Dutlok	77	13,86	10,78	20
95	Lak Vrbjanci	75	12,75	9,75	18
96	Gimnazija	96	16,32	12,48	23
97	Dutlok 3 (Mitre)	86	12,90	11,18	18
98	Abduraim	82	14,76	11,48	21
99	Stara Bolnica	87	14,79	11,31	21
100	Gorceva Vodenica	90	16,20	12,60	24
101	Leska	96	25,92	14,40	19

V. CONCLUSION

1. The biggest danger of transferred potentials in 110kV network in region of the distribution network at Gostivar is when fault to ground appear in TS 110/20kV/kV.
2. Using the MATLAB programming package is estimated distribution fault to ground in to the GS, fault random places on the network.
3. When cables are placed without copper rope, we can see that appear several critical points. Otherwise when cables are placed with copper rope, there are no critical points.
4. Grounding systems do not play a direct part in the normal power flow but are very important in ensuring that insulation failures can be promptly detected and isolated by proper selection of system grounding. The other major function is to ensure that no unsafe voltages appear in any external or extraneous conducting parts of an electrical system. A good knowledge of GS is necessary to design a safe system and ensuring continued safe operation.

REFERENCES

[1] N. Acevski, "Prilog kon modelite za resavanje i analiza na zazemjuvaci i zazemjuvacki sistemi", doktorat, Skopje, 2003.

[2] R. Ackovski, „Zazemjuvaci i zazemjuvacki sistemi vo elektroenergetskite mrezi", septemvri 2008

[3] Zbirka tehnicki preporaki TP-1 do TP-15 EPS, mart 2001

[4] D. Tasic, „ Osnovi elekroenergetski kablovske tehnike", maj 2001

[5] D. Rajicic, „ Distributivni elektroenergetski sistemi", 1995

[6] M. Todorovski, „Upatstvo za rabota vo Matlab" januari 2009

[7] H. Trajkovski, M. Atansovski, L. Trpezanovski, C4„ INTERPSS, OPEN-SORCE Softver za analiza na EES, Model na makedonskiot EES za presmetka na kusi vrski"

Potential Characteristics of Single and Group Earthing Devices

Marinela Yordanova [1] Borislav Hr. Dimitrov[2]

***Abstract* – The paper proposes a model for obtaining the alteration of the potential in the area around a single and a group earthing device in homogeneous and heterogeneous soil, and verification of the adequacy of the model. The aim is to study the possibility of obtaining the potential characteristics of the earthing devices by modelling and their use in the process of earthing equipment design.**

***Keywords* – Earthing Electrode, Potential characterization**

I. INTRODUCTION

Main technical protection measure against indirect contact, which can be used separately or in common with other measures is the protective earthing. Requirements for earthing and protection against direct and indirect contact are subject to many legal documents that are constantly evolving and updated. There is a trend for development and improvement of methodologies for the design of earthing systems, including more accurate models of soil, taking into account the natural earthing, underground and above ground infrastructure, use of modern software, increasing the accuracy in determining the values of foot or contact voltages depending on the configuration of the earthing installation. The paper proposes a model for obtaining the alteration of the potential in the area around a single and a group earthing device in homogeneous and heterogeneous soil, and verification of the adequacy of the model. The aim is to study the possibility of obtaining the potential characteristics of the earthing devices by modelling and their use in the process of earthing equipment design

II. RESEARCH MODEL

To find the potential alteration around individual and group earthing devices (potential characteristics) electrostatic task is solved [1, 2].

A model of a single vertical earthing electrode in a homogeneous environment is shown in Fig.1. The following assumptions are made:

• The soil layer around the electrode is homogeneous, with constant specific electrical resistance and dielectric conductivity;

• The electrode has dimensions: length - 1m and diameter of 0, 08 m;

• The electrode is on the ground surface;

• Current flows through the electrode is 10 A.

• The soil layer around the electrode is homogeneous, with constant specific electrical resistance and dielectric conductivity;

• The electrode has dimensions: length - 1m and diameter of 0, 08 m;

• The electrode is on the ground surface;

• Current flows through the electrode is 10 A.

The boundary conditions are as follows:

• The potential's alteration around the earthing electrode on the ground surface is given;

• At the other boundary surfaces at a distance of 20 m from the electrode potential is taken as zero, which corresponds to the data practices [3].

Fig. 1 shows equipotential lines around the electrode and Fig. 2. – potential's alteration with distance from the electrode according Fig.1.

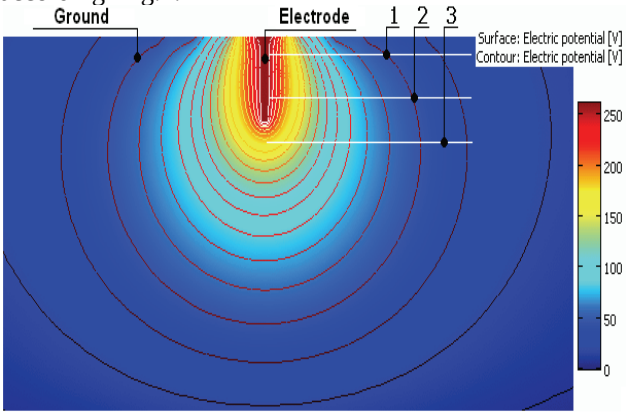


Fig. 1. Model of earthing electrode. Lines 1,2 and 3 – lines of observing potential's alteration

A model of a single vertical earthing electrode, laid on the ground in heterogeneous layer of soil around it is shown in Fig.3, and Fig.4 shows the results obtained in this case. The following assumptions are made:

- The soil layer around the electrode is divided into three layers with different specific electrical resistance and dielectric conductivity, taking into account the real case of moisture in different depth of soil;
- The electrode has dimensions: length - 1m and a diameter of 0,08 m;
- The electrode is on the ground surface;
- Current flows through the electrode is 10 A.

The boundary conditions are the same as in the first case.

[1]Assoc. Prof. PhD Marinela Yordanova – Department of Electrical Power Engineering, TU- Varna, e-mail: mary_2000@abv.bg

[2]Assist Prof. PhD. Borislav Dimitrov – Department of Electrical Engineering and Electrical Technologies,TU- Varna, e-mail: bdimitrov@processmodeling.org

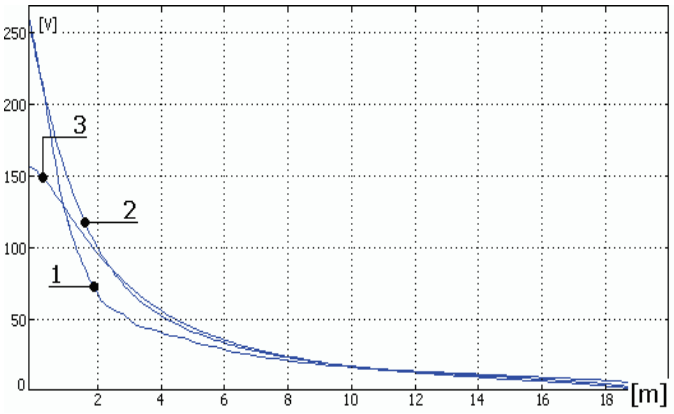


Fig. 2. Potential's alteration. 1,2 and 3 are according to Fig.1.

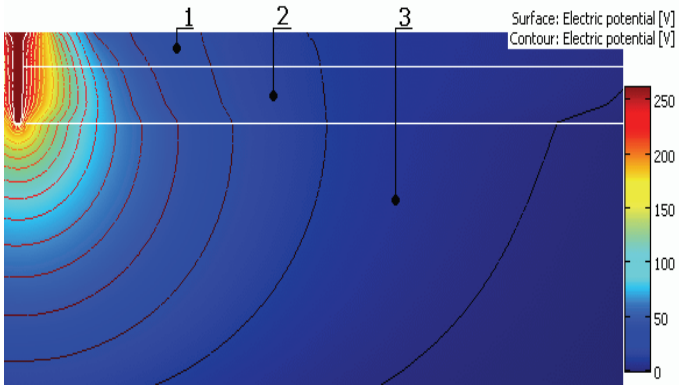


Fig. 3. The field in the heterogeneous soil
1 – wet soil ε=25; 2 – moist soil ε = 14; 3 – dry soil ε = 5

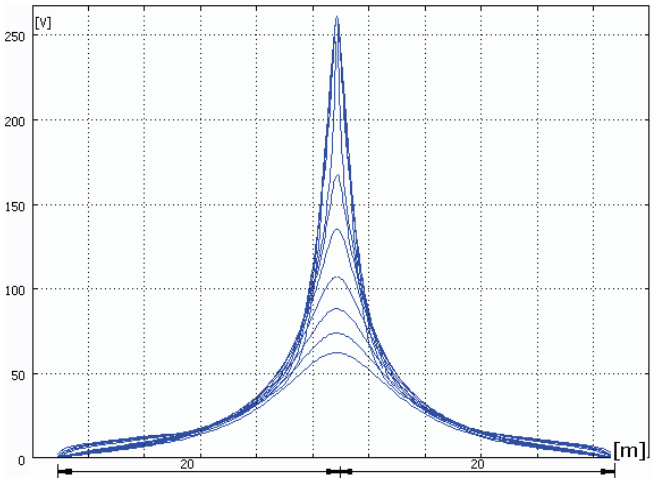


Fig. 4: Alteration of the potential around vertical earthing electrode

Fig.5. shows a model of earthing group of two single vertical electrodes placed on the land surface in the homogeneous soil layer around and Fig.6- the results obtained in this case. Assumptions, dimensions and boundary conditions are as in the first model.

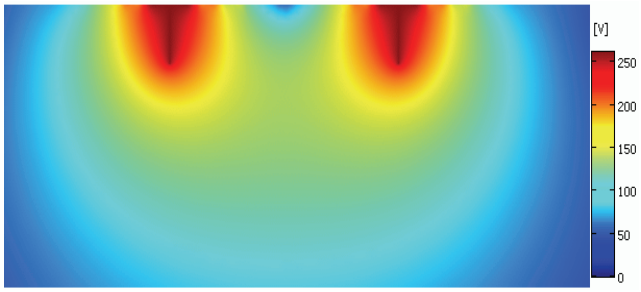


Fig. 5. The field around two electrodes

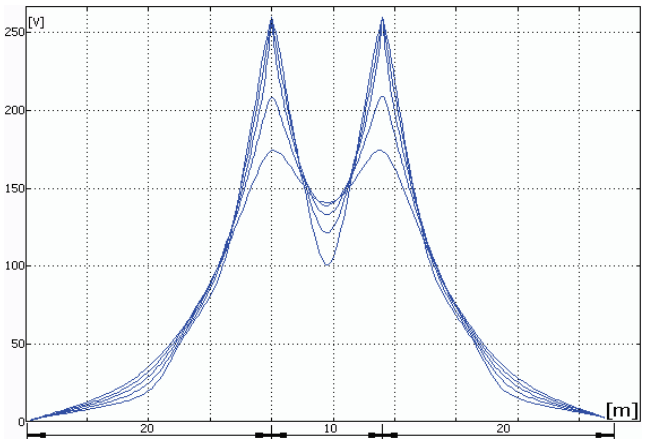


Fig. 6. Alteration of the potential between and out of two electrodes

III. Conclusions

The received results from modeling were compared with calculation results obtained using known in the bibliography and well known formulas for potential characteristics. They are mainly applicable for single earthing rods. The potential characteristics for group earthing device can be obtained graphically using the individual electrode's potential characteristic. The advantage of using modeling to obtain potential characteristic is primarily related to the ability to obtain the characteristics of grounding group comfortably, quickly, in graphic form and accuracy, corresponding to the theoretical calculations.

Other direction, which can be claimed advantage of models, is related to heterogeneity of a soil. Offer complex theoretical dependencies it mostly for double-layer model of soil. Considerably more complex studies of heterogeneous soils with different distribution of layers around the earthing rods can be made using models.

References

[1] Mihailov, M. I., S.A. Sokolov. Earthing devices in telecommunication installations, Moscow, 1971.
[2] Anev G. Harmful action of electricity and the protection of them, Technika, Sofia, 1987.
[3] Ushev, G., M. Yordanova. Technical Safety Notes, Part 2, Varna, 2003.

Analysis of the Residual Current Devices operation in IT and TT systems

Marinela Yordanova[1], Mediha Mehmed-Hamza[2]

Abstract **– The authors of the paper have developed imitation models of one- phase and three- phase residual current devices and the different situations of direct and indirect contact of a person in an electrical network low voltage. TT and IT systems under normal and fault modes are considered. Authors' own model schemes for low voltage networks are used.**

Keywords **– Direct Contact, Electrical safety, Indirect Contact, RCD, IT, TT systems**.

I. INTRODUCTION

Residual current devices (RCDs) are becoming more widely used, together with the implementation of low-voltage network as TN-S system. The protective disconnection is regarded as a protective measure that provides the highest degree of safety. Residual current protective devices have been subject to research by the authors. The results for modeling RCDs as part of the electrical installation low voltage and also their work in other fault situations have been published in [3, 4]. Object modeling was residual current devices, worked by a zero sequence currents and the nominal operating threshold $I_{\Delta n}$= 30 mA.

The protective disconnection is a precaution applied commonly with protective earthing in systems low voltage IT and TT. Using RCD provides protection against direct contact in the low voltage system, but only as an additional protective measure.

Several cases of direct and indirect contact of a person in systems IT, TT and TN-C-S are considered in [1]. They are analyzed and the conclusion is that RCDs do not trip and respectively the person is at risk.

The authors have done a research with developed in [4] simulation model of single-phase two- pole RCD and created for the purpose of this paper such three-phase model. Different situations of direct and indirect contact of a person in normal and fault cases in IT and TT network, listed in [1], have been researched by authors. Model schemes for low voltage networks [2] were used.

[1]Assoc. Prof. PhD Marinela Yordanova – Department of Electrical Power Engineering, TU- Varna, e-mail: mary_2000@abv.bg

[2]Assist Prof. PhD. Mediha Mehmed-Hamza – Department of "Electrical Engineering and Electrical Technologies", TU- Varna, e-mail: mediha.hamza@mail.bg

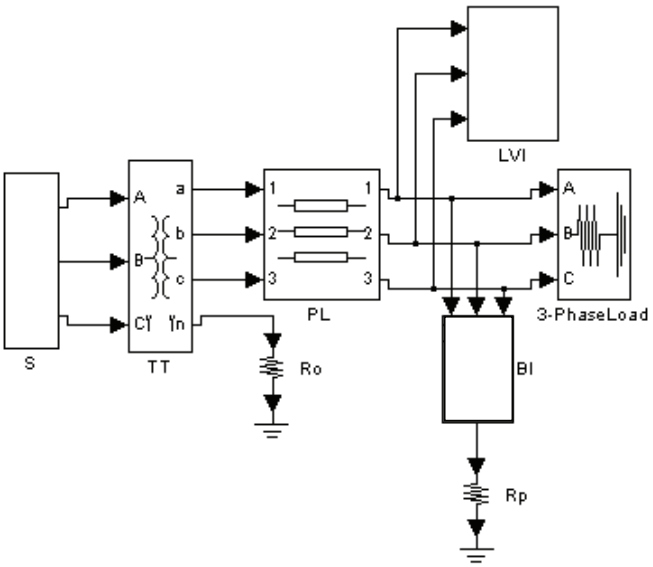


Fig.1. Example for simulation model of network TT; S- Power System; TT- Supply Transformer; PL- Cable Power Line; LVI- Low voltage installation with RCD; BI – Basic Insulation of three phase load; Ro- neutral's earthing; Rp- protective earthing

II. RESEARCH MODELS OF IT AND TT

The block, modeling RCD, is included in the model of an electrical low voltage installation of TT and IT (Fig.2). The block RCD consists of:

- Summing Block SB, monitoring the Rms value of the vector sum of the currents flowing through the phase (or phases) and the neutral conductor (if there is) at a point of the electrical installation;
- Block Logic LB - after SB the signal is given to this block and the time- current characteristic curve Type G is realized for RCD Type AC (Table 1);
- Block „Switching" – interrupts the supply after the time delay. It is in LB.

Table 1 shows the realized tripping curve of RCD and the standardized tripping times.

TABLE I
MAXIMUM TRIPPING TIME

Tripping curve	I_n, A	$I_{\Delta n}$, A	Tripping time, s		
			$I_{\Delta n}$	$2.I_{\Delta n}$	$5.I_{\Delta n}$
G	All values	All values	0,22	0,12	0,06

III. RESEARCHED CASES

The TT system is with earthed neutral. The value of the earthing rod of the supply transformer's neutral is R_0= 4 Ω.

TT network

Case 1 (Fig. 2): Direct contact with a live part (phase L2) and the casing of an equipment (exposed conductive parts). At the same time there is a faulting-to-ground neutral N. The person, contacting with these parts, is replaced with two resistance: a resistance of the way "hand - hand" R_{h1}= 800 Ω between the phase L2 and the casing and a resistance of the way "hands - feet" R_{h2}= 1200 Ω between the casing and the ground.

In TT power systems all the exposed conductive parts have a direct electrical connection to protective earthing installation with resistance R_p = 4Ω, independently from the earthing of the supply system (e.g. utility). In this case [1] a faulting-to-ground neutral downstream of the RCD, can create a parallel path, which again closes via N, and earth leakage current might be not large enough to cause tripping RCD. The simulation results confirmed this reasoning. RCD fails, but the person has a current value in the range below 1 mA.

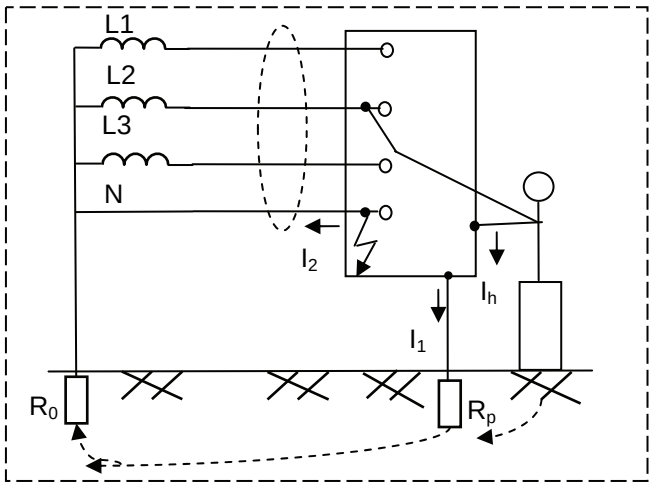


Fig.2 Direct contact with a live part and the exposed conductive parts in TT system with fault- to- ground neutral; R_0 – supply transformer's neutral resistance; R_p- protective earthing system's resistance

Case 2 (Fig. 3): Direct contact with the phase L1. Two phases L2 and L3 are leaking to ground while, at the same time, a direct contact is occurring with the other healthy phase (Fig. 3). The person's body is replaced with resistance Rh = 1000 Ω between L1 and the ground, and the insulation failure of the phases is replaced with a resistance less than 1Ω. The system in this case is with no neutral conductor N. RCD operates by carrying out the vector summation of the three-phase currents I_f on a three-phase circuit and comparing the result to its operating threshold. Since fault currents of the three phases via ground are different, I_{L2} and I_{L3} will be more than I_{L1}. If I_f is less than the RCD threshold, this might not trip, despite a potentially dangerous current via the person [1].

The simulations showed the following results: the current via human's body is in the order of 280 mA. Current in the secondary winding of the transformer of RCD is about 20 A and according to Table 1 tripping occurs in 0, 06 s. The value of the resistance R_f of insulation fault affects the current of the secondary winding of transformer. When R_f = 1000 Ω RCD does not trip and the value of current through the human's body also has a dangerous value - about 230 mA.

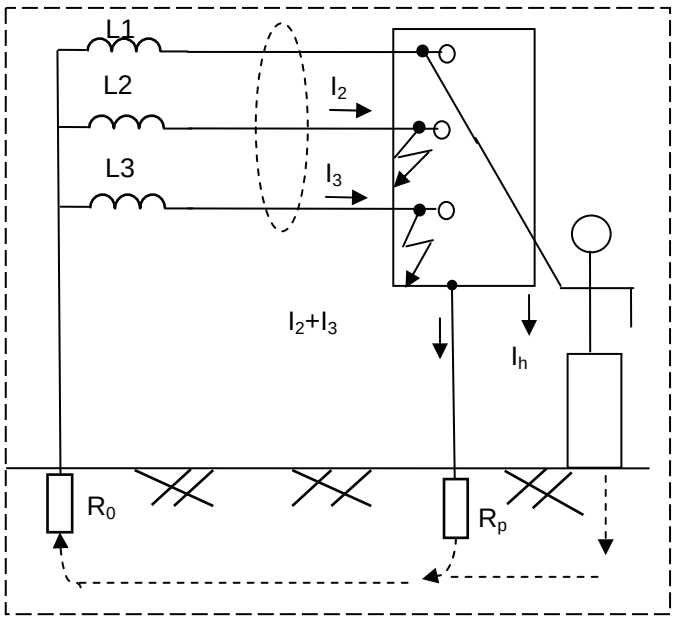


Fig.3. Direct contact in TT system with fault- to- ground phases L_2 and L_3

Case 3 (Fig. 4): Indirect contact with the exposed conductive parts of equipment where there is a fault - to - protective conductor phase. This system is with a neutral conductor and the considered load is a single-phase consumer. The person's body is replaced with resistance Rh = 1000 Ω and the insulation failure of the phase is replaced with a resistance less than 1 Ω.

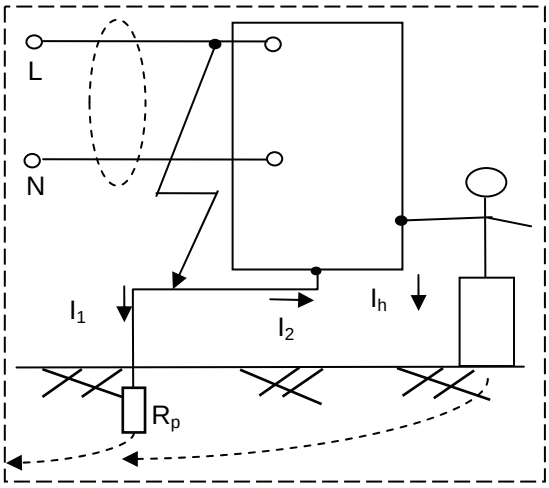


Fig. 4. Indirect contact with the exposed conductive parts in TT system and fault- to- protective conductor phase

In this case the exposed conductive parts are under voltage due to their connection to the protective earthing system.

Thus, in contact with them, the body will be under that voltage, while the current through it may not be sufficient to trip RCD [1]. The simulations showed the following results: the current via person is dangerous; the current in the secondary winding of transformer is sufficient to trip RCD. The tripping occurs in 60 ms. But when the installation point of RCD is after the fault point, it can not be activated.

IT network

Case 4 (Fig. 5): Direct contact with the phase. The person is replaced with resistance Rh = 1000 Ω between L1 and ground. Although the neutral is insolated from ground live conductors have "connection" with the ground through capacitive conductivity, which depend on the length of the line. In the case of large capacitive conductivity RCD can not be activated by direct contact of a person [1].

The simulations showed the following results: current in the secondary winding of transformer of RCD depends on the length of the line and the value of Y.

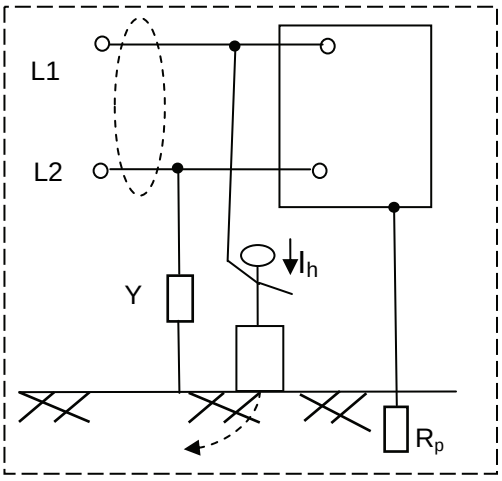


Fig. 5. Direct contact in IT system

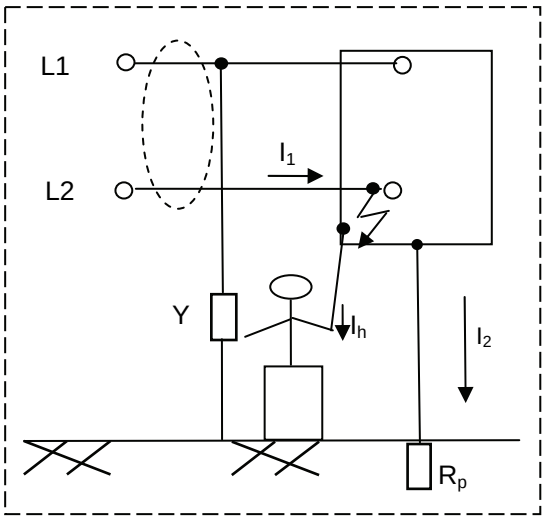


Fig. 6. Indirect contact in IT system

Case 5 (Fig. 6): Indirect contact with the exposed conductive parts of the consumer under insulation fault. Such a situation as described in case 4 may occur in indirect contact with large capacitive values of the conductivity [1].

Simulations showed that the current in the secondary winding of transformer of RCD depends on the length of the line and the value of Y.

IV. Analysis and conclusion

In general, research of IT and TT networks confirmed the presented consideration in [1] about operation of the protection, with the exception of one case. Again confirm the conclusion of [3, 4] that visualizing the values of currents via person and other elements of the network makes it possible to make more accurate analysis of situations of direct and indirect contact in the presence of RCD built. The fulfilled research presented:

- At direct contact with a live part and the exposed conductive parts in TT system with fault- to- ground neutral RCD fails, but the person has a current value in the range below 1 mA;
- At direct contact in TT system with the health phase and fault- to- ground other phases the value of the resistance R_f of insulation fault affects the current of the secondary winding of transformer. When R_f = 1000 Ω RCD does not trip and the value of current through the human's body also has a dangerous value - about 230 mA.
- At the indirect contact with the exposed conductive parts in TT system and fault- to- protective conductor phase the current via person is dangerous and the current in the secondary winding of transformer is sufficient to trip RCD. The tripping occurs in 60 ms. But when the installation point of RCD is after the fault point, it can not be activated.
- At direct and indirect contact in IT system simulations showed that the current in the secondary winding of transformer of RCD depends on the length of the line and the value of Y.

Finally, do not forget that the RCDs are an additional protective measure. The electrical safety is a complex and must include and other requirements such the proper design and the maintenance of electrical networks and equipment.

References

[1] Mitolo M, Shock Hazard in the Presence of Residual Current Devices, Industrial & Commercial Power Systems Technical Conference, 2007. ICPS 2007. IEEE/IAS Volume , Issue , 6-11 May 2007 Page(s):1 - 65.

[2] Yordanova M, M. Vasileva. Research of the work of the surge protective devices in the low voltage electrical grids. Power engineering, Bulgaria, Number 5, 2006 г., p. 20-25, (Bulgarian).

[3] Yordanova, M., M. Hamza, M. Vasileva, Analysis of the residual current devices operation in TN system, SIELA 2009, 4-6 June, Bourgas, Bulgaria, Vol.1, 338- 343.

[4] Yordanova, M., M. Hamza, M. Vasileva, Model research on the operation of Residual Current Devices in electrical networks low voltage, ICEST 2009, 25-27 June, 2009, V. Tarnovo, Bulgaria.

This page intentionally left blank.

Overvoltage Protection of a Cable Part in an Electrical Grid 20 kV

Margreta P. Vasileva[1]

***Abstract* – The problem solved in this paper is determination of the cable length when the Chopped Wave Withstand Level is not exceed in the unprotected cable end.**

A model scheme of the electrical grid is created for variant researches. Numerical and graphic results are given. The variant research depends on the using surge protective devices type and the way of laying of the cables.

***Keywords* – Overvoltage protection, cable line 20 kV.**

I. INTRODUCTION

The question connected whit the right choice of the surge protective devices` type and their installing place has a direct attitude whit the devices insulation coordination.

The insulation levels characterize the specific requirements for insulation systems of devices according to theirs function and voltage class.

The Metal Oxide Surge Arresters (MOSA) reduce the overvoltage sooner as the spark-gap arresters. The last ones are able to conduct after the overvoltage was increased to U_p. That is why their protection distance is shorter in many cases. This means that the overvoltage to the electrical equipment is higher when a spark-gap arrester instead of a MOSA is installed as both types of arresters are at the same distance from equipment to be protected [3].

II. RESEARCH VARIANTS AND RESULTS

The task of this study is to determine the length of power cables (Fig. 1) when the Chopped Wave Withstand Level (CWWL) is not exceed in the unprotected edge of the cable line (CL), depending on the type of used MOSA and of the type laying of the CL. The insulation level will not be exceeded if the selected MOSA between the CL and power line (PL) to provide the necessary protection in the unprotected edge or wave surge has abated sufficiently and it is not dangerous at the end of the power cable (PC).

Figure 1 shows one scheme decision for limiting of overvoltages in the electrical grid 20 kV according to [1].

Model scheme of electrical grid 20 kV is presented in [2].

The results for the length of power cables dependent on MOSA - type MWK 19 and the single-phase power cables whit cross- section 95, 120, 150, 185 mm are summarized in Table 1. The power cables are whit polymeric insulation.

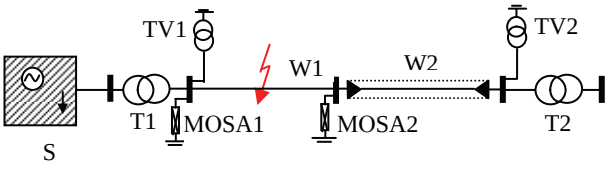


Fig. 1. One-line diagram of power grid 20 kV whit cable connection to the transformer
W1-PL; W2-CL; TV1,TV2 – voltage measure transformers, S – power system 110 kV; T1 – power transformer 110/20 kV; T2 – power transformer 20/0,4 kV.

TABLE I
LENGTH OF POWER CABLE (L_{KEP}), m, WHEN THE CHOPPED WAVE WITHSTAND LEVEL (CWWL) IS EXCEED IN THE UNPROTECTED EDGE OF THE CABLE LINE

Cable type	CAXEκT 95		CAXEκT 120		CAXEκT 150		CAXEκT 185	
MOSA	△	○○○	△	○○○	△	○○○	△	○○○
MWK 19	105÷ 478	70÷ 524	104÷ 475	67÷ 522	102÷ 472	65÷ 521	100÷ 470	63÷ 420

Table 1 shows the values for L_{KEP} depending on the cross-section of the power cables and their way of laying. The results show that L_{KEP} depends largely on the way of laying the power cables and not depend substantially on their cross-section.

Fig. 2 shows the voltages at the cable end depending of its length for MOSA - type MWK 18 and different ways of laying the PC.

Results on L_{KEP}, depending on the type of MOSA and way of laying of PC are summarized in Table 2.

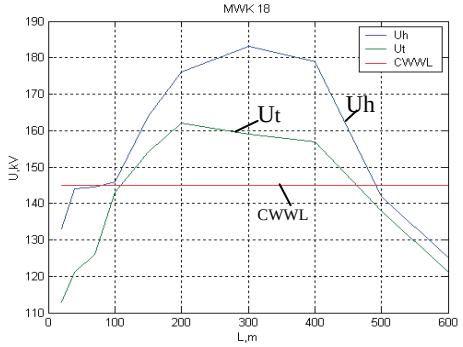


Fig. 2. Voltages at the cable end whit MOSA type MWK 18
Uh - in horizontal installation; Ut - in triangular installation; CWWL - Chopped Wave Withstand Level

[1]Margreta P.Vasileva, Technical University of Varna, Department of Electric Power Engineering, Studentska 1, Varna, 9010, Bulgaria, m.vasileva@tu-varna.acad.bg

TABLE II

LENGTH OF POWER CABLE (L_{KEP}), m, WHEN **THE CHOPPED WAVE WITHSTAND LEVEL IS EXCEED** IN THE UNPROTECTED EDGE OF THE CABLE LINE DEPENDING ON THE WAY OF LAYING THE PC AND THE TYPE OF MOSA.

	MWK 19	MWK 20	MWK 23	MWK 24	MVK 19	MVK 20	MVK 22	PVI 19,5	POLIM DN-20	POLIM DN-22
△	104÷475	125÷450	63÷473	19÷505	130÷445	120÷465	50÷450	120÷454	122÷468	52÷453
○○○	67÷522	57÷505	42÷512	19÷568	53÷493	52÷500	33÷550	57÷500	53÷502	35÷552

Research is done and the scheme from Figure 3.

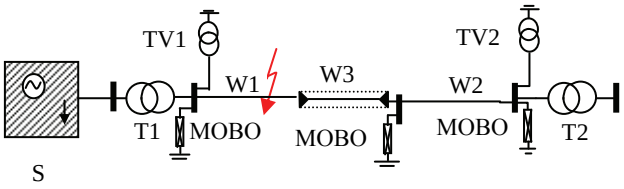


Fig. 3. One-line diagram of power grid 20 kV whit cable line between two power lines (W1, W2 - PL; W3 - CL)

To determine the maximum length of cable line with unilateral protection seen case where surge wave enters from the PL associated with unprotected edge of the cable.

The results for different lengths of the cables have the same character over time as those of fig. 2 and are not shown in the statement. The results for the various options are summarized in Table III.

III. CONCLUSION

When the lengths of PC are within the range specified in Table I and Table II is required MOSA protection at both ends of the PC.

Bilateral protection whit MOSA is recommended when the CL lengths exceed the values in Table 3.

TABLE III

MAXIMUM LENGTH OF THE CABLE LINE (L_{KLMAX}), M, OF WHICH **NOT TO EXCEED CWWL** IN THE UNPROTECTED EDGE OF THE CABLE LINE DEPENDING ON THE WAY OF LAYING THE PC AND THE TYPE OF MOSA

	MWK19	MWK20	MWK23	MWK24	MVK19	MVK20	MVK22	PVI 19,5	POLIM DN-20	POLIM DN-22
△	18	15	11	10	17	12	11	19	11	10
○○○	15	14	10	9	15	10	10	16	10	10

REFERENCES

[1] Ordinance № 3 regarding the structure of electrical devices and power lines, Sofia, 2004.

[2] Vasileva M., “Limitation of overvoltages in electric networks - 20 kV”, Disertation, Varna, 2004.

[3] Dimensioning, testing and application of metal oxide surge arresters in medium voltage networks, ABB Hige Voltave Tehnologies, July, 1999.

POSTER SESSION PO VII

PO VII - Computer Systems and Internet Technologies I

Specific Characteristics of Computer Criminal Offenses With Regard to the Law Regulations

Jelena D. Matijasevic[1] and Zaklina S. Spalevic[2]

***Abstract* – There are different categories of perpetrators of computer criminal activity. This paper puts focus on the profile of a hacker – a perpetrator of computer criminal activity who is not motivated by financial gain. It also deals with current classifications of hackers and emphasizes their important characteristics and principles of hacker ethics. The paper also gives a review of judiciary in The Republic of Serbia in this area. It is clear that a society can adequately confront negative phenomenon, only if all of its characteristics and specificities are recognized.**

***Keywords* – Computer criminal activity, Hackers, Computer technology.**

I. INTRODUCTION

Computers represent one of the most important and the most revolutionary achievements of development of technical and technological civilization. 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.

Growing use of computer technology causes an increase in computer criminal activities, as a new form of criminality in the contemporary society, and a development of its diverse forms. Computer technology is developing very quickly, at the same pace with education and training of persons who intend to misuse it. The press abounds with information about a person (or a group) that had penetrated an important government computer system and had not only acquired specific data, but had also created a possibility for endangering or activation of systems such as nuclear potential of great world forces. This phenomenon is not only a characteristic of the developed western countries – it increasingly becomes typical of the Balkans area [1].

[1]Jelena D. Matijaševic, Faculty of Law, Business Academy, Geri Karolja 1, 21000 Novi Sad, The Republic of Serbia, E-mail: jelena@pravni-fakultet.info

[2]Zaklina S. Spalevic, Faculty of Law, Business Academy, Geri Karolja 1, 21000 Novi Sad, The Republic of Serbia, E-mail: zaklinaspalevic@ymail.com

II. PERPETRATORS OF COMPUTER CRIMINAL ACTS – DIVISION AND BASIC CHARACTERISTICS

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.

In fact, 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 [2].

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 [3].

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 [4].

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 empoyees who have access to the terminals; 16% are cashiers, 11% of them are operators – informaticists, and 12% are persons outside the affected corporation, including the service users [4].

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 who break into other people's computer system, using their computer knowledge and a modem [5].

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.

III. Concept of hacker

The word 'hacker' is very often used in a negative context today, without trying to grasp the essence of activities of these persons, or to adequately analyze the reasons and consequences of their activities.

The term 'hacker'is in its original meaning used to denote a person who deals with research of computer potential and its positive application in everyday life. Hackers are highly intelligent people who explore what is hidden in hardware and software. In simpler terms, they locate something hidden from the public or find randomly made mistakes [6].

Hackers still remain a sort of enigma to the world of psychology and sociology. Understanding their development and motivation has become one of the areas of their interest.

Different authors approach hackers in different ways and analyze them from various viewpoints. The dominant attitude nowadays is the one provided in the explanation of the term 'hacker'- namely, that hackers look for errors in programs and then inform the public about it, so that the manufacturer of the given program can rectify the error and that the public can take necessary steps to protect themselves on time.

With the purpose of ensuring a more positive approach to the term 'hacker', data has supplied so as to indicate that more than 10 000 errors have been found so far and that about five new errors are found daily.

IV. Types of hacker

There are several criteria for hacker classification. According to the criterion of respect for ethics, there is a division into the following types of hacker:

- White Hat Hackers – These hackers respect hacker ethics; they deal with computer system and network protection. They try to improve protection of the information system, so as to avoid penetration into it and causing damage. They are typically rented by companies to break into a system and then inform the owner how it was done and how to improve the flaws.

- Black Hat Hackers – These hackers do not hesitate to steal and destroy data in networks and systems they penetrate into. They interpret the hacker ethics in a way they see fit. The principle that all information should be free grants them an excuse to enter into other people's systems. They often destroy a part of the system. Creating and distributing viruses and worms which damage computers belongs to their activities.

- Gray Hat Hackers – are somewhere in between Black and White Hats. They wish to be distinguished from security testers of a company on one hand, and to disassociate themselves from the negative image of Black Caps. These are mostly hackers who initially violated hacker ethics and then used the acquired knowledge according to all the rules of the ethics.

Another more detailed and precise division, where level of computer skills, sphere of interest and ethical rules are taken as a criterion, distinguishes between the following types of hackers:

- *Old school hackers* – Persons who have dealt with computers from their very emergence belong to this group. These hackers could only rely on themselves in terms of learning about computers, because they were the first to engage in research of computers and their possibilities. Copyright protection was unknown to old school hackers. Their favorite activity was to read other people's programs, to modify and expand their possibilities.
- *Phreakers* – They are hackers whose narrow specialty includes theft of dial impulses, conducting international calls at the expense of another person and all activities related to telephone traffic.
- *Crackers* – The major preoccupation of a cracker is the safety of computer systems. Their main activities include breaking into other people's computers. There is a clear difference between crackers and typical hackers – hackers find loopholes in computer systems in order to patch them, whereas crackers use such flaws to cause damage.
- *Warezdoodz* – They specialize in editing programs, finding serial numbers and their illegal distribution to the users. They are at the top of the piracy chain. The activities of Warezdoodz directly violate copyright laws; they contribute to illegal distribution and copying of programs.
- *Hacktivists* – Hacktivists use their hacking skills to promote political ideology, and thus interpret hacker ethics in their own manner, in the sense that hacking for political goals is not contrary to the ethics.

Using computers to achieve political goals was an exception in the past and it did not attract much attention. However, the development of technology enabled the unlimited access of computers to the world of politics and management, which is why the term 'cyber war' has often

been mentioned lately. The latest example of hacktivism is the so-called cyber war between Serbian and Albanian hackers, which started in the August of 2008. Moreover, Serbian and Croatian hackers led a fierce hacking battle during the year 2004, by crashing websites of TV stations, sport clubs and faculties [7].

Each classification of hackers is conditional and by no means final. It is difficult to determine the exact boundaries for some types of hackers; some types are intertwined, while some cannot function alone, without another type. In any case, the existing classifications provide us with a better overview for analysis of specific characteristics of this group of perpetrators of computer criminal acts.

V. PSYCHOLOGY OF HACKERS

If you are a good hacker, everyone knows you; if you are the best, nobody does! 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. 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 [7].

Hackers are often completely disorganized and clumsy when it comes to communicaton with people around them.

During the years 1994 and 1995, ADD (Attention-deficit disorder) was discovered in people who deal with hacking. ADD is characterized by inability of paying attention, combined with hyper-focusing on things they are interested in.

In 1999, AS (Asperger's syndrome) was discovered. This disorder is also known as „high-functioning autism". It is manifested in the inability to understand face and body language of other people, as well as in inability to express empathy with them. On the other hand, people suffering from AS have high intelligence, great analytical skills and an extraordinary ability to solve technical problems [7].

Some authors even advocate the attitude that perpetrators of compter crimes do not have a developed moral maturity.

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 [8].

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.

VI. 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:

- 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.
- 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 [7].

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.

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, refrainment, 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.

VII. LEGISLATIVE REGIME

In the area of various types of misuse of computer technology, it is necessary to adopt specific legislative regulations, which settle criminal acts committed within this sphere. It is also necessary to understand technology and individuals who commit computer crimes, so as to achieve effectiveness of legislators.

However, due to poorly developed legal regulations and control in the area of information technology, many countries have become a paradise for hackers. Still, in spite of the cyber crime expansion in the eastern countries, the USA remains by far the leading target on the list of countries attacked by hackers [7]. Difficulties that legal systems face in monitoring new criminal trends of the hackers are a source of embarrassment for governments all around the world. A typical example is that of Canada, because its Criminal Law (law of criminal acts) does not clearly define computer criminal acts. This resulted in prosecutors using metaphors to explain the criminal act, due to lack of knowledge about technology.

Computer criminal acts in the legislation of the Republic of Serbia are regulated by regulations of the Law on Organization and Jurisdiction of Government Authorities in the Suppression of High Technological Crimes [9] and in the Criminal Code of the Republic of Serbia [10].

According to the latest novelty, the criminal acts against the security of computer data have been regulated in the chapter twenty-seven (Articles 298-304a) of the Criminal Code of the Republic of Serbia. The legislator included the following criminal acts in this special group of the Code: damage of computer data and programs, computer sabotage, creation and distribution of computer viruses, computer fraud, unauthorized access to a protected computer, to a computer network and to electronic data procession, preventing and limiting access to a public computer network, unauthorized use of computer or computer network, the criminal act of creating, obtaining and providing the other person with means necessary for execution of criminal acts against the security of computer data. The latest amedmends to the Criminal Code envisage yet another form of misuse of computer and computer networks. Due to the fact that the computer network is often misused so as to commit or to conceal criminal acts against sexual freedom of minors, Article 185b in the Chapter Eighteen of the Criminal Code has regulated the criminal act of using computer network or other technical means of [11] communication to commit criminal offenses against sexual freedom of minors. Introduction of legal provisions on computer criminal acts in the criminal legislation of the Republic of Serbia has contributed to making great progress and to creating new opportunities for prevention of illegal activities and practices in this area. However, given the fact that situations dealt with in practice can be highly unpredictable and that the perpetrators of computer crimes are certainly very inventive, it is necessary to constantly renew these regulations, by adoptinng new legal provisions and making amendments to the existing ones.

VIII. CONCLUSION

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.

It was our intention to depict the profile of the perpetrator of computer criminal acts as well as possible in this paper and thus to shed light on all of his/her specific characteristics, because effective steps in eliminating negative effects of a certain phenomenon include not only understanding the phenomenon, but also understanding perpetrators of the criminal activities in that field. Transparency and determined opposition to different forms of criminal activities are two important elements of the aim to reduce different forms of crime, including computer crime, to a previously determined framework, which is endurable for that specific community.

REFERENCES

[1] http://www.sk.co.yu/1998/10/skako1.html
[2] B. Banovic, “Providing evidence in the criminal process of economic crimes “, Police College, Belgrade, 2002.
[3] 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
[4] Z. Aleksic, M. Skulic, “ Crime tactics, techniques, methods ”, „Official Gazette “, Belgrade, 2007.
[5] G. Goldman, H. J. Stenger, “Die ganze Welt als Tatort, Computer Hacking: Modus operandi und Ermittulungsprobleme”, Kriminalistik, 8-9/89, Kriminalistik Verlag, Heidelberg
[6] http://www.mycity.rs/Zastita/ko-su-hakeri.html
[7] http://www.svethakera.com
[8] The socio-psychological profile of the perpetrator of a computer crime, Faculty of Informatics and Computing, http://www.dir.singidunum.ac.rs/
[9] Law on the organization and responsibilities of state bodies for the fight against cyber crime, „ Official Gazette of the Republic of Serbia “, No. 61/2005.
[10] Kriminal Code, „ Official Gazette of the Republic of Serbia “, No. 85/2005, 88/2005, 107/2005 i 72/2009.

A Modular System for Solving Optimization Problems by Testing Them with Genetic Algorithms

Hristo I. Toshev[1], Stefan L. Koynov[2] and Chavdar D. Korsemov[3]

Abstract: – **The paper introduces a modular system for testing with genetic algorithms (GA-s) to solve optimization problems. The system consists of several modules that include blocks with different purpose which are interconnected in various ways. The GA is realized via different methods of GA set-ups, the search process, evolutionary models thus allowing to estimate the practical effectiveness of the algorithm.**

Keywords – **genetic algorithms, optimization, selection, crossover, mutation.**

I. INTRODUCTION

Genetic algorithms (GA) are a method for search based on the selection of the best species in the population in analogy to the theory of evolution of Ch. Darwin.

Their origin is based on the model of biological evolution and the methods of random search. From the bibliographical sources [1], [2] it is evident that the random search appeared as a realization of the simplest evolutionary model when the random mutations are modelled during random phases of searching the optimal solution and the selection is modelled as "removal" of the unfeasible versions.

The main goal of GA-s is twofold:

- abstract and formal explanation of the adaptation processes in evolutionary systems;
- modelling natural evolutionary processes for efficient solution of determined class of optimization and other problems.

Following [3], [4], [5] GA-s differ from other optimization search procedures according to the following below items:

- they operate with a coded set of parameters, not with the problem parameters;
- they realize the search not by improving a single solution but by the simultaneous usage of several alternatives for the given solution set;
- they use the fitness function (FF), not its different increases to estimate the quality of the taken decision;
- they apply probabilistic rules for the optimization problem, not deterministic ones.

During the last years a new paradigm is applied to solve optimization problems GA-based and modifications of GA. EA realize searching a balance between efficiency and quality of solutions at the expense of selecting the strongest alternative solution [1], [2].

[1] Hristo I. Toshev, [2] Stefan L. Koynov and [3] Chavdar D. Korsemov are with the Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 29A, 1113 Sofia, Bulgaria, E-mail: toshev@iinf.bas.bg, , slk@iinf.bas.bg, chkorsemov@iinf.bas.bg

The continuously growing number of publications and also of the practical implementations during the last years is a stable proof of the growing expansion of the scientific and application research in the domain of GA.

In order to give a general fancy for the type of applications, they could be classified in four main directions [1]:

- science – to increase the level of various scientific research [1], [2], [6], [7], [8], [9], [10], [11], [12];
- engineering – to solve specific engineering problems [5], [12], [13], [14], [15], [16];
- industry – to increase the quality and the amount of production for some industrial processes [4], [5], [14], [17], [18], [19], [20],;
- various other directions (miscellaneous applications) [1], [6], [21], [22], [23].

The present paper introduces a way to improve the qualities of applied algorithms for solving various classes of optimization problems. This is done by a modular program system that is GA-based for testing. It allows the application of elaborated various methods for set-ups of genetic operators at execution time with respect to the search process, the different evolutionary models included.

II. THE PROPOSED APPROACHES

The new program modular system to test using genetic algorithms for solving optimization problems is realized as a set of program devices consisting of various program modules each of which including different blocks: (Fig. 1):

- block 1 – Block for Input;
- block 2 – Optimization Problems Editor used for editing different optimizations tasks;
- block 3 – Set-Up Block including three set-up subblocks:
  - 3.1 – Set Up Genetic Operators block;
  - 3.2 – Set Up Search Methods block;
  - 3.3 – Set Up Evolutionary Models block.
- block 4 – Genetic Algorithm block for the already fulfilled set-ups of algorithms with genetic operators, search methods and evolutionary methods;
- block 5 – Genetic Operators Data Base;
- block 6 – Search Methods Data Base;
- block 7 – Evolutionary Models Data Base;
- block 8 – Adaptation Block;
- block 9 – Convergence Analysis Block;
- block 10 – Output Bloc for the Results.

Based on the already presented blocks a synthesis was performed of several program modules that are combinations of the introduced blocks.

MODULE 1 is used to explore various modifications of genetic operators: a recombination operator, a crossover

operator, a mutation operator, an inversion operator, a segregation operator, a translocation operator, of their modifications and also of their joint operation.

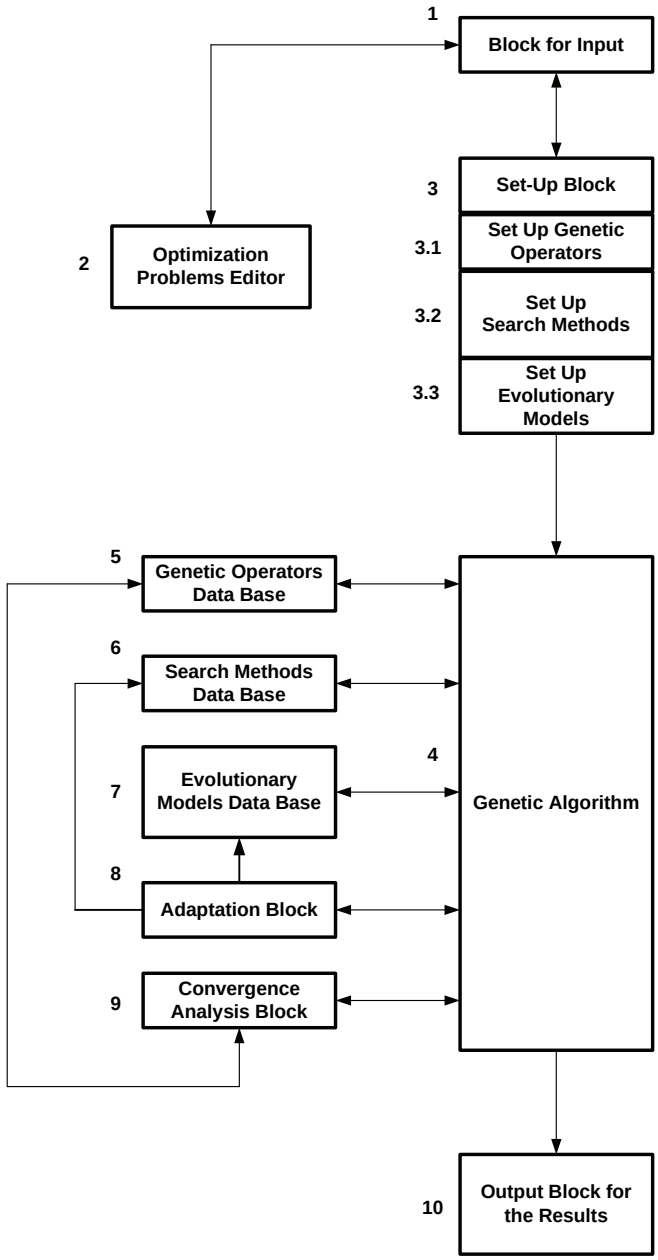


Fig.1

The introduced module is realized based on GA-s and it is used to compare original properties of one and the same GA for different genetic operators. The result of this comparison includes the operative time of the algorithm, the best solution from the execution, the algorithm convergency estimate. The general structural graph of the program module includes block 1, block 2, block 3.1, block 4, block 5, block 10 (Fig. 2).

Block 1 is the block for the data input about the problem (input of variables and parameters of the investigated block, combination matrices, various criteria, etc.);

The input of the set-up data for the algorithm is done in block 3.1 (population size, types of applied operators, probability for their usage, etc.) and the data base (DB) for various types of genetic operators is realized in block 5.

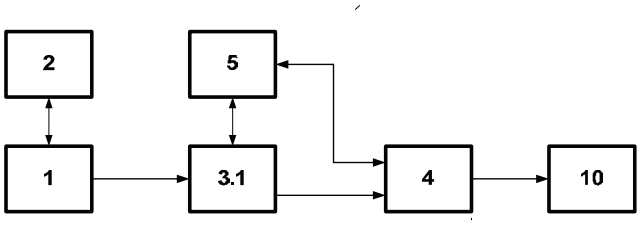


Fig. 2

Different set-ups of GA are performed in block 4 and the results are obtained in block 10.

MODULE 2 is applied to explore the enabled heuristic search methods. It is realized analogously to MODULE 1 and it consists of blocks with numbers 1, 2, 3.2, 4, 5, 6, 10 (Fig. 3).

Blocks with numbers 1 and 2 operate in the already known way, with the introduced functions.

Block 3.2.inputs the set-up parameters for the algorithm (the search automation functions included) so that various methods to improve the search process can be enabled or disabled.

Block 4 is oriented to GA with its set-up about the search optimization to produce information for the current value of the best FF and also the graph of the solution modifications along all generations of the algorithm.

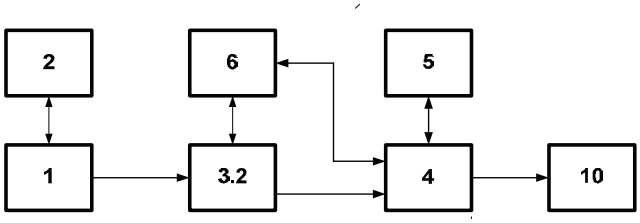


Fig. 3

Block 6 is related to the data base with different search methods.

During the process of operation block 10 produces an output file with the code of all chromosomes, all iterations of the current generation and the result about the global extremum.

MODULE 3 concerns the research of the applied in the algorithm evolutionary models. It includes blocks with numbers 1, 2, 3.3, 4, 5, 6, 7, 10 (Fig. 4).

More specific are the listed below blocks:

Block 3.3 – to input the set-up parameters for the algorithm (they enable the search optimization functions). The block enables or disables various methods to improve the search process and also various evolutionary models;

Block 4 is oriented to GA with its set-up about the search optimization to produce information for the current value of the best fitness function (FF) and also the graph of the solution modifications along all generations of the algorithm.

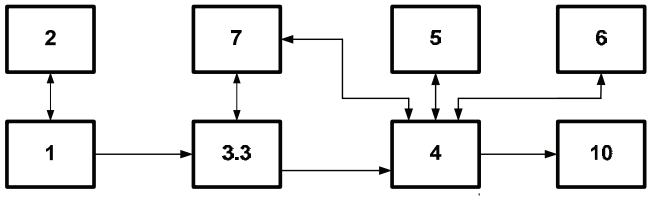


Fig. 4

Block 7 is the Evolutionary Models Data Base block.

Block 10 is the Output Bloc for the Results. The output data are the same as the features of MODULE 2.

MODULE 4 is used during the research process of the applied in the algorithm search methods, evolutionary models, the iterative and statistical improvement, the adaptation and the analysis of convergence. The realization of this scheme is analogous to the rest of the MODULE-s and it contains blocks with numbers from 1 to 10:

- Input of the task data and of the algorithm set-up parameters (the search optimization included) – block 1;
- Optimization task editing– block 2;
- Set-up of genetic operators, search methods and evolutionary methods – block 3 (3.1, 3.2, 3.3 included);
- GA with all set-ups – block 4;
- Three types of DB-s (genetic operators DB, search methods DB and evolutionary models DB) are realized as blocks with numbers 5, 6, 7;
- The feedbacks and the balance between the procedures are achieved in block 8 for adaptation;
- The analysis and the convergence are realized in block 9;
- The obtained results are recorded in block 10.

The execution of a given program MODULE of some subsystem begins with the input of information about the performed task(s). Then a set-up is performed of the global and the private parameters of the algorithm, namely:

- the initial distribution of the coefficients and the criteria;
- the initial population size;
- the upper limit of the iteration number for the algorithm (the number of steps);
- the upper limit of the number of generations (the number of algorithm (re)starts);
- the limits (upper and lower) for the FF values (the global optimum, if neither is given);
- the probability for applying the genetic operators.

Private parameters for the MODULE 1 are various types of operators:

- for crossover – single dot, two dots, cyclic, universal, etc.;
- for mutation – single dot, two dots;
- for translocations;
- for segregation;
- for selection – randomly, according to a given criterion, elite, tournament, etc.

The execution of all established genetic operators is followed by a check of the FF. The research process restarts iff the FF value is unsatisfactory.

The search parameters about the DB methods (MODULE 2, block 6) are enable flags to use heuristics based on:

- optimization statistical methods;
- gradient methods;
- dichotomy methods;
- Fibonacci methods;
- golden-section methods;
- fractal sets, etc.

GA is executed after the performance of all established search methods. If the stop criterion is reached then the algorithm finishes (the execution), else the research process restarts.

The MODULE 3 Evolutionary Models Data Base program block 7 parameters are as it is listed below:

- Darwin-evolution model;
- *Lamarckism* – Lamarck-evolution model;
- *saltationism* – de Vries model of evolution;
- K.Popper's model of evolution;
- synthetic theory of evolution.

The GA is realized via the best operator set that is determined during the test. The cited above elaborated methods are applied during the execution time of the algorithm. The main idea is representation of the general improvement embedded by the heuristics during the search which includes their behavior for various types of problems.

This is followed by the execution of GA according to the result about the operator set from the execution of all heuristic procedures and all evolutionary models included. The operative set-up influences the Adaptation Block no. 8 that realizes the balance between the procedures and the Convergence Analysis Block no. 9 to overcome the algorithm convergence. This is the way to achieve the best features at execution time and also raising the algorithm robustness.

III. CONCLUSIONS

The basic idea of the performed testing was to receive better features at run time for the set of algorithms, and an increase in their stability. The determination of the improvement parameters includes approaches, algorithms, software modules, DB-s, etc. They are performed by three basic series: testing genetic operators and the Gas, testing methods for searching and testing of the evolutionary models.

The conception of the presented system covers testing of the cited below aspects:

- algorithms for a specific problematic case by a successive enabling of all elaborated heuristics from a standard problem and the determination of the values of the improved parameters. In this way it is possible to estimate the effectiveness of the applied in the algorithm methods to optimize the search.

- algorithms of some standard tests followed by comparing the obtained results with already existing ones for a given benchmark that are calculated by another algorithm(s). So it is possible to estimate the effectiveness of the algorithm and of the program modules.

REFERENCES

[1] C. A. Coello Coello, and G. B. Lamont, *Applications of Multi-Objective Evolutionary Algorithms*, Vol. 1, World Scientific, 2004, 761 pp. Hardcover, ISBN: 981-256-106-4, 2004.

[2] V. V. Emelyanov, V. M. Kureychik, V. V. Kureychik, *Theory and Practice of Evolutionary Modelling*, Moscow. (in Russian), 2003.

[3] J. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control and Artificial Intelligence*, University of Michigan, 1975.

[4] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Inc., 1989.

[5] Handbook of Genetic Algorithms / Edited by Lawrence Davis, New York: Van Nostrand Reinhold, 1991.

[6] E. Falkenauer, *Genetic Algorithms and Grouping Problems*, New York: Wiley, 1998.

[7] Y. Wang, Z. Cai, Y. Zhou and W. Zeng, An Adaptive Tradeoff Model for Constrained Evolutionary Optimization, *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, February 2008, pp. 80-92, 2008.

[8] S. Gustafson and L. Vanneschi, Crossover-Based Tree Distance in Genetic Programming, *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 4, August 2008, pp. 506-524, 2008.

[9] D. Goldberg, *Web Courses*, http://www.engr.uiuc.edu/OCCE, 2000.

[10] H. I. Toshev, S. L. Koynov and Ch. D. Korsemov, Evolutionary Theories and Genetic Algorithms, In: *Proceedings of the XLI International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST 2006*, (Editor Dimitrov, D.), Sofia, Bulgaria, pp. 236-239, 2006..

[11] A. Sóbester, P. B. Nair, and A. J. Keane, Genetic Programming Approaches for Solving Elliptic Partial Differential Equations, *IEEE Transactions on Evolutionary Computation*, vol.12, no. 4, August 2008, pp. 469-478, 2008.

[12] V. Guliashki, H. Toshev and Ch. Korsemov, Survey of Evolutionary Algorithms Used in Multiobjective Optimization, *Problems of Engineering Cybernetics and Robotics*, (Ed. Sgurev V.), Sofia 2009, ISSN: 0204-9848, Vol. 60, 2009, pp. 42-54, 2009.

[13] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Design*, New York: Wiley, 1997.

[14] H. I. Toshev, S. L. Koynov, and Ch. D. Korsemov, , Application of Genetic Algorithms for Supply Management of Multi-range Warehouse In: *Proceedings of the XIX International Scientific Conference on Robotics and Mechatronics RM 2009*, (Editor Zahariev, R.), Varna, Bulgaria, pp. 12-17, 2009.

[15] A. Sóbester, P. B. Nair, and A. J. Keane, Genetic Programming Approaches for Solving Elliptic Partial Differential Equations, *IEEE Transactions on Evolutionary Computation*, vol.12, no. 4, August 2008, pp. 469-478, 2008.

[16] V. Guliashki, H. Linkov „An Accelerated Genetic Reconstruction Algorithm for Inverse EIT Problems", XV International Symposium on Theoretical Electrical Engineering ISTET 2009, 22-24 June 2009, Lübeck, Germany, ISSN: 0932-6022, pp. 379-382, 2009.

[17] P.V. Afonin, System for Rational Cutting Out of Materials Applying Genetic Optimization Algorithms, *4th International Summer Workshop in Artificial Intelligence for Students and Ph.D. Students*, Proceedings, Minsk, 2000, pp. 125-128 (in Russian), 2000.

[18] M. Nicolini, A two-level evolutionary approach to multi-criterion optimization of water supply systems, In Carlos A. Coello Coello et al, editor, *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, pp. 736-751, Guanajuato, Mexico, Springer, Lecture Notes in Computer Science Vol. 3410, Mar. 2005.

[19] E. Mininno, F. Cupertino and D. Naso, Real-Valued Ifmpact Genetic Algorithms for Embedded Microcontroller Optimization, *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 2, April 2008, pp. 203-219, 2008.

[20] C. Mattiussi and D. Floreano, Analog Genetic Encoding for the Evolution of Circuits and Networks, *IEEE Transactions on Evolutionary Computation*, vol. 11, no.5, October 2007, pp. 596-607, 2007.

[21] F. Schlottmann and D. Seese. Financial applications of multi-objective evolutionary algorithms: Recent developments and future research directions. In Carlos A. Coello Coello and Gary B. Lamont, editors. *Applications of Multi-Objective Evolutionary Algorithms*, pp. 627-652. World Scientific, Singapore, 2004.

[22] A. Malossini, E. Blanzieri, and T. Calarco, Quantum Genetic Optimization, *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 2, April 2008, pp. 231-141, 2008.

[23] N. Wagner, Z. Michalewicz, M. Khouja, and R. R. McGregor, Time Series Forecasting for Dynamic Environments: The DyFor Genetic Program Model, *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 4, August 2007, pp. 433-452, 2007.

Ontology-Based Deep Web Search For E-Science[1]

Tatyana I. Ivanova[2]

Abstract: **This paper makes a brief exploration of Deep Web search technologies and proposes a new semantic ontology-based approach for personalized searching scientific publications in digital libraries, books in web catalogs of scientific-content books, and other scientific data in web databases. Our main aim is to investigate main deep web search tools and digital libraries and in the base of them develop a conceptual model of personalized searching tool for scientists.**

Keywords: Deep Web Search, Personalized Search, Searching Web databases, Ontology-based Search, Searching digital libraries

I. INTRODUCTION

User queries on the Web can be classified into three types according to user's intention [3]: informational query (The intent is to acquire some information assumed to be present on one or more web pages), navigational query (The immediate intent is to reach a particular site) and transactional query [2] (The intent is to perform some web-mediated activity, as downloading or purchasing). General search engines usually don't recognize the user intent and disregarding the result type, return the mixed result list. Sometimes, it is difficult to make a strict classification of user queries according to his intent. For example, searching the digital library for scientific information is informational query, but downloading the chosen paper is transactional operation.

Web search engines can't index most of the possible pages, that can be returned by the dynamic web sites, or data, stored in Web databases (so called Deep Web[6]) and it is difficult to find such information if (the location of) source site is unknown. Google scholar for example is very useful for searching free scientific publications, but it has indexed only a little part of all of the scientific papers, published in the Web. It is very important for scientists to be able to find easy all the new research papers, related to his subject, purchase new issues of valuable books or download needed software. Another drawback of Web search engines is that during searching or ranking results they do not take into account personal user preferences or interests. Federated search tools help users to identify the databases that are best suited to the subjects they are searching. It allows users to search across multiple resources: subscription databases, library catalogs, and other types web databases.

[1] The research, presented in this publication is funded by the Internal Research Project 102ни013-10 Research and Development sector at Technical University of Sofia for 2010

[2]Tatyana I. is from the Technical University of Sofia, Bulgaria , E-mail: tiv72@abv.bg

In this paper a new semantic ontology-based approach for personalized searching scientific publications in digital libraries and web catalogs of scientific-content books is proposed. Our aim is to develop a conceptual model and as a future research, a tool for personalized searching of scientific publications in digital libraries and scientific books in web catalogs (for purchasing). As such resources are stored in full text databases or web catalogs, and are intended for users with specific research interests, we have to made research about and develop a specialized personalized Deep Web search tool. It will be used as part of the virtual scientific laboratory to facilitate the search for scientific publications, books, or specific information in Internet databases relating to scientific research.

The paper is organized as follows: Section 2 discusses earlier research in Deep Web searching; Section 3 proposes a new semantic ontology-based approach for searching scientific objects; Section 4 discuss the expected problems, strengths and drawbacks of proposed approach and it's further realization; Section 5 concludes the article.

II. DEEP WEB SEARCH STATE OF THE ART

Most of Search engines rely on programs known as crawlers (or spiders) that gather information by following the trails of hyperlinks that tie the Web together. Traditional search engines [5] such as Google, or Yahoo can be searched, retrieved and accessed only sources that have been indexed by the search engine's crawler technology. That approach works well for the pages that make up the surface Web, but for online databases that are set up to respond to grand amount of typed queries it is practically impossible to index all possible responses. The large volumes of documents that compose the Deep Web are not open to traditional Internet search engines because of limitations in crawler technology.

There are two main approaches for Deep Web search: searching previously harvested metadata (in search engine indexes, as in surface web), and federated search. Deep Web indexing methods are very different from those in surface web, as all indexing process is based on automatically querying and retrieving data behind web database search interfaces. Search engines, indexing deep web content (as Google, or Yahoo) use specific deep web crawlers. They detect the index features by issuing probe queries against the search and build a sample of the queried database by issuing a set of queries. Next, they select the most frequent words in the documents in samples to crawl the database, assuming they also have a high frequency in the actual database/index. Interface.

Federated search makes deep web documents in databases searchable by sending queries directly to native search interfaces of these databases. Additionally, federated search provides a singular search interface to numerous underlying

deep web data sources. Federated search is the technology of simultaneously searching multiple content sources from one search form and aggregating the results into a single results page. This reduces the burden on the search patron by not requiring knowledge of each individual search interface or even knowledge of the existence of the individual data sources being searched. Federated search process consists of four phases: 1. transforming a query and broadcasting it to a group of disparate databases or other web resources, with the appropriate syntax; 2. merging the results collected from the databases; 3. presenting them in a unified format with minimal duplication; 4. providing a means, performed either automatically or by the portal user, to sort or cluster the merged result set. Federated search is a type metasearch. We can build our own metasearch engines for federated search, using database or other federated search engine interfaces.

For our purposes, it is important to build our metasearch personalized tool for searching scientific digital libraries, e-commerce book catalogs and some type specific scientific databases.

A digital library is a library in which collections are stored in digital formats and accessible by computers. Many academic and government organizations provide libraries, some of which are actively involved in building institutional repositories of the institution's books, papers, theses, and other works which can be digitized. Many of these repositories are made available to the general public with few restrictions, in accordance with the goals of open access, in contrast to the publication of research in commercial journals, where the publishers often limit access rights.

Digital libraries frequently use the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to expose their metadata to other digital libraries, and search engines like Google Scholar, Yahoo! and Scirus. The OAI-PMH compliant digital repositories in the world may be find on [15]. Metadata according to this protocol are represented in XML format. The main drawback of this representation is the lack of explicit formal semantic. As digital libraries are Deep Web recourses (mainly textual databases), there are three general strategies for searching digital libraries: Searching, using the library search interface, Federated search and Searching previously harvested metadata.

Almost every one digital library proposes internal searching tools. It provides full access to all stored in the library resources and make full use of specific library metadata. The main disadvantage of this approach is the need for the user to know in which library to search, as there are many different metadata standards [1]. There are a lot of digital libraries and choosing the best one for concrete search is a problem.

For building effective federated search engines, the knowledge of internal architecture and metadata standard of used libraries is needed. For example, DSpace architecture has three layers and two APIs : Storage layer to store digital objects and their metadata in databases and file systems; Business logic layer for key operations such as searching and browsing services; and Application layer for users to access DL system through its user interface; Networked Digital Library of Theses and Dissertations (NDLTD) is based on Federated Architecture, with MARIAN as a mediation middleware; CiteSeer uses a Service-Oriented Architecture (SOA), Open Digital Libraries uses Component-based DL architecture. Various architectures and metadata standards are the main sourse of problems in federated search engine building.

Federated search engines searching in digital libraries perform typical vertical search, as most of libraries contain resourses, related to one or few domains. For example, CiteSeer.IST[9] search engine (and digital library) search information within scientific literature, Scopus finds academic information among science, technology, medicine, and social science categories, GoPubMed searches for biomedical articles in PubMed, PubFocus searches Medline and PubMed for information on articles, authors, and publishing trends, Scitation search over one million scientific papers from journals, conferences, magazines, and other sources, Scirus moves beyond journal articles and also includes searches among such resources as scientists' webpages, courseware, patents, and more, Sesat is an open sourced Search Middleware with federation capabilities and a built-in search portal framework, CompletePlanet uses a query based engine to index 70,000+ deep Web databases and surface Web sites, WorldWideScience is composed of more than 40 information sources, several of which are federated search portals themselves. One such portal is Science.gov which itself federates more than 30 information sources representing most of the Federal government articles. This approach of cascaded federated search enables large number of information sources to be searched via a single query. For effective searching user have to have some knowledge about digital library search engines, mainly which libraries they search, papers, related to which domains store corresponding libraries and what metadata is important in searching. List of important academic databases and search engines can be found in [4].

Big search engines as Google, Yahoo, or Bing index a nearly every web site (web developers take care of this by complying with search engine optimization rules), and one may rely on them for finding emerging digital libraries before choosing the best tool for search them. Strength of general purpose search engines, having deep web searching capabilities is that they (for example Google Scholar) can offer many of freely available in the internet scientific publications in nearly every domain.

There are three main ways to search web catalogs: direct usage of building search engines, using general purpose or deep web (e-commerce) meta search engines, or making own federated search engine to search in many (directly chosen from the user) catalogs simultaneously.

For efficient direct usage of building search engines user can be previously informed about type and coverage of catalog content and corresponding search engine capabilities (accuracy, relevancy of returned results; misspelling correction capabilities; ability of searching and sorting according to different criteria, as price, brand, availability; ability in finding related words and common synonyms for terms; helping in query formulation).

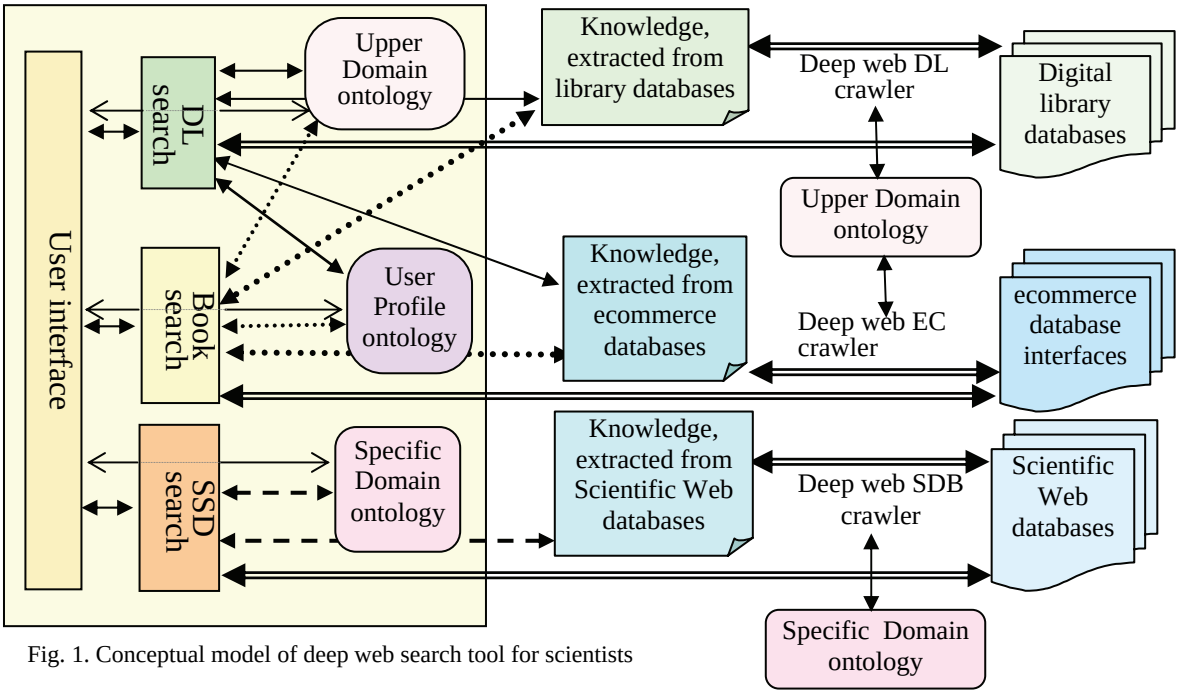


Fig. 1. Conceptual model of deep web search tool for scientists

Ecommerce Meta Search Engines (EMSEs) [8] provide a unified access to multiple ESEs allowing users to search and compare products from multiple sites with ease. A number of EMSEs have been created for some application areas on the Web. For example, addall.com searches books from multiple book sites.

Search engine technology had to grow up dramatically during the last few years [7], in order to keep up with the rapidly growing amount of information available on the web. Despite of all the continuous improvement, usually search engines return thousands results, but it is difficult to find those that are needed for the user, or they are not among the returned results. We believe that for easy and successful finding of needed information on the web, the search has to be personalized focused and semantic-based. User has to be advised or supported in full text unambiguous query formulation and choosing the best for his concrete purpose search engines. In the second chapter we propose a conceptual model of semantic ontology-based tool for scientists. It is intended for personalized searching scientific publications in digital libraries, books in web catalogs of scientific-content books, and other scientific data in web databases.

III. Semantic Ontology-Based Approach For Searching Scientific Objects

Recent research in both Data Integration, Semantic Web, or EScience, witness increasing needs for semantically driven data access, and in particular for the so called Ontology Based Data Access (OBDA). The aim of OBDA is to use ontology, i.e. a formal conceptualization of the application domain, to mediate access to data. Ontologies provide a semantic access to domain specific application data and the expression of constraints allow overcoming incompleteness that may be present in the actual data. Our idea is to model the user profile of the scientist and his scientific domain semantically using ontologies to achieve more flexibility in choosing the right search engine, decrease query ambiguity and in such a way increase the precision and recall in searching scientific publications, books or another type of scientific data. We also may rerank the returned (from one or several similar search engines) results according to particular user profile. Conceptual schema of the proposed deep web search tool for scientists is shown on figure 1.

The main differences between our tool and other Deep Web tools are:

– It is specialized for scientific papers, books and scientific data;

– It is intended to search three main Deep Web recourse types and recourse type is explicitly specified in the sending query.

The tool will offer three type services: Searching specific digital libraries for reading or purchasing scientific papers, searching book catalogs for purchasing scientific books, and searching scientific databases for integrated circuits scientific data.

We propose rich user interface, allowing selection of query intention (informational, for reading papers, or transactional, for book purchasing), selection of preferred libraries or repositories, as well as searching, dependent from chosen data or metadata. As the tool use semantic metadata (domain and profile ontologies), we will experiment some query expansion techniques [12], as well as automatic library selection or returned results reranking according to particular user profile.

Searching specific digital libraries everything uses various metadata, related to keywords, bibliographic and other specific to concrete library metadata. We will manually explore widely used technical and scientific digital libraries (for example [16], [17], [18]) and supply user with a tool for

automatic library selection according to the query and his profile information. User also may manually choose a library before sending a query. As a whole, stored in digital libraries paper metadata are not sufficient for efficient searching. We propose on the fly annotation of selected papers before it ranking and recommending to the users. For such annotation, domain ontologies, scientific experiment ontology (EXPO [14]) and paper structure ontology may be used. Annotation will be discussed in another paper. We also may use a Deep Web crawler for extracting metainformation from digital libraries or finding automatically ones, which we don't explore manually.

For searching book catalogs for purchase scientific books we will experiment several Ecommerce Meta Search Engines. Their drawback is that they query many databases (not only these that are preferable for scientists) and syntactic search approach may cause appearing the best for concrete user results very backward in the result list, or even disappearing. That is why we plan using specialized Deep web Electronic catalog semantic crawler to extract specific metadata from electronic catalogs or find emerging catalogs (figure 1).

For searching Web databases, containing scientific data, related to our electronic circuit testing domain, we firstly will exporting database schema as ontology, representing semantic of our data and then will develop and test semantic scientific data search tool, which realize Deep Web crawling for similar databases, metadata extraction and searching. Query-based sampling [10] can be used to discover the content available at a remote database server. Database translation as ontology will discuss in another paper. We expect that access rights to analogous Web databases nay be serious problem for deep web crawling and data extraction.

IV. CONCLUSION

More than a half of Web data are hidden from the surface-web search engines in databases of financial information, shopping catalogs, medical and other research in digital libraries. It is of great importance for scientific research to have easy and continuous access to the latest developments in the scientific area (presented in publications, books, and other scientific resources, usually stored in web databases).

As a result of rapidly growing number of scientific publications and books in electronic catalogs, the search precision and accuracy are becoming more and more important. One of the main trends for improving search quality is increasing the recourse metadata quality by using collaborative or semantic web technologies for metadata extraction, representation, and usage. Another important trend is digital library and web catalog standardization, exporting recourse metadata in mashine-processable format, development of the more and more effective deep web search engines.

In this paper after an analytical survey of deep web tools and approaches, we propose a conceptual model of specialized personalized Deep Web search tool for scientific information, stored as publications in digital libraries or specific databases. It uses ontology-based semantic search approach to improve search quality. It relies on rich collection of metadata, extracted from repositories, or by using methods of direct automatic otology-based annotation of textual resources to propose a flexible user friendly search interface and user query disambiguation capabilities. After analyzing the query and taking into account user profile, domain ontology and explicitly selected from the user options, the tool may reformulate the query and take a decision to which search engine (s) forward it. The tool will be implemented and tested as part of our research project. We plan to experiment dynamic selection of search strategy among several variants: direct forwarding the query to one or more scientific database or e-commerce search engines, manage user feedback and store processed information in user profile ontology for future usage in strategy-selection process.

REFERENCES

[1] Links to metadata standards http://archive.ifla.org/II/metadata.htm

[2] Y. Liz, R Krishnamurthy, S. Vaithyanathan, "Getting Work Done on the Web: Supporting Transactional Queries", http://www.almaden.ibm.com/cs/projects/avatar/sigir06.pdf, 2006

[3] A. Broder, " A taxonomy of web search" , IBM Research http://www.sigir.org/forum/F2002/broder.pdf, 2002

[4] Wikipedia's List of academic databases and search engines, http://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines

[5] Wikipedia's List of search engines, http://en.wikipedia.org/wiki/List_of_search_engines

[6] Deep Web Research Resources and Sites http://deepwebresearch.blogspot.com/

[7] Search Tools News, http://www.searchtools.com/info/database-search.html, 2010

[8] Q. Peng, W. Meng, and Hai He, "WISE-Cluster: Clustering E-Commerce Search Engines Automatically", *WIDM'04*, November 12-13, 2004, Washington, DC, USA, 2004

[9] http://citeseer.ist.psu.edu/

[10] A. S. Tigelaar,, D. Hiemstra, "QueryBased Sampling: Can we do Better than Random?", CTIT Technical Report , 2009

[11] K. C. Chang, B. He, C. Li, M. Patel, and Z. Zhang." Structured databases on the web: Observations and implications", *SIGMOD Record*, 33(3):61–70, Sept. 2004.

[12] M. Shokouhi, L. Azzopardi, P. Thomas, "Effective Query Expansion for Federated Search", SIGIR'09, July 19–23, , Boston, Massachusetts, USA, 2009

[13] A. S. Tigelaar , D. Hiemstra, "Query-Based Sampling: Can we do Better than Random?", CTIT Technical Report, http://wwwhome.cs.utwente.nl/~hiemstra/papers/tr-ctit-10-04.pdf, 2004

[14] L. Soldatova1, R. D. King, " An Ontology of Scientific Experiments", Journal of the Royal Society Interface, December 22; 311): 795–803. , 2006

[15] Web site to OAI-PMH compliant digital repositories in the world, http://www.openarchives.eu/home/home.aspx

[16] WorldWideScience.org

[17] Science.gov

[18] Scitopia.org

RDB to RDF or Ontology Mapping – Approaches, Tools and Problems[1]

Tatyana I. Ivanova[2]

Abstract: **Web based relational databases are secure, reliable, and widely used, but search engines index only a little part of it content. To facilitate the search and collaborative usage of Web – based scientific data, exporting database schemas as ontologies are needed. This paper explores current approaches and tools for relational databases to ontology translation and mapping to find suitable tool for automatic representation of our scientific data in mashine-processable format.**

Keywords: **database to ontology mapping, relational database to ontology transformation tool, OWL, ontology**

I. INTRODUCTION

Most of data on the Web are stored in relational databases and are accessible for humans through Web browsers. Web applications, as crawling-based general search engines like Google are not fully capable of searching them as their contents, known as deep Web or invisible Web, are hidden behind their Web search interfaces and not effectively crawlable. The form-based interface to relational Web databases, used by humans, is not suitable for intelligent agents. If a software agent was directed to use the e-commerce system for example, it would need to interpret the instructions for filling out the form, enter the appropriate parameters on the form, submit the form, and parse the results. All of the above may be difficult if not impossible for a software agent to do, especially if the requirement is to compare results from multiple e-commerce sites, all with different instructions, forms, and returned data formats. Semantic Web technologies and standards have been developing for making all kind of data, available on a web site or web service, accessible and easy to use from both humans and computers. For that purposes, data semantics have to represent by ontologies and the methods for querying ontologies, return semantic web data formats have to be used. However, relational databases are the best known tools for storing, managing and accessing data, as they are reliable, secure and well-working. Therefore, it is important to establish interoperability between relational databases and ontologies.

Working on our scientific project, we will store valuable testing data in relational database and we wish to make them easily accessible from web for other scientists. For our investigations, comparing our data to another scientific research results is of great importance, and we also would like to simplify finding of related data on the web.

[1] The research, presented in this publication is funded by the Internal Research Project 102ни013-10 Research and Development sector at Technical University of Sofia for 2010

[2]Tatyana I. is from the Technical University of Sofia, Bulgaria , E-mail: tiv72@abv.bg

Effective ways to achieve interoperability between databases are finding mappings between relational database schemas and ontologies-mediators to develop semantic database wrappers, or exporting database schemas as ontologies and dynamically mapping these ontologies. A lot of approaches [3], methods and tools [3] for semantic interoperability of relational databases have been developing during the last years. Because of some significant differences between relational and semantic knowledge models and semantic web technologies immaturity, no one of them can guarantee automatic free of errors disambiguous database to ontology export, or high precision and recall in web searching. We will explore, compare approaches and tools, using described them research papers, documentations, or making our own tests (if systems are available for downloading). Our main aim is to choose the best approaches, methods, find suitable open source and free tools that after some customization we may use to expose our database to semantic web and make possible both it usage from other scientists and finding similar databases in the Web.

II. METHODS FOR EXPOSING RELATIONAL DATABASE DATA TO SEMANTIC WEB

There are grand variety of methods for exposing relational database data to semantic web, differing from each other in used models (annotation or translation), languages and additional database manipulation techniques. Some extract the schema from the database and convert it to semantic web format; others use annotations, or wrappers. Used formats for storing semantic data are usually RDF(S) or OWL. Extracted semantic data can be stored together with database schema in new Semantic Repositories, or remain the relational database, but store Semantic Metadata Extractions in a Separate Repository, or Adding Semantic Markup to the Existing data Repository.

In order to annotate database data, it is necessary to assign the meaningful labels to them. Existing automatic data annotation techniques [10] can be divided into three categories: approaches based on Web interface pages information (for example Arlotta presents an automatic annotating approach which basically annotates data units with the closest labels on result pages), interface schema (DeLa uses some heuristic rules to construct some basic annotators, Yiyao Lu utilizes new features that can be automatically obtained from the result page including the content and data types of the data units to align data), and domain ontology.

The ontology annotation-based approach [2] suppose that database owner produces server side web page markup (usually XML) describing the database's information

structures; The searching client after database location use his own client-side ontology, describing the semantic of his domain and his annotator to produce client-side annotations that conform to his ontology and the server-side markup. Then he can send semantic queries to server-side database, using his ontology and mapping rules through the Web-service API. If there is no server side XML markup, describing the database, client may use deep web crawler for sending random queries to a server to obtain a sample of documents of the underlying collection. The sample represents the entire server content. This representation is called a resource description.

OntoMat-Annotizer is a user-friendly interactive webpage annotation tool (may be used to annotate directly database Web interface forms). It supports the user with the task of creating and maintaining ontology-based OWL-markups i.e. creating of OWL-instances, attributes and relationships. It includes an ontology browser for the exploration of the ontology and instances and a HTML browser that will display the annotated parts of the text. It is Java-based and provides a plug-in interface for extensions. It is freely available and can be downloaded from http://annotation.semanticweb.org/ and come with documentation and tutorial for usage.

Annotation-based approach is easy for database-developers, but he never represents unambiguously the semantic of database content and that sometimes leads to extraction of non-relevant data, or strictly limited reasoning capabilities.

Extraction of the database schema and representing it as ontology ensure semantic access to database data, but there are some problems, related to it. First, there are significant differences in RDB, RDF and OWL models. We will discuss them bellow. Second, manual representation is difficult, time consuming and assume knowledge-management skills, whereas automatic methods are far from his maturity. Third, a lot of web databases are accessible only thought HTML forms, what make additional problems. Extraction of metadata from the database schema is a common method used by OntoKnowledge [5].

Difficulties in database schema extraction methods usage depends from the amount of information about database schema. When we develop our own database, we have full access to it data and schema, and if we plan to make it easily accessible from the web for human and we have to make semantic interface to data by automatic or semiautomatic generation of semantic description as ontology. When we will have an access to not semantically described web database, we can learn about it structure and data only by direct querying through it form-based interface.

There are two main difficulties in translating from RDB to OWL: how to capture and represent all implicitly used in database domain knowledge and how to manage with different logical foundations of RDB and OWL.

Data models, such as database or XML-schemes, typically specify the structure and integrity of data sets, and the semantics of data models often constitute an informal agreement between the developers and the users of the data model and which finds its way only in application programs that use the data model. Ontologies, in contrast, should make explicit all the semantic of data model and make him as much generic and task-independent as possible. There are two main sources for acquisition of all this implied semantic during RDB and OWL mapping process – domain knowledge, represented in machine-processable format as ontologies (automatic approach) and human user or expert (manual approach). Other mismatches between RDB and OWL data models that affect a transformation system are related to inheritance modeling, property characteristics, underlined logical systems and open/closed world assumptions [19].

Deep web crawling and Web information extraction are the two main important areas, related to extracting data and metadata through web interfaces from databases for simplifying the database access. A first prototype deep Web crawler, presented to automatically extract and analyze the interface elements and submit queries through these query interfaces was HiWe. Many independent efforts are proposed for keyword query selection techniques for downloading the textual content from Web repositories.

There has been an active research interest in understanding the semantics of the query interfaces of the structured Web databases [4], [12]. WISE-integrator [12], for example, extracts element labels and default value of the elements to automatically identify matching attributes, [4] uses statistical models to find the hidden domain-specific schema by analyzing co-appearance of attribute names.

Three main models and markups are used for storing extracted knowledge: XML, RDF(S) and OWL. We will discuss them from knowledge representation point of view in details separately lather.

The disadvantage of generating an ontology, based on the database schema approach and converting all the database in a new semantic web format is that any other applications that to interface with the legacy database will need to change. In the tangled network of databases in a corporation or other information organization, this option may be too costly and disruptive to contemplate in the near term. Moreover, response time of the knowledge base strictly depends from his logical model and richness and in some cases checking and querying the base may be very slow (OWL full for example is undesidable, which means that in some cases database couldn't respond within a finite time).

The schema of a database can be extracted and converted into a semantic format such as RDF-S. This semantic version of the schema can be mapped to ontology or published via UDDI or WSDL to make the data available to semantic applications. The semantic metadata and mappings can then be stored in a central repository for the purpose of making queries across multiple data sources.

Semantic markup can be provided at the web page or web service accessing the data or on the repository itself. This approach is used mainly in deep web annotation methods. If a system uses a high level of semantic encoding, there will be greater richness and precision in the semantics available to capture the relationships between concepts that the logical reasoning of agents requires. *Levels of Semantic Encoding* (from lowest level to highest level) are: XML; XML Schema; RDF; RDF-S; DAML + OIL ; OWL Lite ; OWL DL; OWL Full.

There are several tools available for transforming relational databases to ontologies [21], [22], 23]. There are three main approaches, using in transformation: data-mining based, knowledge-based and rule-based. DataGenie [24] is rule-based Protégé's plug-in that is capable of importing data from a relational database and representing it in ontology. This import is simple: each table maps to a class, each column maps to a data type property and each row maps to an instance. The drawback of this simplicity is that DataGenie and similar tools fail to discover inheritance, restrictions, symmetric and transitive properties, object properties and restrictions. It also ignore constraints that capture additional semantics and do not analyze loss of semantics caused by the transformation. The RTAXON learning method combine the most robust rules for exploiting relational schemas with data mining focused on the specific problem of concept hierarchy identification. It is implemented in the RDBToOnto tool, which can be downloading free from [25]. Another similar free java-based tool, RDB2Onto converts selected data from a relational database to a RDF/OWL ontology document based on a defined template. It is intended for ontology population. DB2OWL [14], is another tool for automatic generation of ontologies from database schemas. OntoWrapper [17] exposes external semi-structured data to an ontology repository.

METAmorphoses processor [8] is a tool for the data transformation from a relational database into RDF documents. It is implemented in Java and is based on the two-layer data transformation model: the mapping layer and template layer. In the mapping layer, a given database schema is mapped into a structure of a given ontology. The template layer uses this mapping and produces RDF documents in the way driven by templates.

III. Discovering Mappings Between Relational Database Schemas And Ontologies

The difference between transformation of relational databases to ontologies and database-to-ontology mapping is that the transformation generates ontology, corresponding to database schema, whereas mapping assumes the existence of both a relational database and ontology and produces a set of correspondences between the two. Two main logical models of semantic data representation are used in the Web: RDF – based (including RDF and RDF Schema) and OWL-based (including OWL Lite, OWL DL, OWL Full, and their extensions).

The RDF data model is a directed labeled graph, which consists of nodes and labeled directed arcs linking pairs of nodes. RDF is more expressive than the relational data model and data represented in RDF can be interpreted, processed and reasoned over by software agents. Two main approaches for mapping generation between RDB and RDF are used: Automatic domain-independent Mapping Generation, and Domain Semantics driven Mapping Generation.

Automatic Mapping usually generate mappings between RDB and RDF with RDB table as a RDF class node and the RDB column or relation names as RDF predicates Even though these automatically generated mappings often do not capture complex domain semantics that are required by many applications, these mappings can serve as a useful starting point to create more customized, domain specific mappings, or enable Semantic Web applications to query RDB sources.

The Domain Semantics driven Mapping Generation approach generates mappings from RDB to RDF by incorporating domain semantics that is often implicit or not captured at all in the RDB schema. The explicit modeling of domain semantics, often modeled as domain ontology, enables software applications to take in mind valuable facts or relations, concerning data, that users implicitly assume working with database. There are freely available ontologies in the internet (such as the National Center for Biomedical Ontologies (NCBO) at http://bioportal.bioontology.org/; Gene Ontology GO and so. on.) in almost all domains, that may by used (usually after customization).

The mappings between RDB and RDF may be represented as XPath rules in a XSLT stylesheet, in a XML-based declarative language such as R2O [5], D2RQ [16], D2R MAP[1] or as "quad patterns" defined in Virtuoso's [6] metaschema language. The mappings, especially if they are created by domain experts or reference domain ontology, may have wider applicability.

Mapping of RDB to RDF may be either a static Extract Transform Load (ETL) implementation (called "RDF dump"), and implemented in almost all such tools, or a query-driven dynamic implementation. The dynamic approach, (for example in D2RQ, or Virtuoso systems) implements the mapping dynamically in response to a query.

Tools from the OntoKnowledge project [17] and KAON project [9] can be used for mapping a database schema to an existing ontology or generating an ontology based on the database schema.

Virtuoso RDF View [6] uses the unique identifier of a record (primary key) as the RDF object, the column of a table as RDF predicate and the column value as the RDF subject in the mapping process. Other similar tools are D2RQ [16] and SquirrelRDF [Seaborne et al., 2007]. D2RQ platform is freely available and can be downloading from http://sourceforge.net/projects/d2rq-map/. SquirrelRDF provides access to relational databases, by providing a SPARQL interface to a non-RDF store by extending the basic ARQ - query engine for Jena. This approach ensures a full SPARQL implementation over the foreign data source. SquirrelRDF is freely available and can be downloading from http://sourceforge.net/projects/jena/files/. Triplify [18] is an approach to publish RDF and Linked Data from relational databases. It transforms the resulting relations into RDF statements and publishes the data on the Web in RDF serializations, as Linked Data. Triplify can be easily integrated and deployed with Web applications. It is complemented by a library of configurations for common relational schemata and a REST enabled data source registry.

Creating mappings between database schema and Web ontology is a preconditioning process in the generation of ontological annotations for dynamic Web page contents extracted from the database.

In OWL, a class can be mapped to a relational table. Properties of a class can be mapped to the attributes of a relational table. Inheritance (*subClassOf*) relation between classes can be realized by the foreign key (acting as a primary

key) between relational tables, and foreign key, disjoint with primary key can be mapped to object property. Declarative Languages as D2R MAP may be used to describe mappings between relational database schemata and OWL ontologies, or mappings may be stored as pare of initial ontology [15]. Two main approaches may be used to discover semantic mappings: statistical and knowledge-based.

The mapping process, based on statistical approach [13] starts with a relational schema and an ontology, constructs virtual documents for the entities in the relational schema and the ontology to capture their implicit semantic information, discovers simple mappings between entities by calculating the confidence measures between virtual documents via the TF/IDF model, uses mappings between relations and classes to validate the consistency of mappings between attributes, and properties a set of simple discovered mappings.

Knowledge-based approaches use knowledge sources as WordNet or previously developed domain ontologies for extracting shared concepts between RDB and ontology. These approaches are semiautomatic, or complementary to rule-based and statistical, as the quality of knowledge processing is relatively low.

IV. Discussion And Conclussions

As shown above, we have to expose the schema of our scientific data, using XML – based syntaxes for easy usage from the Web. We may do this, using two different approaches: expose the whole database schema or only propose annotations. Making annotations is easy and they can be easily used from the Web software, but such type of data representation lack of formal semantic and natural language ambiguity can become an obstacle to the proper use of data.

Database schema representation as ontology provides both metadata, related to our data, and formal semantic, and make possible for software agents reasoning about the semantic of the data. Moreover, this representation will be used when we search related to our data in the Web. We will map our ontology concepts to metadata or concepts, representing considered Web databases. Only database schema representation ontology may not be sufficient for performing this mapping process and we will expose additionally our domain ontology (for handling synonymy, one to many domain relationships…) and local context ontology (for explicit representation of all the domain knowledge, implicitly implied, but not explicitly represented in the database), used in the process of automatic building of database ontology. Moreover, queries using semantic web query languages can be imposed to our database through its connection to ontology.

RDBToOnto is a free open source tool for automatically generation of fine-tuned ontologies from relational databases. We plan to use it (may be after some customization) for automatic exporting our database schema to ontology.

Using Protégé, we were able to map ontology instances into relational databases and retrieve results by semantic web query languages. The key idea is that, instead of storing instances along with the ontology terminology, we can keep them stored in a database and maintain a link to the dataset. VisAVis is a open source java-based Protégé plug-in for mapping ontologies to databases, can be download from [15] VisAVis maps the relational database contents to the TBox of the ontology. We plan to use it (may be after some customization) for mapping of external database data to our ontology in the process of searching the related to ours data in the Web.

V. References

[1] C. Bizer, “D2R MAP – A Database to RDF Mapping Language”, In Proceedings 12th International WWW Conference, 2003

[2] S. Handschuh and R. Volz , “Annotation for the Deep Web”, EEE Intelligent Systems archive. Vol. 18 , Issue 5 , 2003)

[3] S.S.Sahoo et al., “A Survey of Current Approaches for Mapping of Relational Databases to RDF”,W3C RDB2RDF Incubator Group, 2009

[4] Z. Zhang, B. He, K. C. Chang, “Understanding Web Query Interfaces: Best-Effort Parsing with Hidden Syntax”,SIGMOD, 2004.

[5] J. Barrasa, A. GómezPérez,, “Upgrading relational legacy data to the semantic web”, In Proc. of 15th international conference on World Wide Web Conference (WWW 2006), pages 1069-1070

[6] C. Blakeley, “RDF Views of SQL Data (Declarative SQL Schema to RDF Mapping)”, OpenLink Software, 2007.

[7] http://www.searchtools.com/info/database-search.html

[8] http://metamorphoses.sourceforge.net/METAmorphoses_processor/

[9] Karlsruhe Ontology Project (KAON). Online. Internet. 2/15/2005. Available at:http://www.KAON.SemanticWeb.org.

[10] C. Xiao-Jun, P. Zhi-Yong, W. Hui, "Multi-source Automatic Annotation for Deep Web," csse, vol. 4, pp.659-662, 2008

[11] S. Handschuh, R. Volz, S.Staab, “Annotation for the Deep Web”, IEEE Intelligent Systems, September/October 2003.

[12] H. He, W. Meng, “WISE-Integrator: A System for Extracting and Integrating Complex Web Search Interfaces of the Deep Web”, VLDB'03, pp.357-368, Berlin, Germany, 2003

[13] W. Hu and Y. Qu, “ Discovering Simple Mappings Between Relational Database Schemas and Ontologies”, 2007, http://dit.unitn.it/~p2p/RelatedWork/Matching/Discovering.pdf

[14] N. Cullot, R. Ghawi, and K.Yétongnon, “DB2OWL: A Tool for Automatic Database-to-Ontology Mapping”, CiteSeerX, 2008

[15] N. Konstantinou, et al. “VisAVis: An Approach to an Intermediate Layerbetween Ontologies and Relational Database Contents http://www.cn.ntua.gr/~nkons/essays_en.html#t

[16] C. Bizer, R.Cyganiak, “D2RQ — Lessons Learned”, W3C Workshop on RDF Access to Relational Databases, 2007.

[17] “The OntoKnowledge Toolset,” Online. Internet., 2004. http://www.ontoknowledge.org/tools/toolrep.shtml.

[18] S. Auer et al., “Triplify Lightweight Linked Data Publication from Relational Database”, WWW 2009, Madrid, Spain

[19] SS Bhowmick, J. Küng, and R. Wagner, “Translating SQL Applications to the Semantic Web” LNCS 5181, pp. 450–464, 2008.

[20]Ontostudio web site http://www.ontoprise.de/en/home/products/ontostudio

[21]. M. Li, X. Du, S.Wang, “Learning Ontology from Relational Database”. ICMLC, Vol. 6, 2005

[22]. G.Shen, et al.,. “Research on the Rules of Mapping from Relational Model to OWL”. Workshop on OWL: Experiences and Directions.Vol. 216 (2006)

[23] I.Astrova, A. Kalja, “Towards the Semantic Web: Extracting OWL Ontologies from SQL Relational Schemata” IADIS International Conference WWW/Internet(2006) 62–66

[24] DataGenie: http://protege.cim3.net/cgi-bin/wiki.pl?DataGenie, 2007

[25] RDBToOnto download page, http://www.tao-project.eu/researchanddevelopment/demosanddownloads/RDBToOnto.

Software Cost Estimation - a Practical Approach

Violeta T. Bozhikova[1]

***Abstract* – Software cost estimation is considered as one of the most challenging tasks in software project management. The process of software estimation includes estimating the size of the future software product, estimating the effort required, estimating the duration of the project and finally – the people required. This paper gives an overview of the most powerful cost estimation models, discusses their advantages and weakness and finally a hybrid cost estimation approach that combines their strengths is recommended**

***Keywords* – Software Cost Estimation, Software Cost Estimation Methods, Software Cost Estimation Tools.**

I. INTRODUCTION

Software cost estimation [1-5] is a continuing activity which starts at stage of the project proposal and continues through the overall life time of the software project. The goal of this continual cost estimation is to ensure that the expenses will not exceed the budget provided.

Considerable research has focused on development and evaluation of universal software cost estimation models and tools suitable for all software projects. After 20 years research, we could claim that there are many software cost estimation methods available, but no one method is suitable for all software projects. In fact, their strengths and weaknesses are often complimentary to each other. To understand their strengths and weaknesses is very important for the software estimators. The estimators are increasingly convinced that accurate software estimation is impossible using a single method and increasingly believe that a combination of methods will allow a more accurate and reliable software cost estimate.

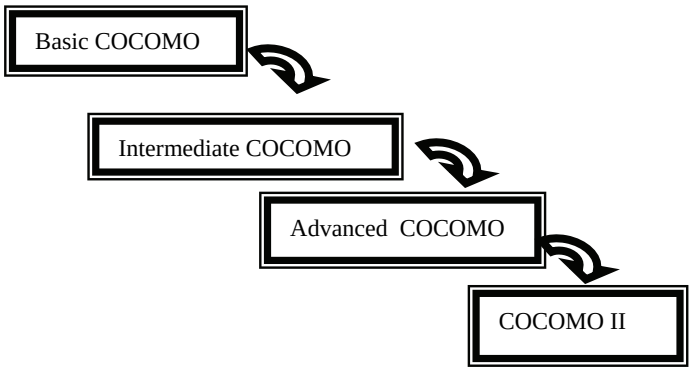
This paper gives an overview of COCOMO hierarchy and Function Points cost estimation models, discusses their advantages and disadvantages and finally a practical cost estimation approach that combines their strengths is recommended as a way for efficient cost estimation.

II. COCOMO MODELS AND FUNCTION POINT ANALYSIS

A. COCOMO Family

One of the most commonly used software cost estimation methods are the **CO**nstructive **CO**st **MO**dels (COCOMO models). These methods for software estimation are considered as algorithmic because provide mathematical equations to perform software estimation. The COCOMO mathematical equations are based on extensive historical research and use inputs such as Source Lines of Code (SLOC), number of functions to perform, and other cost drivers such as language cost drivers, design methodology, skill-levels, risk assessments, etc. As algorithmic methods the COCOMO models have a lot of advantages. The most important are the objectivity, stability and the sensitivity of the results produced. Using such models the estimator gets repeatable results. In the same time, it is easy to modify input data, refine and customize formulas. The general disadvantage of these models is the strongly dependence of the estimations on the inputs. Some inputs can not be easily quantified. As a result, poor sizing inputs or|and inaccurate cost driver rating will result in inaccurate estimation

Basic COCOMO [1] is the first from the family COCOMO models. It is designed by Barry W. Boehm as a model for estimating effort, cost, and schedule for software projects in 1981. Now, a hierarchy of COCOMO models is available:



Basic COCOMO model computes software effort applied "PM" (development effort i.e. development cost) in "person-months" as a function of program size expressed in estimated thousands lines of code KLOC. Person month is the amount of time one person spends working on the software development project for one month. This number is exclusive of holidays and vacations but accounts the weekends. The basic Cocomo equations are:

$$PM = a_b (KLOC)^{b_b} [person - months]$$

$$TDEF = c_b (PM)^{SchedExp} [months]$$

$$Average\ Staffing = PM / TDEF\ [people]$$

The coefficients a_b, b_b, c_b and SchedExp depend of the type of the project (organic, semi-detached or embedded) and are given in the next table:

Software Project	a_b	b_b	c_b	SchedExp
Organic	2.4	1.05	2.5	0.38
Semi-Detached	3.0	1.12	2.5	0.35
Embedded	3.6	1.20	2.5	0.32

[1]Violeta T. Bozhikova is with the Faculty of Computing and Automation, Technical University of Varna, 9000 Varna, Bulgaria, E-mail: vbojikova2000@yahoo.com

TDEF is the Development Time in chronological months and Average Staffing is the People required for the whole project development This модел is good for quick, early and rough estimates of software costs, but its accuracy is limited because it doesn't account the influence of a number of well known factors such as hardware constraints, personnel quality and experience and so on that have a significant influence on software costs.

Intermediate COCOMO is an extension of the Basic COCOMO. This model computes software developement effort PM as a finction of program size and set of "cost drivers" that include subjective assessements of 15 cost driver attributes that are grouped into 4 major categories "Product attributes", "Hardware attributes", "Personnel attributes", "Project attributes". Each of the 15 attributes is rated on a 6-point scale that ranges from "very low" to "extra high" (in importance or value). Effort adjustment factor (EAF) for a given project is calculated as the product of the fifteen effort ratings (EM_i, i=1…15). Typical values for EAF range from 0.9 to 1.4. The Intermediate Cocomo formula for PM now takes the form:

$$PM = EF\,(KLOC)^{ee}\,EAF\,[person - months]$$

Where:

$$EAF = \prod_{i=1}^{15} EM_i$$

Where the coefficient EF and the exponent ee are given in the following table:

Software project	EF	ee
Organic	3.2	1.05
Semi-detached	3.0	1.12
Embedded	2.8	1.20

The Development time (TDEF**)** and People required (Average Staffing) are calculated from PM in the same way as with Basic COCOMO.

Advanced COCOMO can be seen is an extension of the Intermediate COCOMO version. It calculates PM the same way as Intermediate COCOMO but with an assessment of the cost driver's impact on each stage (analysis, design, etc.) of the software engineering process.

The development of the new **COCOMO II** model by the Boehm's team is based on a study of about sixty projects at TRW (a Californian automotive and IT company) in 2002 and is the latest major extension to the original COCOMO. This model is turned to the newer software paradigms (for example OOP) and the modern software life cycles. For comparison, the previous COCOMO models have been very successful for projects up to 100000 lines of code, based mostly on the waterfall model of software development and for programming languages ranging from assembly to PL/I. In addition, the previous COCOMO versions were defined in terms of estimated lines of code LOC (and thousands of LOC, i.e. KLOC). The COCOMO II model bases the calculation of required effort PM on the software project's size measured in SLOC (and thousands of SLOC, i.e. KSLOC). The difference between LOC and SLOC (single Source Line of Code) is that a SLOC may include several physical lines. Each structured construction, for example the "if-then-else" statement would be counted as one SLOC. For comparison, in basic COCOMO model this statement might be counted as several LOC.

The first equation below (PM_{nom}) is the base model for the Early Design and Post-Architecture cost estimation of the software project. The inputs are the Size of software development in KSLOC, a constant A and a scale factor – B [3]. The size is in KLOCS is derived from estimating the size of software modules that will constitute the application program. It can also be estimated from unadjusted function points (UFP), converted to SLOC then divided by one thousand. The scale (or exponential) factor B derived from five scale drivers, such as Team Cohesiveness factor, Process maturity factor, Precedentness, Flexibility and Breakage factor and accounts for the relative economies or diseconomies of scale encountered for software projects of different sizes [3]. The constant A depends on the size of the project. The nominal effort PM_{nom} and the adjusted effort $PM_{adjasted}$ calculations for a given size project and expressed as person months are presented by the next equations:

$$PM_{nom} = A(KLOC)^{B}\,[person - months]$$

$$PM_{adjasted} = PM_{nom}\,(EAF)\,[person - months]$$

Where:

$$EAF = \prod_{i=1}^{17} EM_i$$

COCOMO II has 17 cost drivers attributes (Analyst Capability, Applications Experience, Programmer Capability, Use of Software Tools, Multisite Development, Required Development Schedule, Required Software Reliability, Database size, Product complexity, Personnel Experience, Language and Tool Experience, Personnel Continuity, Execution Time Constraint, Main Storage Constraint, Platform Volatility, Required Reusability, Documentation match to life-cycle needs) which rating (expressed as a number EM_i, i=1…17) the estimator has to determine with the goal to calculate the value of effort required EAF .

B.Function Point Analysis

Although counting lines of code is the first and most common software sizing methodology this sizing method is no longer practical due to the great advancements in software engineering and modern programming languages. Another commonly used sizing method is the IFPUG method [5] called Function Point Analysis (FPA). It is another method of quantifying the size in terms of the functions that the system delivers to the user. The function point measurement method was developed by A. Albrecht at IBM in 1979. The main advantages of function point analysis based model are:

- function points (FP) can be estimated from requirements specifications or design specifications, so using FPA it possible to estimate development cost in the early phases of development.

- function points are independent of the programming language or the methodologies used for software implementation.
- since function points are based on the system user's external view of the system the non-qualified users have a better understanding of what function points are measuring

Different variations of Function Points have emerged over the years, such as Object Oriented Function Points, Use Case Function Points and so on. Function point estimation approach is widely used within COCOMO II because COCOMO II is oriented to the newer software paradigms and to the modern software life cycles.

III. A Hybrid Software Cost Estimation Approach

Our approach is a combination between almost all COCOMO models: Basic COCOMO, Intermediate COCOMO, and COCOMO II with Function Point Estimation features. The raison to develop such hybrid approach collecting all these mentioned above methods is to give the estimators an opportunity for a suitable choice of cost estimation model, depending of the concrete project type and the specific and often incomplete initial knowledge about the software product in the early stages of its development.

The two basic steps, required to accomplish software estimation are:

- Estimate product size,
- Estimate effort applied, project's duration and resources needed.

A. Estimate product size

Our approach bases the calculation of required effort PM on the software project's size measured in COCOMO II SLOC (and thousands of SLOC, i.e. KSLOC). The calculation of SLOC (KSLOC) may be based on the expert's estimation of the size of software project (if is possible to make such estimate) or on FP estimation. The usual Function point's estimation procedure is based on information that is available early in the project life cycle. It begins with determining and classifying (by complexity level) the user functions as Inputs, Outputs, Files, Interfaces, and Queries (figure 1.). As a result, the Unadjusted Function Points (ΦT) quantity is calculated (figure 1). Next a Translation of Unadjusted Function Points (ΦT) into SLOC is realized. The unadjusted function points are converted into equivalent SLOC depending of a LangFactor of the language used. For example, the LangFactor [3] for Assembly language is 320SLOC/UFP, for C++ - 29SLOC/UFP, for Fortran 77 – 105SLOC/UFP, for Lisp – 64SLOC/UFP, for Pascal – 91 SLOC/UFP and so on.

$$SLOC = \Phi T \times LangFactor \quad (1)$$

The usual Function Point procedure accounts the degree of influence DI (2) of fourteen application characteristics (figure 2), such as distributed functions, performance, reusability, etc. The ratings of these 14 characteristics (rating scale of 0.0 to 0.05 for each characteristic) are added together, and added to a base level of 0.65 to produce a general characteristics adjustment factor that ranges from 0.65 to 1.35.

$$DI = \sum_{1}^{14} rating_i \quad (2)$$

Our approach has respected this described above usual Function Point procedure to calculate the size of the project. The final equation that is used for cost estimates is shown below:

$$SLOC = \Phi T \times (0.65 + (0.01 \times DI)) \times LangFactor \quad (3)$$

Fig. 1. FP calculation

B. Estimate effort applied, project's duration and resources needed

The general equation that we have used to calculate the effort needed (PM) for a given size project development, expressed as person months is given below:

Where:

$$PM = EF \times EAF \times KSLOC^{ee}[person-months] \quad (4)$$

$$PM = \begin{cases} PM_{nom} & if \quad EAF = 1; \\ PM_{real} & if \quad Effort\ AdjasmentFactor\ is\ caclulated \end{cases}$$

If the effort adjustment factor EAF is 1 (it is its default value) PM is interpreted as the nominal effort PM_{nom} needed for a given size project development, expressed as person months. The values of the coefficient EF and the exponent ee in this case are based on Intermediate COCOMO model.

The calculation of the effort adjustment factor EAF (5) is related with the calculation of the adjusted effort PM_{real}. EAF estimation could be based on the fifteen COCOMO Intermediate cost drivers or on the seventeen COCOMO II Cost Drivers plus one. Total of eighteen Cost Drivers in the latter case are grouped into 3 major categories "Personnel attributes", "Project attributes" and "Product attributes". An additional user defined cost driver, named USER is added to the classic COCOMO II Cost Drivers. It gives estimators an opportunity to recognize the impact of a chosen project-specific factor, other than the provided in COCOMO II.

EAF for a given project is calculated as the product of the effort ratings of these attributes.

$$EAF = \prod_{1}^{CDN} EM_i \quad (5)$$

Where,

$$CDN = \begin{cases} 18 & if \quad COCOMO\,II\ is\ used; \\ 15 & if \quad COCOMO\ Intermediate\ COCOMO\ is\ used. \end{cases}$$

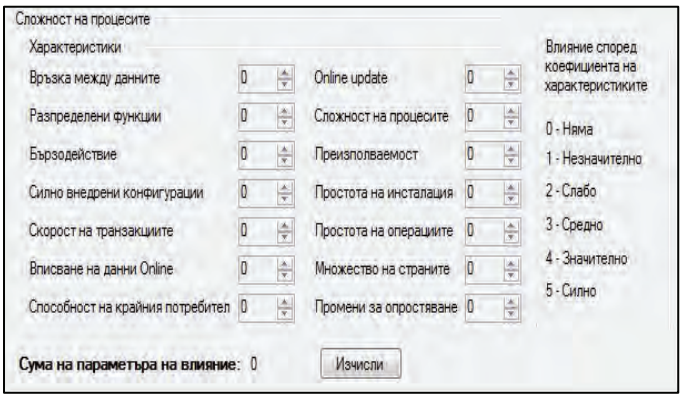


Fig. 2. Application characteristics and DI calculation

The calculation of the duration TDEF of the project is based on the effort predicted by the effort equation:

$$TDEV = EF \times (PM)^{SchedExp}[months] \qquad (6)$$

Where:

PM is the effort (nominal or real) that is calculated, SchedExp is the schedule exponent derived from Basic COCOMO model and EF is a coefficient derived from Intermediate COCOMO model.

The average staffing is calculated as follows:

$$Average\ Staffing = {}^{TDEV}\!/_{PM}\,[people] \qquad (7)$$

IV. Conclusion

This paper gives a comparative overview of COCOMO and FPA models, discussing their advantages and disadvantages and proposes a hybrid cost estimation approach that combines their strengths. Our observation is that an approach that collects all these mentioned above methods gives the estimators an opportunity to choose the appropriate estimating method in a situation of often incomplete specifications and unclear requirements in the early stages of the project life cycle.

An interactive and flexible tool (figure 3) that implements the software estimation approach, discussed above, was developed. Depending on the specific characteristics of the project, the estimator can choose the appropriate sizing metric and method of cost estimation. The experiments prove that is not reasonable to use SLOC as sizing metric, but it is not also reasonable to use Function points as sizing metric for low level language projects estimation or for legacy system's estimation. Although the results are encouraging and match expectations for the tested projects, research must continue in the direction of evaluating large and complex projects.

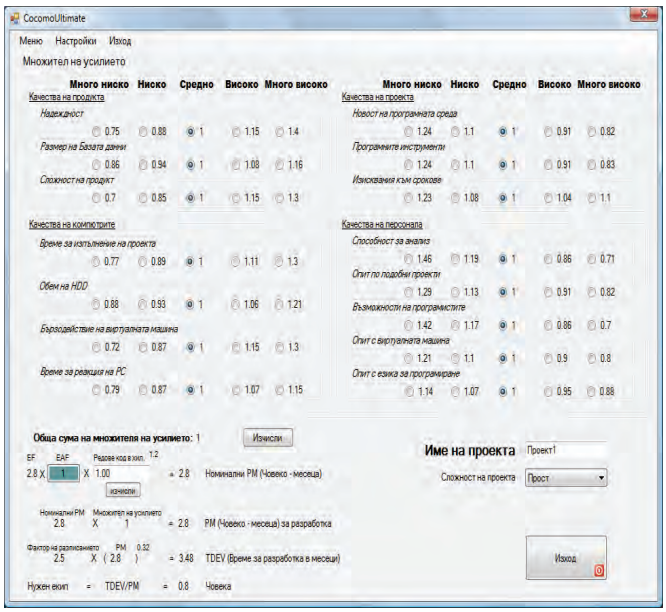


Fig. 2. A tool for Software Cost Estimation (the main window)

References

[1] B.W. Boehm et al, "The COCOMO 2.0 Software Cost Estimation Model", American Programmer, 1996, pp.2-17.

[2] Boehm, B.W. "Software Engineering Economics", Prentice_Hall, 1981.

[3] COCOMO II Model Definition Manual, ftp://ftp.usc.edu/pub/soft_engineering/COCOMOII/cocomo97docs/modelman.pdf.

[4] Karen Lum et al. "Handbook for Software Cost Estimation", Jet Propulsion Laboratory, Pasadena, California, 2003.

[5] Liming Wu, "The Comparison of the Software Cost Estimating Methods", ttp://www.compapp.dcu.ie/~renaat/ca421/LWu1.html

Web-based Interactive System for Multicriteria Decision Analysis

Mariana V. Vassileva[1], Krassimira B. Genova[2] and Boris A. Staykov[3]

***Abstract* – The multicriteria decision analysis problems are decision problems, in which a finite number of decision alternatives are evaluated in terms of multiple conflict criteria. A web-based software system is presented in the paper, which implements an interactive optimizationally motivated method, appropriate for solving problems with a big number of alternatives. The information, which the decision maker (DM) has to set, includes the desired or acceptable changes in the values of some criteria and the desired or acceptable changes in the direction of other criteria. In this way the DM is able to manage the solving process of finding the most preferred solution, which makes him/her feel more confident in the final results obtained.**

***Keywords* – Multicriteria decision analysis, Interactive method, Web-based system.**

1

I. INTRODUCTION

In problems for multicriteria analysis (MA) (also called multicriteria problems with discrete alternatives), several conflicting criteria have to be simultaneously optimized in a feasible set of alternatives [20]. In the general case there does not exist one alternative which could be optimal for all the criteria. However, there does exist a whole set of alternatives, called non-dominated set of alternatives, which possess the following property: every improvement in the value of one criterion leads to deterioration in the value of at least one other criterion. Each alternative from this set is called a non-dominated alternative and can be the final Pareto optimal solution of the multicriteria analysis problem.

The multicriteria analysis problems can be formulated in different areas of resources management /financial, natural, etc./, communications, production, commerce, services, education and others. In these problems the set of alternatives usually consists of a not very large set of variants for choice, ordering or ranking. However, there also exist such, that might have hundreds of alternatives [4, 7]. In some MA problems the evaluations of the alternatives with respect to the criteria have got an uncertain value. The evaluations may be quantitative, qualitative or ranking.

[1]Mariana V. Vassileva is with the Institute of Information Technologies-BAS, "Acad. G. Bonchev", str. Bl. 29A 1113 Sofia, Bulgaria, E-mail: mvassileva@iinf.bas.bg .

[2]Krasimira B. Genova is with the Institute of Information Technologies-BAS, "Acad. G. Bonchev", str. Bl. 29A 1113 Sofia, Bulgaria, E-mail: kgenova@iinf.bas.bg .

[3]Boris A. Staykov is with the Institute of Information Technologies-BAS, "Acad. G. Bonchev", bl. 29A 1113 Sofia, Bulgaria, E-mail: boris@staykov.net .

The methods for solving multicriteria analysis problems, developed by now, can be grouped into three groups. Each one of these methods has its own advantages and disadvantages, which are connected with the ways of setting the preference information, given by the DM.

The first class of methods includes the multiattribute utility theory methods [8] and the analytical hierarchy process (AHP) methods [16]. There are differences in the way in which the DM's global preferences are aggregated in the two subclasses of the methods. In the first one a generalized functional criterion is directly synthesized, whereas in the second subclass it could be said that such a criterion (additive form) is indirectly synthesized. The two subclasses of methods are based on the assumption that there does not exist limited comparability among the alternatives. They use a DM's preference model, which does not allow the existence of incomparable alternatives and the preference information, obtained by the DM, is sufficient to determine whether one of the alternatives is to be preferred or whether the two alternatives are equal for the DM. The second class of methods are called outranking methods [1, 2, 14, 15] and they use a DM's preference model which allows the existence of incomparable alternatives, and the preference information, obtained by the DM, may be insufficient to determine whether one of the alternatives is to be preferred or whether the two alternatives are equal for the DM. In these methods the DM does not make comparison of the criteria or alternatives, but he/she has to provide the so called inter- and intra-criteria information. To solve multicriteria analysis problems with a large number of alternatives and a small number of criteria, the "optimizationally motivated" interactive methods have been suggested [5, 9, 10, 12, 18]. The creation of this kind of methods for solving problems for multicriteria analysis has been inspired by some well-known methods for solving problems of multicriteria optimization. Their aim is to elicit and use implicit information about the decision maker's preferences in order to help steer the decision maker to his most preferred solution. An important result of the interactive process is that the DM can realize better the decision making problem which is to be solved, as well as enrich his/her knowledge about the preferences set, which he/she may alter during the process. The DM usually needs to modify his/her preferences in order to find solutions that are close enough to the goals (values of the criteria), that are both attainable and have trade-off between the criteria that fit the user preferences.

The usage of the different methods for solving real-life problems of multicriteria analysis depends not only on the efficiency of these methods, but also on the development of user-friendly software systems, in which these methods are implemented. The case studies with application of MADSS

(multicriteria analysis decision support system) of general purpose (Expert Choice [17], HIVIEW [13], ELECTRE III-IV [15], PROMCALC and GAIA [2], Decision Lab [3], VIMDA [9]) are most often published. In them one method or several methods, which belong to of the above described group, are implemented. MADSS ensures successive approach of problem structuring – definition of the overall goal, of the essential criteria and alternatives, evaluation of each one of the alternatives with respect to every criterion, and it also aids the deriving of DM's preferences. The development of Internet communications and World Wide Web has considerably extended the possibilities for public interaction, interactive extraction of data and results exchange [19]. The first web-based general purpose MADSS, which provides instruments for structuring of problems, setting preference information and sharing of results over Internet, is Web-HIPRE [11]. Nowadays many researchers apply the methods they have developed in Internet environment [22].

A web-based software system, called Web-MKA, intended for multicriteria analysis, is presented in the current paper. The system is developed on the basis of the interactive partition-based method [12] and the system for multicriteria analysis MKA-1 [21], which operates in Windows environment. The system was developed, using Visual Studio.NET and programmed in Visual Basic. The architecture is client-server oriented. The system is easy-to-use software, which does not require deep knowledge in MA methodology. Users with different levels of qualification can easily enter their problems and set the preference information in terms of desired or acceptable changes of the values, directions or intervals of change of the values of some or all the criteria on the basis of comparisons with the current non-dominated alternative.

The rest of the paper is organized as follows: a short presentation of the main characteristics of the interactive partition-based algorithm for discrete multicriteria problems [12], which is the core of the web-based software system developed, is given in Section II. In Section III the architecture and functions of the Web-MKA software system is developed. The results achieved and the trends for future work are described in the last Section.

II. FEATURES OF THE INTERACTIVE PARTITION-BASED ALGORITHM

The multicriteria analysis problem is defined with the help of the so called $n \times k$ decision matrix $A = \{a_{ij}\}$, $i = 1,...n$, $j = 1,...k$, where I is a set of $n\,(>1)$ deterministic alternatives and J is a set of $k\,(\geq 2)$ criteria. The element a_{ij} of the matrix A denotes the evaluation of the alternatives $i \in I$ with respect to the criterion $j \in J$. From a mathematical point of view there exists a set of so called non-dominated or Pareto optimal solutions, and each of them could be the final solution of the multicriteria analysis problem being solved. The alternative $i \in I$ is called non-dominated if there is no other alternative $s \in I$, for which $a_{sj} \geq a_{ij}$ for all $j \in J$ and $a_{sj} > a_{ij}$ for at least one $j \in J$ (the confirmation is true, when maximal values for the criteria $j \in J$ are sought). From a practical point of view, the solving of a problem for multicriteria analysis is finding of one non-dominated alternative, which satisfies the preferences of the DM to the highest extent.

At every iteration of the interactive partition-based method [12], the DM has the possibility to choose from the current ranked set of the alternatives the one, which satisfies mostly his/her preferences (the most preferred alternative) or the current preferred alternative a_{hj}. In order to obtain the current ranked set, we use a Tchebychev type optimization scalarization problem that is a discrete analog of the scalarization problem, described in [22]. This scalarizing problem is based on the information given by the DM for the desired changes of the values for some or all the criteria in relation to their values in the current preferred alternative. The DM has to choose one from the following options:

- improvement by desired (aspiration) values Δ_{hj};
- improvement as a desired direction of change;
- acceptable deterioration by no more than δ_{hj};
- acceptable deterioration as a desired direction of change;
- the criteria value to lie within an interval, ($a_{hj} - t_{hj}^{-} \leq a_{hj} \leq a_{hj} + t_{hj}^{+}$) around the current value a_{hj}
- to either preserve or improve the current value of the criteria;
- the DM is indifferent about the value of these criteria and as such they may be altered freely.

Using this scalarizing problem, the alternatives are ranked in an increasing order by the value of the objective function of the scalarizing problem. The smaller the value of the objective function of a given alternative, the closer the alternative is to the preferences set by the DM. The first *l* alternatives in this ranking order establish the current set of alternatives that are shown to the DM for evaluation and choice of the current preferred alternative, where $l << n$ is specified by the decision maker. It is possible that there may not exist *l* alternatives that satisfy the requirements of the DM.

This interactive partition-based method uses the advantages of the interactive methods and provides opportunity for the DM to control the process of finding the most preferred alternatives choosing from sets of current ranked alternatives. The main advantage of the method is the reduced burden of the DM, connected with the necessity of direct comparisons of two or more alternatives at each iteration.

III. STRUCTURE AND FUNCTIONS OF WEB-BASED SYSTEM WEB-MKA

The Web-based software system is developed, using MS Visual Studio .NET and MS SQL Express database server.

It contains three main modules:

- a user module - used to store user personal information, like names, contact information, login information, etc.;

- a solving module - used for the interactive process of solving multicriteria analysis problems;
- an archive module - used to store and keep all user defined problems and their decisions in the database for later revision or resolving.

The general system structure is given in Fig 1.

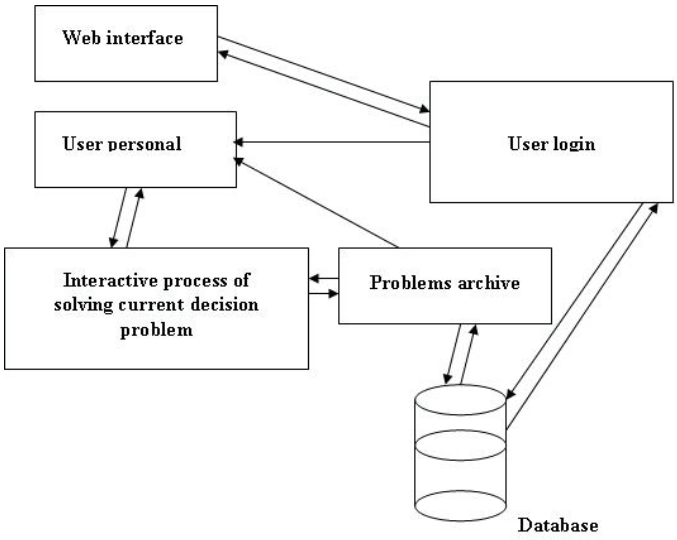


Fig. 1. Web-MKA system structure

The solving module contains a solving core that implements the interactive partition-based algorithm. When calculations are needed, all data information is formatted and passed to the solving core, which on its hand returns a current problem solution or a message that no feasible solution, satisfying the decision maker preferences, can be found.

In order to use the system, one must create a user account and log in the system entering the username and password.

The whole user account information, problem definitions and solving results are stored in the database.

The problem definition is done in a step-by-step manner, as the user enters all the alternatives and criteria types and values. When the definition is done, the solving process can be started. At this point all the information data about the problem is passed to the solving core of the system and an initial solution and ranking are generated and output to the user interface.

One solution contains the following information:

- the current preferred alternative;
- the current inadmissible alternatives;
- the minimal criterion value (rating) for each criterion in each alternative;
- the maximal criterion value (rating) for each criterion in each alternative;

Along with this information, full ranking of the alternatives is also generated and output to the user interface.

At this point, if the current solution satisfies the DM, the process can be stopped. Otherwise the user must define new preferences for each or some of the criteria and generate a new solution.

The decision making process can be described by the block scheme given below (Fig. 2):

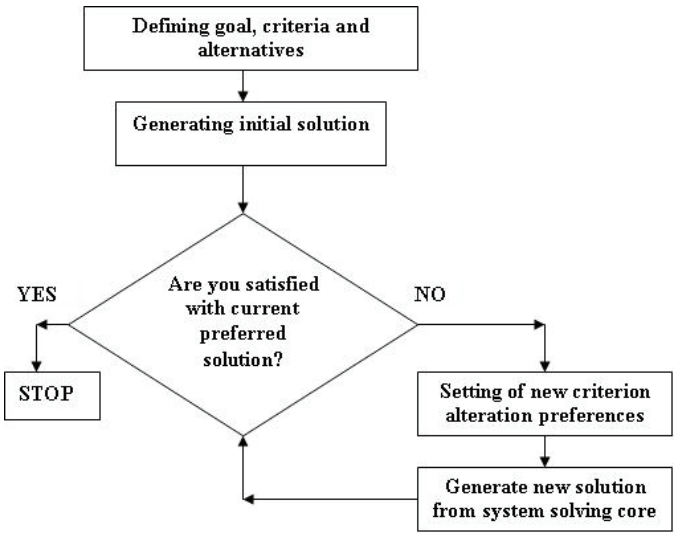


Fig. 2. The working process in Web-MKA

CONCLUSION

An on-line version of the software support system MKA-1, called Web-MKA, is presented in the paper. The architecture and the user interface are developed in such a way that assists the DM in learning about the problem and in evaluating systematically the set of alternatives. At every iteration the DM sets his/her preferences, not using comparison or estimation of the priority of the criteria, but applying a more understandable and easy way. He/she evaluates the current non-dominated alternative and sets the changes in the criteria values, which are acceptable for him/her. In this way the DM can express his/her wish with much more flexibility and select the final non-dominated solution of the problem being solved more confidently.

ACKNOWLEDGEMENT

This work is supported by the Bulgarian National Science Fund, Grant No DTK02/71.

REFERENCES

[1] J. P. Brans, B. Mareschal, "The Promethee Methods for MCDM: the Promcale, Gaia and Bankadviser Software", *Readings in Multiple Criteria Decision Aid (A. Carlos, C. Bana Costa, Eds.)*, pp. 216-252, Springer-Verlag, Berlin, 1990.

[2] J. P. Brans, B. Mareschal, "The PROMCALC & GAIA Decision Support System for Multicriteria Decision Aid", *Decision Support System*, vol. 12, pp. 297-310, 1994.

[3] J. P. Brans, B. Mareschal, "How to Decide with PROMETHEE?", http://www.visualdecision.com, 2000.

[4] C. Zopounidis, M. Doumpos, "Multicriteria classification and sorting methods: A literature review", European Journal of Operational Research, vol. 138, pp. 229-246, 2002.

[5] A. Jaszkiewicz, R. Slowinski, "The LBS-Discrete Interactive Procedure for Multiple Criteria Analysis of Decision Problems", Multicriteria Analysis (J. Climaco, Ed.), pp. 320-330, Springer-Verlag, Berlin, 1997.

[6] T. Hanne, H. L. Trinkaus, “knowCube for MCDM –Visual and Interactive Support for Multicriteria Decision Making”, Published reports of the Fraunhofer ITWM, vol. 50, www.itwm.fraunhofer.de/rd/presse/berichte, 2003.
[7] Ch. Giannoulis, A. Ishizaka, “AWeb-based decision support system with ELECTRE III for a personalised ranking of British universities”, Decision Support Systems, vol. 48, pp. 488–497, 2010.
[8] R. Keeney, H. Raiffa, Decisions with Multiple Objectives, Preferences and Value Trade Offs,. John Wiley & Sons, New York, 1976.
[9] P. Korhonen, “A Visual Reference Direction Approach to Solving Discrete Multiple Criteria Problems”, European Journal of Operational Research, vol. 34, pp. 152-159, 1988.
[10] V. Lotfi, T. J. Stewart, S. Zionts, “An Aspiration-Level Interactive Model for Multiple Criteria Decision Making”, Computers and Operations Research, vol. 19, pp. 671-681, 1992.
[11] J. Mustajoki, R. P. Hamalainen, “Web-HIPRE: Global Decision Support by Value Tree and AHP Analysis”, INFOR, vol. 38, pp. 208-220, 2000.
[12] S. C. Narula, V. Vassilev, K. Genova, M. Vassileva, “A Partition-Based Interactive Method to Solve Discrete Multicriteria Choice Problems”, Cybernetics and Information Technologies, vol. 2, pp. 55-66, 2003.
[13] C. R. Peterson, “HIVIEW – Rate and Weight to Evaluate Options”, OR/MS Today, April, 1994.
[14] B. Roy, “The Outranking Approach and Foundations of ELECTRE Methods”, Readings in Multiple Criteria Decision Aid (A. Carlos, C. Bana Costa, Eds.), pp. 156-183, Springer-Verlag, Berlin, 1990.
[15] B. Roy, “The Outranking Approach and the Foundations of ELECTRE Methods”, Theory and decision, vol. 31, pp. 49-73, 1991.
[16] T. S. Saaty, The Analytic Hierarchy Process, McGraw-Hill, New York,1980.
[17] T. S. Saaty, L. G. Vargas, “Models, Methods, Concepts and Application of the Analytic Hierarchy Process” *Boston, Kluwer Academic Publishers*, 2000.
[18] M. Sun, R. Steuer, “InterQuad: An Interactive Quad Free Based Procedure for Solving the Discrete Alternative Multiple Criteria Problem”, European Journal of Operational Research, vol. 89, pp. 462-472, 1996.
[19] P. Valente, G. Mitra, “The evolution of web-based optimization: From ASP to e-Services”, Decision Support Systems, vol. 43, pp. 1096-1116, 2007.
[20] P. Vincke, *Multicriteria Decision-Aid*, John Wiley & Sons, New York, 1992.S. Haykin, *Neural Networks*, New York, IEEE Press, 1994.
[21] V. Vassilev, K. Genova, F. Andonov, G. Chepilev, M. Vassileva, S. Konstantinova, “Software Decision Support System MKA-1”, *Working Papers IIT/WP-185*, 2004.
[22] M. Vassileva, K. Genova, V. Vassilev, “A Classification Based Interactive Algorithm of Multicriteria Linear Integer Programming”, *Cybernetics and Information Technologies*, vol. 1, pp. 5-20, 2001.
[23] J. Wang, S. Zionts, “WebAIM - an online aspiration level interactive method”, Journal of Multicriteria Decision Analysis, vol 13, number 2-3, pp. 51-63, 2005.

Constructing Data Cube as an Object Oriented Class

Antoaneta I. Ivanova[1]

***Abstract* – The goal of this article is to describe Object Oriented Conceptual Model Data Cube using different kind of data – cancer register, experiments in toeing tank. Problems of data analyzing, saving and summarizing at various levels of detail and on various combinations of attributes are explained. Future tasks to develop appropriate methods to analyze and save consequence data are outlined.**

***Keywords* – data warehouse, data cube, OLAP, object oriented classes**

I. Introduction

OLAP databases often need to summarize data at various levels of detail and on various combinations of attributes [3], [4]. Data warehouse, multidimensional database (MDB), and OLAP applications emphasize multidimensional modelling, which offers two benefits:

- closely parallels how data analyzers think and therefore, helps users understand data;
- helps to predict what the final users want to do, thereby facilitating performance improvements.

The objective of this work is:

- to present a method of structuring and description of different kind of data, providing a faster and efficient data analysis by user defined aggregation functions, as well as consequent storage of the processed data;
- to analyze outlined models of classes and demonstrate that offered model of dynamic object oriented class is universal and useful for completely different data.

II. Model of Data Cube

A. Related works

Trujillo, Palomar and Gomez [5] propose an OO approach to accomplish the conceptual modeling of data warehouses, MDB, and OLAP applications. They use UML to design Data Warehouses because it considers, at conceptual level, the structural and dynamic properties of the information system more naturally than do the classic approaches such as the Entity-Relationship model. Further, UML provides powerful mechanisms—such as the Object Constraint Language and the Object Query Language—for embedding Data Warehouse constraints and initial user requirements in the conceptual model. This approach to modeling a Data Warehouse system yields simple yet powerful extended UML class diagrams that represent main data warehouse properties at the conceptual level.

[1]Antoaneta I. Ivanova is with Department of Computer Science and Engineering, Technical University of Varna, Studentska 1 Str., 9010 Varna, Bulgaria, E-mail: antoaneta_ii@yahoo.com

Some recent papers [1] and [2] as a previous work, describe examples for use of a similar class without emphasis however, on the algorithms for construction of a data cube, the OLAP operations and methods about cancer register and toeing tests – completely different data.

B. Attributes of DATA CUBE

A data cube is constructed from a subset of attributes in the database. Certain attributes are chosen to be **measure attributes**, i.e., the attributes whose values are of interest. Other attributes are selected as dimensions or **functional attributes**. The measure attributes are aggregated according to the dimensions.

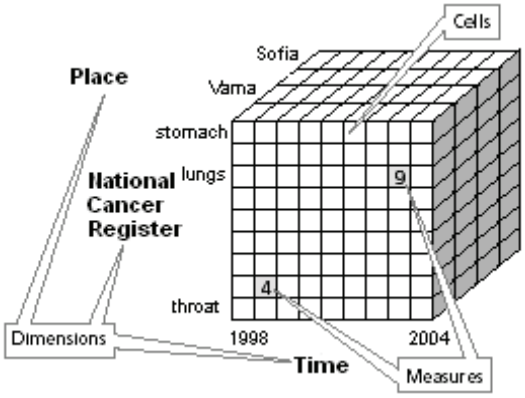


Fig. 1. Model data cube of Cancer Register

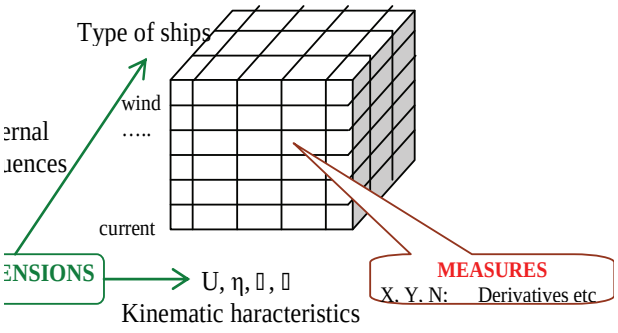


Fig. 2. Model dta cube of PMM ship tests

The measure is additive along a dimension. The User Defined aggregation operator is used to aggregate attribute values along all hierarchies defined on that dimension. The

aggregation of some fact attributes—called roll-up in OLAP terminology— might not however, be semantically meaningful for all measures along all dimensions.

By selecting cells, planes, or subcubes from the base cuboid, we can compare and analyze the feature attributes which are stored in the data cube during data processing. In total, a d-dimensional base cube is associated with 2d cuboids.

B. Classification hierarchies

The definition of the **classification hierarchies** of certain dimension attributes is crucial since these classification hierarchies provide the basis for the subsequent data analysis. As a dimension attribute can be aggregated also to more than one other attribute, multiple classification hierarchies and alternative path hierarchies are also relevant.

Defining the classification hierarchies of certain dimension attributes is crucial because these classification hierarchies provide the basis for the subsequent data analysis. Because a dimension attribute can also be aggregated to more than one other attribute, multiple classification hierarchies and alternative path hierarchies are also relevant. For this reason, directed acyclic graphs provide a common way of representing and analyzing dimensions with their classification hierarchies.

The main of first model on Cancer Register is outlined on Fig.3 – National Cancer Register and Patient.

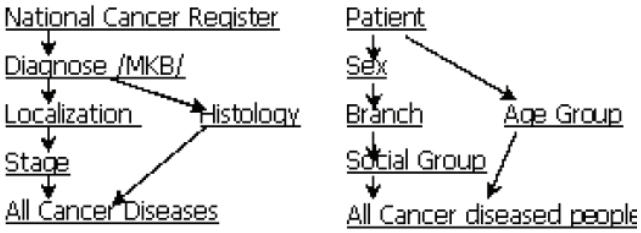


Fig. 3. The classification hierarchies with alternative path - National Cancer Register, Patient

In this model there are other classification hierarchies -***Time*** dimension- 1) Time–quarter–year–5 years 2) Time–season.

Place dimension - classification hierarchy without an alternative path have been defined:

Place–Town/Village–Municipality–Region–Country.

For the second example Fig 4 shows the different classification hierarchies defined for the type of ship dimension. On this dimension, a multiple classification hierarchy is defined so that data values can be aggregated by user defined function along different hierarchy paths.

Ship dimension

1)Ship–Types Depending on navigation region (see, river, etc.)

2) Ship–Hull

3) Ship – Control means – Number of control means – Type

4) Ship – Propulsion means - Number of propulsion means - Type

The data cube may store data, witch are derived from other tests. An alternative path classification hierarchy has been defined with two different paths that converge into the same hierarchy level, for the ***Type of Test*** dimension

1) Type of tests–Free running maneuvers

2) Type of tests–PMM

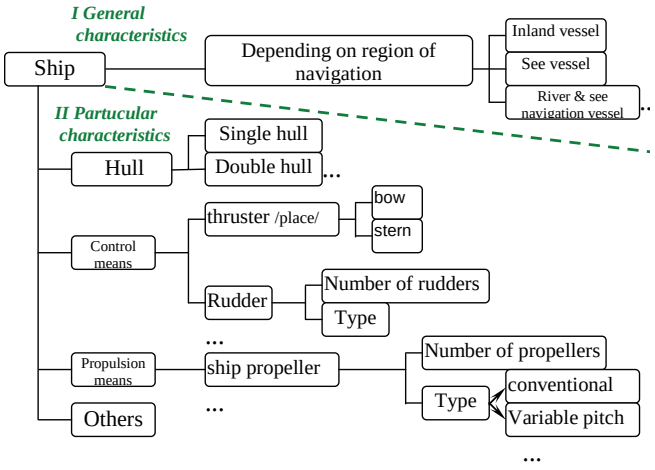


Fig. 4. *Type of ship* - the classification hierarchy display the dimension that defines the cube

In most cases, however, classification hierarchies are not so simple. The concepts of strictness and completeness are important for both conceptual purposes and for further multidimensional modeling design.

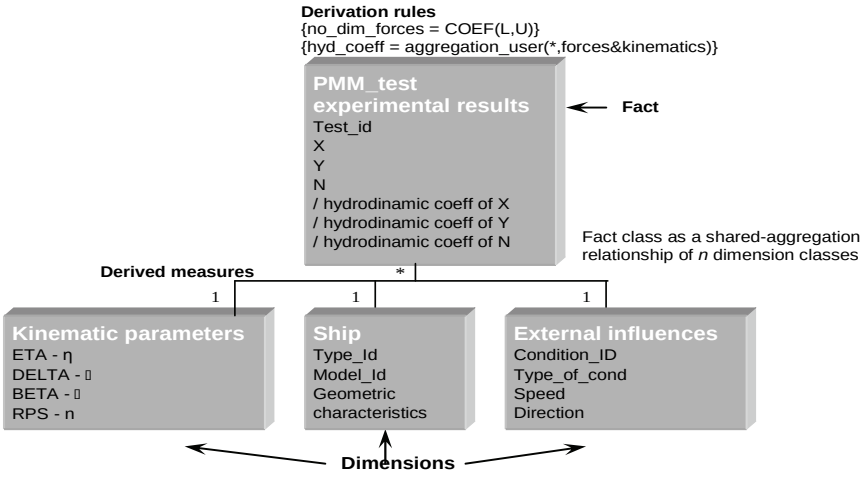


Fig. 5 The PMM_test class consists derivation rules, derived measures and has shared – aggregation relationship with the Kinematics parameters, Ship, Conditions

III. Object Oriented Conceptual Model - Class Definition

OO approach can elegantly represent multidimensional properties at both levels.

A. Structural level

This OO approach is not restricted to flat UML class diagrams in order to model large, complex data warehouse systems. UML's package grouping mechanism groups classes into higher-level units, creating different levels of abstraction and simplifying the final model. In this way, a UML class diagram improves and simplifies the system specifications, created with classic semantic data models such as the Entity-Relationship model. This approach clearly separates the structure of a multidimensional model specified with a UML class diagram into facts and dimensions. Fig 5 shows the example on PMM tests.

A. Dynamic Level

These cube classes are used to represent initial user requirements as the starting point for the subsequent data-analysis phase.

The basic components of the cube classes include the:

- head area, which contains the cube class's name;
- measures area, which contains the measures to be analyzed;
- slice area, which contains the constraints to be satisfied;
- dice area, which contains the dimensions and their grouping conditions to address the analysis; and
- cube operations, which cover the OLAP operations for a further data-analysis phase.

B. Examples

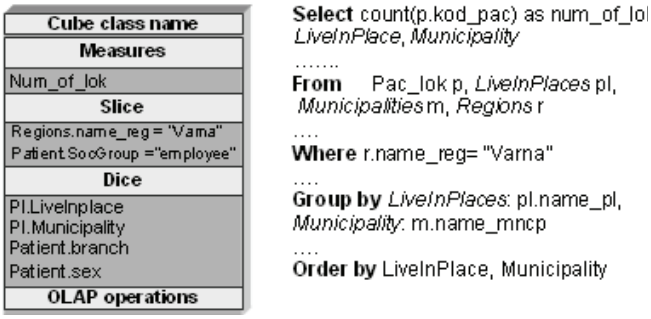


Fig. 6. Cube class example with parameters specified in the measures, slice, dice, and operations areas, and the class's corresponding Object-Query Language specification on model of Cancer Register

In the first model about Cancer Register can add additional methods which calculate for example: 1) disease by

localization, location (place), Age group 2) , 2) death-rate on year or 5 year, for different kind of social group.

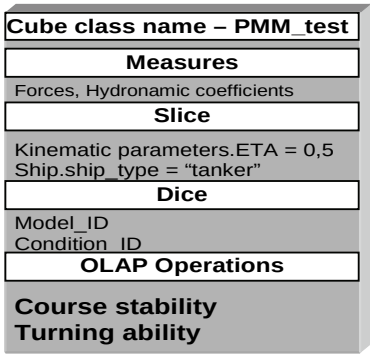


Fig. 7 Cube class definition on PMM tests

Into the second model is represented an extension of created class about OLAP operations – calculate course stability and turning ability on different models.

On the basis of aggregation function over data we can calculate and analyze new data which save on Database.

IV. Conclusion

OO Conceptual models of data cube are a very interesting direction for constructing and using efficiently information extracted from data cube.

OLAP tools implement a multidimensional model from two different levels:

• Structural—the structures that form the database schema and the underlying multidimensional model—also known as the metadata—that provides the model's key semantics (facts, measures, dimensions).

• Dynamic—refers to the definition of final user requirements—also known as method and OLAP operations for further analyzing data.

The offered model of dynamic object oriented class is universal and useful for completely different data.

For a further work will be to develop additional properties or methods in class definition for analyzing data and saving results.

Acknowledgement

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

References

[1] Ivanova A., V.Chotukova, "A model of Object Oriented Class for Extraction and Use of Data Cube Unit Information", Proceedings of CompSysTech'06, Rousse (Bulgaria), 2006.

[2] Ivanova A., "Constructing Object Oriented Class for extracting and using data from data cube", Proceedings of CompSysTech'05, Varna (Bulgaria), 2005.

[3] Lewis P., "Database and Transaction Processing: An Application – Oriented Approach", Addison Wesley, New York, 2003

[4] Molina.H, J. Ullman, J. Windom, "DataBase Systems: The Complete Book", Prentice Hall, 2002, 1119p.

[5] Trujillo J., M. Palomar, J. Gomez, Y. Song, "Designing Data Warehouses with OO Conceptual Models", Computer, 66-75, December 2001, IEEE

Virtual System for Generating Analog and Digital Signals

Tsvetozar St. Georgiev[1] and Georgi N. Krastev[2]

***Abstract* – The report examines the implementation of a virtual system for generating analog and digital signals. The system consists of hardware module NI USB 6008 and specialized software developed using the programming environment LabVIEW. Conducted experimental studies show that the system can be used successfully to test the efficiency of dynamic systems.**

***Keywords* – Virtual systems, Virtual instrumentation, Signal generation.**

I. INTRODUCTION

The rapid development of personal computers in the last 20 years accelerates the changes in the tools for measuring, testing and automation. One of the most important outcomes is the creation of the concept of virtual instruments, which offer many advantages such as increased productivity in developing, higher accuracy and better performance.

The virtual instrument (VI) is a combination of hardware and software elements which under the control of PC has the functionality of the traditional instrument. The virtual instruments consist of an industrial PC or workstation equipped with powerful software, hardware and appropriate drivers. Together all these components perform the functions of traditional instruments. VI represents a fundamental shift from the traditional hardware-oriented systems to instrumental-oriented software using computer power, productivity and opportunities for better visualization. Although PC and integrated circuits are very important in implementation of VI, the computer based software provides the implementation of the concept of VI. While users of traditional tools are limited by decisions made by the manufacturers of the equipment, the VI can meet precise needs of the users.

The signal generators are devices that generate signals with certain characteristics. They are most commonly used for testing different equipment in the process of design or repair. There are different types of signal generators, depending on the type and the supported features of the generated signals. They may be separate devices or those managed by computer.

The report is considered an inexpensive version of the generator of analog and digital signals, built on a module NI USB 6008, a PC and software developed using LabVIEW.

[1]Tsvetozar St. Georgiev is with the Department of Computing, University of Ruse, 8 Studentska Str., 7004 Ruse, Bulgaria, E-mail: TGeorgiev@ecs.uni-ruse.bg

[2]Georgi N. Krastev is with the Department of Computing, University of Ruse, 8 Studentska Str., 7004 Ruse, Bulgaria, E-mail: GKrastev@ecs.uni-ruse.bg

II. ARCHITECTURE OF THE VIRTUAL SYSTEM

A. Main components of the virtual system

The developed virtual system will be a part of a system for automobiles dynamic characteristics investigation. This system includes the following units (Fig.1): a hardware module for generation of signals and for data transfer to personal computer (PC); a PC or Notebook and a virtual instrument for signal generation and data acquisition. The application is developed by programming the NI USB-6008 using LabVIEW and using NI-DAQmx driver software for Windows. A NI-DAQmx virtual channel consists of a physical channel on a DAQ device and the configuration information for this physical channel, such as input range and custom scaling [3].

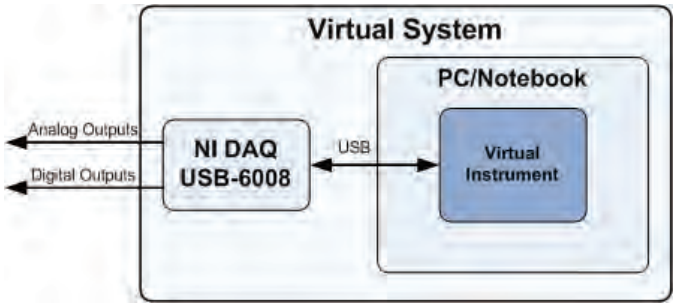


Fig. 1. Architecture of the virtual system

B. Data acquisition (DAQ) hardware module

DAQ hardware module acts as an interface between the computer and the outside world. Devices for data collection may contain analog inputs and outputs, digital inputs and outputs, counters and timers. Information received by them through NI software is analyzed, processed and visualized. DAQ devices are integrated into various hardware platforms, remote desktops, laptops, and industrial computers.

The data acquisition hardware used in this paper is NI USB-6008 multifunction I/O device, which interfaces to the PC through a USB connector. It has 8 differential analog voltage inputs, 2 outputs, 12 channels which can be used as either DI (digital inputs) or DO (digital outputs), and 12-bit resolution [5].

Figure 2 shows the main functional components of the NI USB-6008 [5].

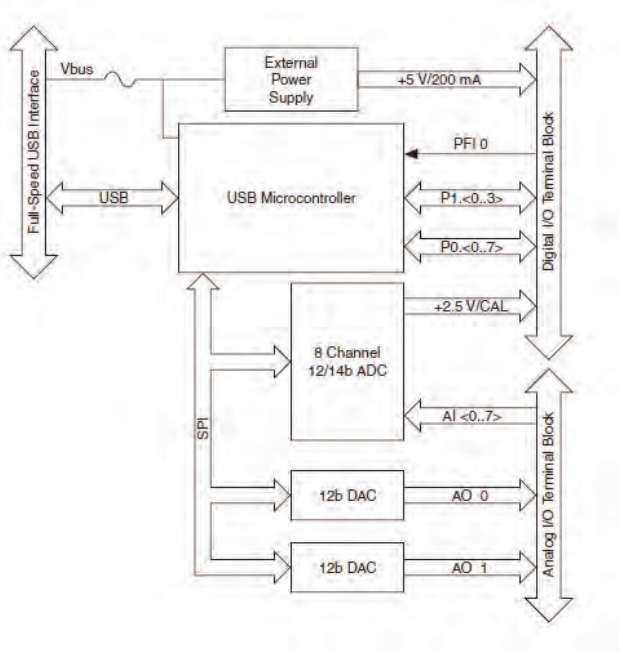


Fig. 2. USB-6008 block diagram [5]

The National Instruments USB-6008 provides basic data acquisition functionality for applications such as simple data logging, portable measurements, and academic lab experiments, as well as abilities for generation analog and digital signal.

C. Virtual instrument

The appropriate environments for development of virtual instruments are LabVIEW, LabVIEW Express, LabWindows/CVI or Measurement Studio of the National Instruments Company, USA.

The LabVIEW environment allows the users to develop test and measurement, data acquisition, instrument control, data logging, and measurement analysis applications. LabVIEW includes a set of VIs that let users to configure, acquire data from, and send data to DAQ devices. Often, one device can perform a variety of functions: analog-to-digital (A/D) conversion, digital-to-analog (D/A) conversion, digital I/O, and counter/timer operation.

The programs which are created by LabVIEW are named virtual instruments because their appearances imitate the real devices [1]. The virtual instruments have an interactive user interface, block diagram of data flow (program code) and icon with connections [2]. VIs can be used as subprograms in other virtual instruments.

The interactive user interface of VI is named Front Panel because it looks like the panel of the physical device. The Front Panel may include buttons, sliders, digital indicators, controls, graphical displays, etc. The user can input information manually by the keyboard and/or import information from data files. The results from the execution of the program code can be seen on the Front Panel. In LabVIEW is used G program language. It has many functions for processing, working with various file formats, analysis, and generation of signals and visualization of data.

The developed virtual instrument is separated on two parts – one for generation of analog signal and other for generation of digital signals. These parts have the following options:

1) Part for generation of analog signals (Fig.5 and Fig.6):

- An ability of the user to select the channel, which led to the generated signal;
- An ability of the user to select the minimum and maximum output value (in V);
- An ability of the user to set parameters of the generated signal - output rate (in ms) and the number of points generated per cycle;
- A possibility to stop the generation of signals (button Stop);
- An ability of the user to clear the chart from previous values.

The hardware module USB-6008 supports only software-timed update of digital-to-analog converter (DAC). The maximum supported refresh rate is 150 reports per second [4]. Therefore, the output rate can not be less than 6 ms. If this value is reduced, it leads to the resolution of the generated sinusoidal signal, but it reduce the frequency of the signal.

The virtual instrument calculates the approximate frequency of generated signal using the following formula:

$$f_{aprox} = \frac{1}{O_R * P_C}, \text{Hz}$$

where f_{aprox} is the approximate frequency of generated signal, O_R is the output rate (ms) and P_C are the points per cycle.

2) Part for generation of signals on digital outputs (Fig.3):

- An ability of the user to set the level (0 or 1) of digital outputs using push buttons (for example from “Digital Out 0” through “Digital Out 5”).
- An ability of the user to control the execution of the program using the Stop button.
- An ability of the user to monitor the change of the signal levels by using a special widget Digital Waveform Graph. On this graph with different colors are presented different signals and the level of each signal is indicated with 0 or 1.

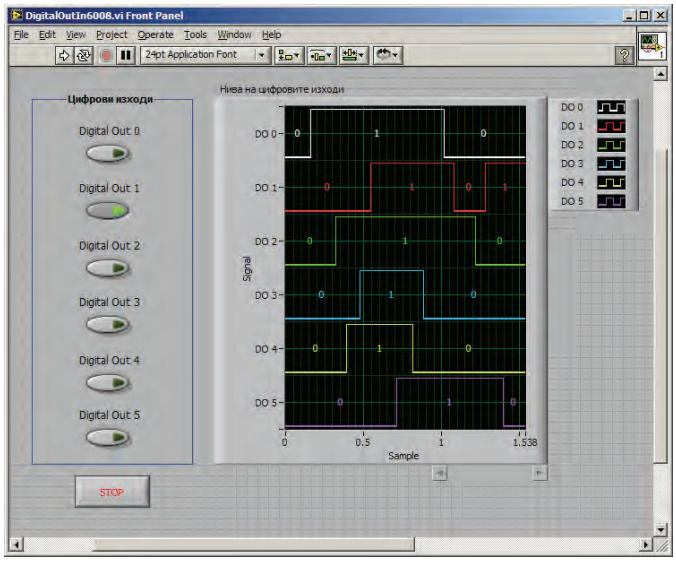


Fig. 3. Front panel of the part for digital signals generation

III. TESTING OF THE SYSTEM

For the purposes of the system testing the following approach was used. The analog output (AO0), where the signal is generated was related to one of the analog inputs (AI0) and six digital outputs were connected to six digital inputs (fig.4).

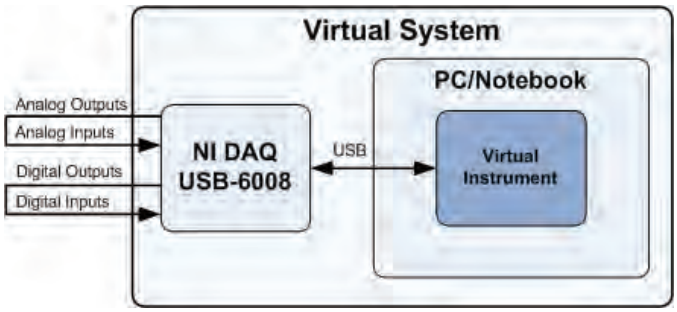


Fig. 4. Architecture for testing the virtual system

The analog part of virtual instrument was modified in such a way to allow selection of input and output channels and to display generated and acquired analog signals. Figure 5 shows the testing results of generation and acquisition of analog signal with an approximate frequency of 33 Hz (the output rate is 6 ms and the number of points per cycle is 5). The maximal value of the generated signal is 5 V and the minimal value is 0 V.

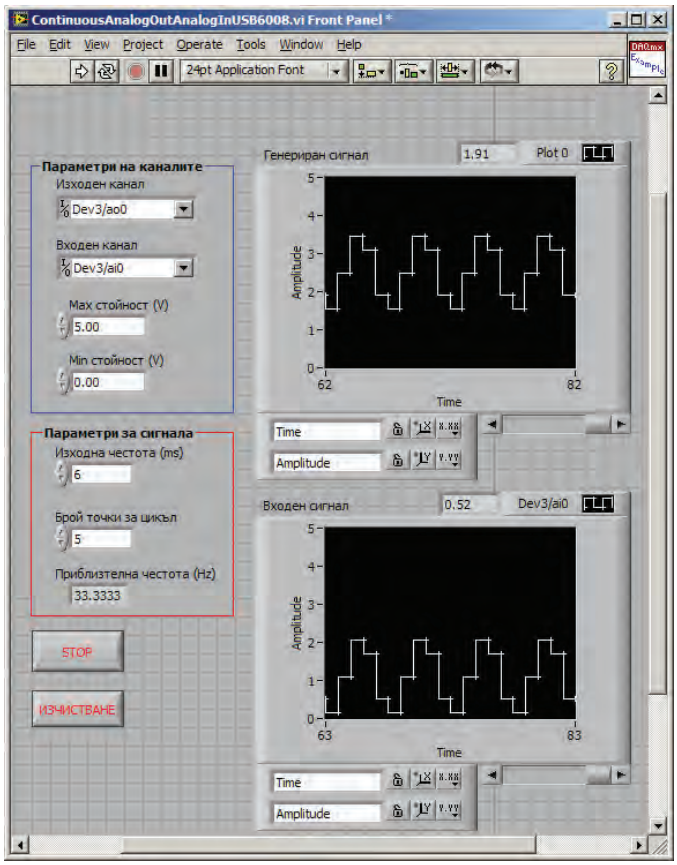


Fig. 5. Generation and acquisition of analog signal with an approximate frequency of 33 Hz

Figure 6 shows the results of generation and acquisition of analog signal with an approximate frequency of 1.67 Hz. In this case the output rate is also 6 ms but the number of points per cycle is 100.

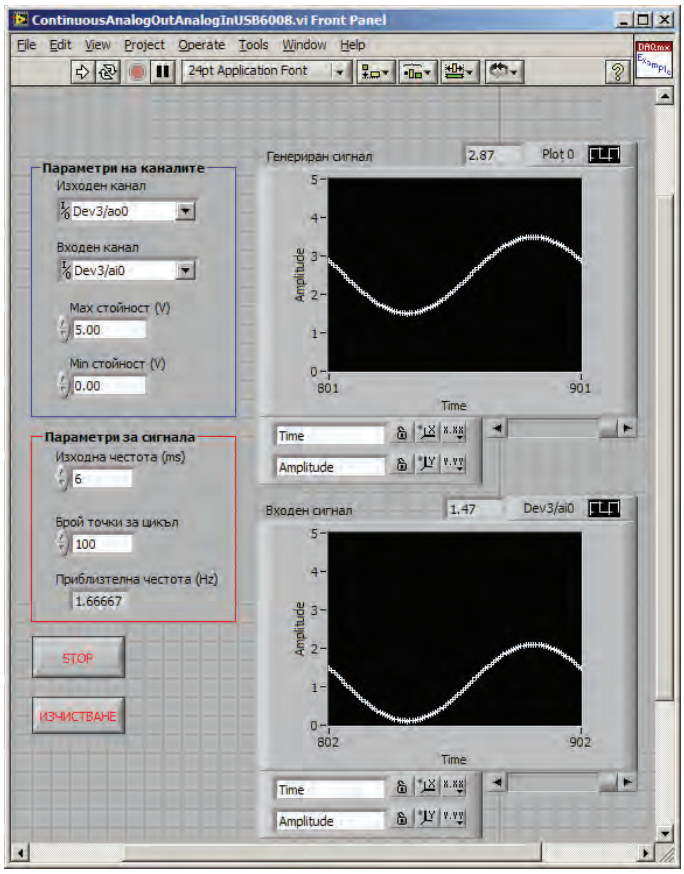


Fig. 6. Generation and acquisition of analog signal with an approximate frequency of 1.67 Hz

It is easily to find that the hardware module USB-6008 is unable to generate analog signals with high frequency and it is suitable for generating signals with a low frequency (up to 15 Hz). The reason for this is the limit of refresh rate which is 150 reports per second.

The digital part of the virtual instrument was modified to display the generated and acquired digital signals. To test the virtual instrument 12 digital input-outputs are divided into two parts of 6 channels. The first part (from P0.0 to P0.5) is initialized as digital outputs, which generate signals. The second part (from PO.6 to P1.3) is initialized as digital inputs. Thus initializing digital outputs are connected to digital inputs using a cable. On the front panel of the virtual instrument are added LEDs which indicate the state change of the inputs. On the front panel a second Digital Waveform Graph element is also added, whereby the user can monitor the change history of the signals.

The test generation and acquisition of six digital signals is given on Fig.7.

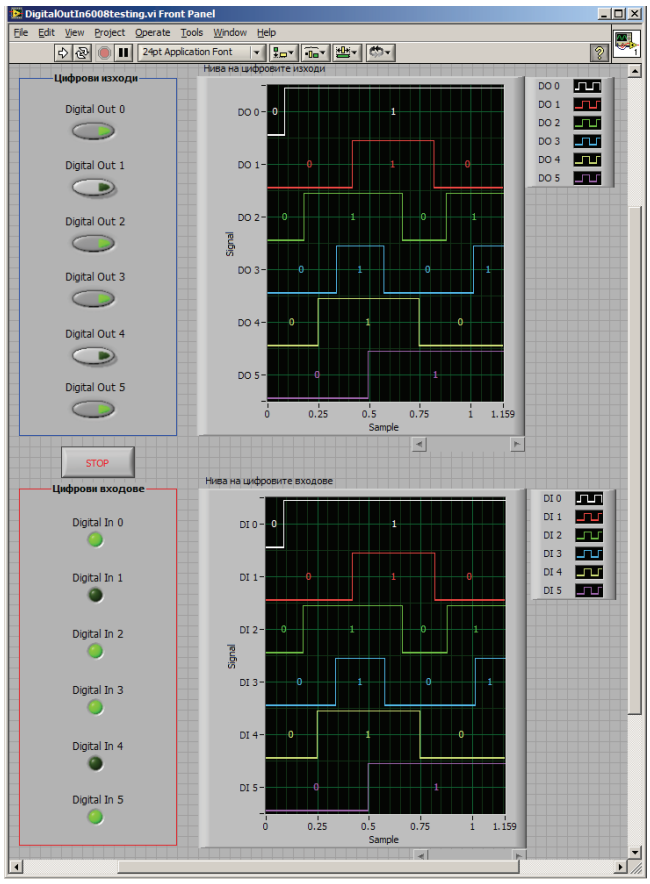


Fig. 7. Generation and acquisition of digital signals

IV. Conclusion

The results from the investigations show that the developed virtual system can be used for generating of low frequency analog signals and generating signals on digital outputs of the hardware module. It can be easily integrated in more complex software for automobile investigation which is based on the virtual instrument technology.

It is also possible to use the system for testing the efficiency of other virtual instruments developed for data acquisition of analog and digital signals.

References

[1] Georgiev, T., G. Krastev, Systems for Automation of Scientific Research, Rousse, Bulgaria, 2002.

[2] Travis, J., LabVIEW for Everyone, Prentice Hall, USA, 2002.

[3] Data Acquisition and Signal Conditioning Driver Software Options, http://www.ni.com/pdf/products/us/20056083301101d.pdf

[4] Low-Cost, Bus-Powered Multifunction DAQ for USB – 12- or 14-Bit, up to 48 kS/s, 8 Analog Inputs, http://www.ni.com/pdf/products/us/20043762301101dlr.pdf.

[5] User Guide and Specifications NI USB-6008/6009, www.ni.com/pdf/manuals/371303l.pdf.

Asynchronous Micro-Pipeline With Multi-Stage Sections

Dimitar S. Tyanev[1], Stefka I. Popova[2]

Abstract: **The interface of multi-stage micro-pipeline sections building continuous micro-pipelines is defined and analyzed. As the multi-stage micro-pipeline sections have own memory, such micro-pipelines don't need additional registers. In these conditions there is pipeline asynchronous protocol and implementing control unit synthesized. The protocol's operation is shown in cases, arising from combined work of neighbor multi-stage micro-pipeline sections. Possible problems of combining one- and multi-stage sections are indicated.**

Keywords: **Computational devices, Micro-pipeline, Race conditions, Synchronization**.

I. INTRODUCTION

Micro-pipelines contain consecutively connected micro-pipeline sections which structure is made of register and logic (for example [6÷18] or another). The register supports the data and the logic implements the necessary computations, but it is not required. If the delays describing the particular sections are relatively the same, there is common control and the micro-pipeline is defined as synchronous. If the delays are significantly different, the control is distributed and the micro-pipeline is determined as asynchronous. After every registration impulse new data enters and is processed at certain section in the both types of micro-pipelines. So after each impulse the intermediate results are moving from section to section. In these terms the micro-pipeline sections in such kind of pipelines are defined as one-stage. The stage period at particular sections is set by the switching time of the logic.

Data shifting from section to section in asynchronous micro-pipelines of mentioned type is implemented after the hand-shake principle, in 2-phase or 4-phase protocol. The protocol is realized by control block, containing some version of well-known Mueller C-element [5÷18]. The nature of control is asynchronous because the shifting of current results to the next section is possible only if the last is not busy. This is the main reason such type of micro-pipelines to be defined as asynchronous.

Micro-pipeline sections with internal feedback are presented in [1÷4]. These sections implement iterative computations and are designed as synchronous devices. They work as synchronous because of local clock. Such micro-pipeline sections can be determined as multi-stage on account of the internal (local) clock. The delays generated from these sections are significantly different amongst themselves, as well as compared to the delays from one-stage sections, so as devices they can be included only in asynchronous micro-pipelines. As the multi-stage micro-pipeline sections have their own memory, the micro-pipelines with such sections don't need additional pipeline registers. This paper presents the interface of this kind of multi-stage micro-pipeline sections and the control possibility with serial inclusion.

[1]Dimitar S. Tyanev is with the Faculty of Computer Sciences and Technologies, Technical University of Varna, Bulgaria.
e-mail: dstyanev@yahoo.com
[2]Stefka I. Popova is with the Faculty of Computer Sciences and Technologies, Technical University of Varna, Bulgaria.
e-mail: s_ivanova@abv.bg

II. MULTI-STAGE MICRO-PIPELINE SECTION

Micro-pipeline sections with internal feedback can be presented by the following general structure:

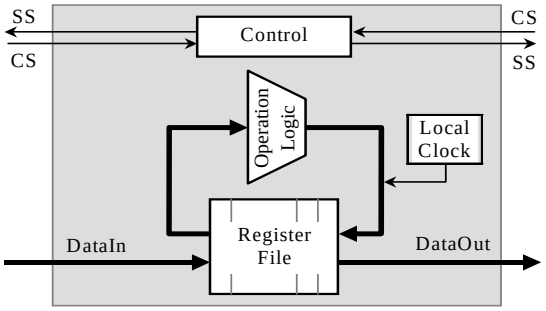


Figure 1 Structure of multi-stage micro-pipeline section

The structure contains three basic elements – register file (*Register File*), which consists of one or more registers (pipeline fixers) and set of combinatorial logic (*Operation Logic*) implementing necessary computations. The main characteristic of this structure is the internal feedback. The third element (*Control*) is integral part of such sections and realizes their internal control. Some of the tasks of the internal control will be discussed in this paper. The most important function of the internal control is to carry out the communication between sections and to process and generate the respective signals (*Status Signals*, *Control Signals*).

Multi-stage micro-pipeline sections are intended to implement various types of cyclic algorithmic structures. The iterative computations carried by such sections lead to long detention of the main computational process, which allows them to be defined as very asynchronous regarding to it.

Micro-pipeline sections of presented type can be stable in one of the next three states:

1. "Free" state. It means that the result of computations in the current section *k* is sent through the output bus *DataOut* to the next section *k+1* and the last confirm the reception. In this state the logic connections in the structure are determined so the section is ready for the next start. In these terms, the

state “Free” is the information necessary to section *k-1*, because there is no sense to be started if the next is not free;

2. “Busy” state. Current section *k* is in this state during the implemented cyclic computations. At this time the input data bus *DataIn* is switched off and the data on it does not have any impact on its structure. The data on its output bus *DataOut* is not valid so it shouldn’t be accepted and used by the next section;

3. “Ready” state. It is alternative to the previous state. It occurs in the current section *k* when its computations finish and the true value of the result are set on the output bus *DataOut*. In this state the section supports the obtained result on the output bus so it is still only there.

Note: All micro-pipeline sections in certain micro-pipeline should be forced to “Free” state after the power switching, as well as in other situations that require this state. The last is defined as initial or last state for the particular section and for the pipeline in general.

At the time of the pipeline operation the order of the states in each section is as follows:

... “Free”, “Busy”, “Ready”, “Free”, “Busy”, “Ready”, ...

States in which every section of the micro-pipeline could be are declared by the following signals (signals of SS type):

1. Signal F_k (Free). It is produced after switching of the section in “Free” state. This signal is conditionally directed to the back, i.e. to the previous section *k-1*;

2. Signal B_k (Busy). It is produced after switching of the section in “Busy” state. The signal is directed to the back as well;

3. Signal R_k (Ready). This signal is produced after switching of the section in “Ready” state. It is conditionally directed to the front, i.e. to the next section *k+1*.

III. CONSECUTIVE INCLUSION OF MULTI-STAGE MICRO-PIPELINE SECTIONS

Multi-stage micro-pipeline sections are included in exact sequence according on the current algorithm. In order to the pipeline organization there is certain control required, depending on the signals SS and CS. The control of the connection between each pair neighbor sections is assigned to control automation (CA) which must recognize the states of the both sections and to manage their dialogue. In other words, this automation has to synchronize the common work of two neighbor sections using signals of SS type and in response to produce the necessary CS-signals (Figure 1). In conformity with Section 1, the pipeline automation should be asynchronous. Figure 2 presents pipeline from the discussed type.

As it seen from the figure, the control automation CA generates two control signals:

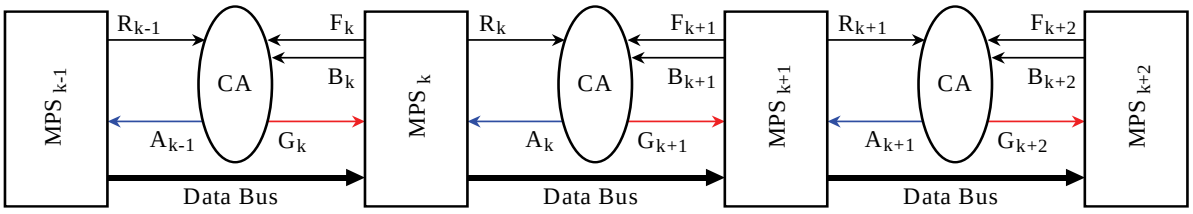


Figure 2 Micro-pipeline with multi-stage sections

1. Signal G_{k+1} (*Go*). With this signal the automation starts computations into the next micro-pipeline section, i.e. the signal is directed to the front. The emission of this signal must be possible only if the previous section is in “Ready” state and next – in “Free” state. This is situation in which the previous section finished its computations and supports the results on the output bus. At the same time the next section is free and waits for new data;

2. Signal A_k (*Acknowledgement*). With this signal the automation informs the previous section that transferred data is successfully received by the next one. The signal is conditionally directed to the back. The previous section must announce “Free” state in response. The logic of such state is presented as:

$$F_k = R_k \cap A_{k+1} \quad . \qquad (3.1)$$

The control automation’s operation is showed by the graph at Figure 3. The graph shows that the automation has two states. The initial state is marked as *S0*. In this state the automation supports the signal A_k which is a reason for the “Free” state of section *k*. From this state automation is switched to state *S1* only when the two neighbor sections complete the required transition condition: $R_k \cap F_{k+1} = True$. Once automation is in *S1* state it produces the signal G_{k+1}. This signal appears to be initial for the next section *k+1*.

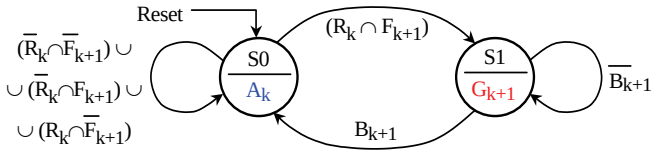


Figure 3 Transition graph in CA

After section *k+1* begin its operation, it passes to “Busy” state and forms signal B_{k+1}. This signal causes the switching of the synchronizing automation back to the initial state *S0*. From this state is produced signal A_k which informs the previous section about the successful transfer of its data to the next section. It is the end of the exchange session at this stage of the micro-pipeline.

The synthesis of the pipeline automation in terms of the transition graph from Figure 3 leads to the principal logic structure presented at Figure 4 in two variants:

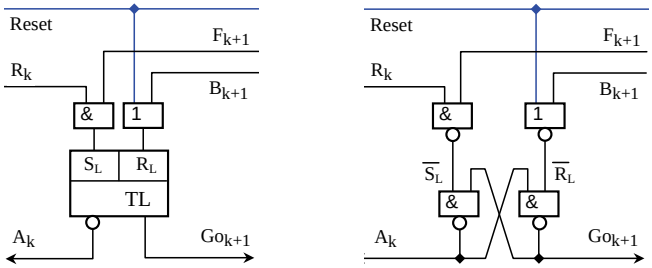


Figure 4 Principal structure of CA

The automation is implemented following Moore's structure by one asynchronous RS-Latch flip-flop. Internal states of the automation are coded as follows:

$$S0 = \overline{Q}\ , \qquad S1 = Q\ . \tag{3.2}$$

So the right input implements signal *Go* and the inverse input – signal *Acknowledgement*.

Because of the different duration of the computations in two neighbor sections there are two possible situations for the control automation, for example:

1. Section *k+1* is free and waits for the end of section *k*'s computations;
2. Conversely, the section *k* is ready and waits for the end of computations in section *k+1*.

The time-diagram from Figure 5 shows these two cases of the pipeline automation switching. In the first situation (the left half) section *k+1* waits for the data from section *k*. The automation is in *S0* state, waiting for signal R_k .

In the second case (the right half) section *k+1* still works while the previous section *k* finish computations and is in "Ready" state at the same time, producing signal R_k . With this signal the automation is switched to *S1* state from the *S0* state.

The transitional process corresponding to automation's graph is presented twice – into the left and into the right side of the time-diagram at Figure 5 and shows the beginning, work, final and repeated start of the micro-pipeline section *k+1*. Analyzing this time-diagram can be concluded that the control automation implements 4-phase protocol.

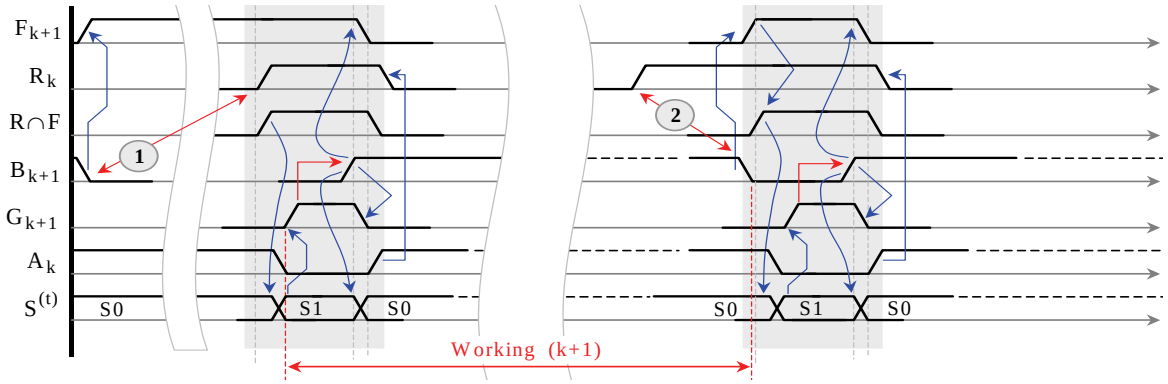


Figure 5 Switching of the synchronization automation

IV. SYNCHRONIZER

For starting the computational process in multi-stage micro-pipeline sections [1], [2], [3], [4] is required start impulse, conditionally called *Enable* which must be synchronous with the rising edge of the local clock impulses and to have duration up to one period. From discussion in Section 3 is clear that the parent of the start signal will be the asynchronous control automation which generates signal *Go*. This signal is asynchronous regarding to the local clock impulses. Thus, there is a task for the signal *Go* converting into signal *Enable*.

The task of converting asynchronous signal into synchronous is illustrated by the time-diagram at Figure 6:

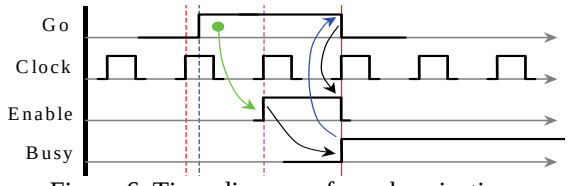


Figure 6 Time-diagram of synchronization

Can be seen that the signal *Go* appears asynchronously in the time of clock impulse from the Clock sequence. The start impulse *Enable* to micro-pipeline section follows as response. With appearance of the *Busy* signal disappear the signals *Go*, which is function of the pipeline automation, and *Enable*, which should be function of the synchronization schema. Signal *Busy* characterizes the state of the micro-pipeline section, as it was described in Section 2.

The schema, which implements expressed logic, is presented at the following figure:

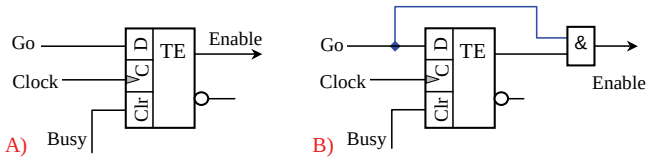


Figure 7 Principal structure of the synchronizer

There is dynamic D flip-flop with Edge structure used, which is basic approach for synchronization of asynchronous signals [15, 19, 20, 21, 22]. Fixing of the signal *Go*'s logic value is made with the rising edge of the clock impulse. If it is missed, as it shown at Figure 6, it could be done with the next

impulse. For reliable fixing there must be restriction on the initial asynchronous value. This value should be kept in time during the following period:

$$t_{Go} \geq T + t_1 \quad , \qquad (4.1)$$

where t_{Go} is the duration of the signal Go;

T is the period of signal Clock;

t_1 is the duration of single impulse into signal Clock.

If the signal *Go* has significantly bigger duration, in the name of the necessary duration of *Enable* signal is used forced cleaning of the flip-flop by the *Clr* (Clear) input. Notice that this input is with high priority and if the signal *Busy* is connected to it (*Busy* is active through all computational cycle), the forced keeping of the flip-flop in zero state is guarantee for reliability and makes impossible false values of the signal *Enable*.

There are two options for the synchronizations schema (look Figure 7). The variant A synchronizes by the rising and by the falling edge; the variant B synchronizes only by the rising edge. In the last variant, because the *Enable* signal is function of the input signal *Go*, it is the direct reason for its disappearance. The synchronizer should be assumed as part of the logic of every multi-stage micro-pipeline section.

V. CONCLUSION

Discussed in this paper type of micro-pipeline sections and their serial inclusion in certain micro-pipeline is only one special case. As it was told in the beginning, there are different types of possible micro-pipeline sections. It is possible their serial inclusion in various combinations, for example: one-stage section followed by multi-stage section or multi-stage section, followed by one-stage. If we consider longer sequence, the possible combinations will be more. The only one known case, which was mentioned in the introduction, corresponds to sequence of one-stage micro-pipeline sections, where the control automation is based on Mueller C-element.

While one-stage sections are served only by two signals, usually called *Request* and *Acknowledgment*, for the multi-stage sections is defined a few possible signals. However, knowing the structure of the multi-stage sections we can claim that there are preconditions for other interpretations, as well as for possibilities for generalization. Admitting the mentioned combinations of serial arrangement, it will lead to necessity of different kinds of control automation. Of course, these possibilities are topic of future work.

REFERENCES

[1]. Тянев, Д., Колев, С., Янев, Д., Метод за реализация на апаратни самоуправляващи се циклически структури - част II, "Компютърни науки и технологии", ТУ-Варна, ISSN 1312-3335, година V, брой №2/2007, стр. 23-30.

[2]. D. Tyanev, S. Kolev, D. Yanev, Micro-pipeline Section For Condition-Controlled Loop, CompSysTech'09, 18-19 June 2009, Ruse, Bulgaria.

[3]. D. Tyanev, D. Yanev, S. Kolev, Method for realization of self-controlling loop apparatus structures, Fifth ISCCS'09, 5-6 November 2009, Sofia, Bulgaria.

[4]. Тянев, Д. С., Колев, С. И., Йосифов, В., Метод за реализация на апаратни самоуправляващи се циклически структури, ТУ-Варна, ЮС "45 години ТУ-Варна", 2007, ISSN 1311-896X, стр. 130-135.

[5]. Миллер, Реймонд Е., Теория переключательных схем, том 2 – последовательностные схемы и машины, Москва, Издательство "Наука", 1971.

[6]. Sutherland, Ivan E., Micropipelines.

[7]. Tiberiu Chelcea, Girish Venkataramani, Seth C. Goldstein, SelfResetting Latches for Asynchronous MicroPipelines, Proceedings of the 44th annual ACM/IEEE Design Automation Conference, June 2007.

[8]. Mannakkara, C., Yoneda, T., Asynchronous pipeline controller based on early acknowledgement protocol, National Institute of Informatics: DI, Graduate University for AS, Tokyo, Japan, NII-2009-015E, Sept. 2009.

[9]. GALAXY-Project, Milos Krstic, , Specification of optimized GALS interfaces and application scenarios, GALAXY, 12-2008.

[10]. Jens Muttersbach, Globally asynchronous locally synchronous, architecture for VLSI Systems, PhD thesis [PDF], ETH Zurich, Diss. ETH №14155, 2001.

[11]. Milos Krstic, Eckhard Grass, New GALS Technique for Datapath Architectures, in Integrated circuit and system design: power and timing modeling, by Jorge Juan Chico, Enrico Macii, p.161, books.google.com, 2003; Lecture Notes in Computer Science, ISSN 0302-9743, Volume 2799/2003.

[12]. Eckhard Grass, Frank Winkler, Miloš Krsti, Enhanced GALS Techniques for Datapath. Applications.

[13]. Xin Fan, Miloš Krstić, Eckhard Grass, Analysis and Optimization of Pausible Clocking based GALS Design.

[14]. Kenneth Yun, Peter A. Beerely, Julio Arceo, High-Performance Asynchronous Pipeline Circuits, In Proc. International Symposium on Advanced Research in Asynchronous Circuits and Systems, IEEE Computer Society Press, 1996.

[15]. Ran Ginosar, Fourteen ways to fool your synchronizer - Asynchronous Circuits, IEEE Pr. of the 9th ASYNC'03.

[16]. Stefan Hirschmann, Asynchronous processors, Seminar Embedded System Design, Institute of Computer Science, University of Innsbruck, February 25, 2008.

[17]. Chang-Jiu Chen, Wei-Min Cheng, Hung-Yue Tsai, Jen-Chieh Wu, A quasi-delay-insensitive microprocessor core Implementation for Microcontrollers, Journal of information science and engineering 25, 543-557 2009.

[18]. Montek Singh, Chapel Hill, Steven M. Nowick, US Patent, 2005/0156633, Circuits and methods for high-capacity asynchronous pipeline processing.

[19]. Ronald J. Tocci, Neal S. Widmer, Digital systems: principles and applications, 8th ed., Prentice-Hall Inc., ISBN 0-13-085634-7, 2001.

[20]. Pong P. Chu, RTL Hardware Design Using VHDL: Coding for Eflciency, Portability and Scahbility, Wiley IEEE Press, ISBN-13: 978-0-471-72092-8, 2006.

[21]. Daniel Page, Practical Introduction to Computer Architecture, Springer, ISBN 978-1-84882-255-9, 2009.

[22]. Richard F. Tinder, Asynchronous Sequential Machine Design and Analysis, Morgan and Claypool, ISBN: 9781598296907, 2009.

POSTER SESSION PO VIII

PO VIII - Computer Systems and Internet Technologies II

Collaborative Environment Aimed to Promote Knowledge Creation

Tania K. Vasileva[1] and Vassiliy P. Tchoumatchenko[2]

Abstract – The paper presents the Knowledge Practices System, a virtual environment aimed to promote collaborative knowledge creation and continuous work around shared objects. The environment makes use of recent Web 2.0 and Semantic Web applications. It is designed to provide specific affordances for joint development of shared objects as well as organizing related tasks and user networks and interactions. The main focus in the present paper is on describing basic functionalities of the system and integrated tools dedicated to support innovative knowledge practices in educational and workplace settings.

***Keywords* – Collaborative learning, Virtual environment.**

I. INTRODUCTION

A fundamental challenge for modern society is to organize work with knowledge in a way that supports the shared efforts of innovation, that is, collaborative efforts for developing novel objects like commercial products, scientific theories, and new forms of working or technology.

A challenge for education is to prepare learners for the emergent knowledge society through appropriate pedagogical practices that promote competencies for sharing, creating and working with knowledge and knowledge artifacts in an innovative way.

The current approaches of working with knowledge in educational and workplace settings are still focused on individuals' skills and knowledge structures (knowledge acquisition) on the one hand, or on social and cultural interaction (participation) on the other hand. The rapidly growing body of scientific and professional knowledge as well as dramatically changing work-practices call for new pedagogical approaches suitable to prepare students but also practitioners to actively participate in a knowledge-based society. Towards this end there is a need for new methods but also tools which support both students as well as practitioners in sustained efforts of innovation and knowledge creation and social participation around shared objects of activity.

Real knowledge advancement and knowledge creation is a long-term process in which progress takes place through series of inquiry cycles. The challenge is to design an environment that supports such long-term work besides individual courses or projects. To answer this challenge we present the KP-Lab System as an innovative collaborative environment aimed to promote knowledge creation and continuous work around shared artifacts. The KP-Lab System is currently under development in the Knowledge-Practices Laboratory (KP-Lab) project [1] funded by the European Commission.

[1]Tania K. Vasileva is with the Faculty of Electronic Engineering and Technologies, 8, Kliment Ohridsky bul., 1000 Sofia, Bulgaria, e-mail: tkv@tu-sofia.bg

[2]Vassiliy P. Tchoumatchenko is with the Faculty of Electronic Engineering and Technologies, 8, Kliment Ohridsky bul., 1000 Sofia, Bulgaria, e-mail: vpt@tu-sofia.bg

II. THE KNOWLEDGE PRACTICES ENVIRONMENT

Knowledge Practices Environment (KPE) is designed to provide specific affordances for joint development of concrete objects as well as for planning, organizing and reflecting on related tasks and user networks [2]. KPE is a virtual environment that includes a set of basic, integrated tools (e.g., working spaces with real-time and history-based awareness, wiki, note editor, commenting, chat, semantic tagging and semantic search) for working with the shared knowledge objects.

KPE enables viewing of knowledge objects and their relations from several perspectives. Three basic perspectives are *content*, *process* and *community* views. Various tools and functionalities are highly integrated in the basic views to enable versatile and flexible connection, organization and reflection on all information related to the knowledge objects, processes and people concerned. Below, the basic functionalities of KPE that provide affordances for collaborative knowledge creation practices are described.

A. Work with knowledge Artefacts

In KPE, user groups can create 'Shared spaces' through which various knowledge artifacts can be shared and co-constructed. But instead of providing only a space to store and manage vast number of documents, KPE enables the users to organize knowledge objects (represented by graphical icons) through flexible, visual representations. A central view in KPE for working on knowledge artifacts is the *Content view* that allows free visual arrangement and linking of its content (Fig. 1).

In addition to a possibility to upload files in a Content view, some specific tools are built in or integrated in KPE to support easy production of texts and sketches as well as co-editing and comparison of text versions to facilitate idea exchanging when working on shared knowledge objects. [3].

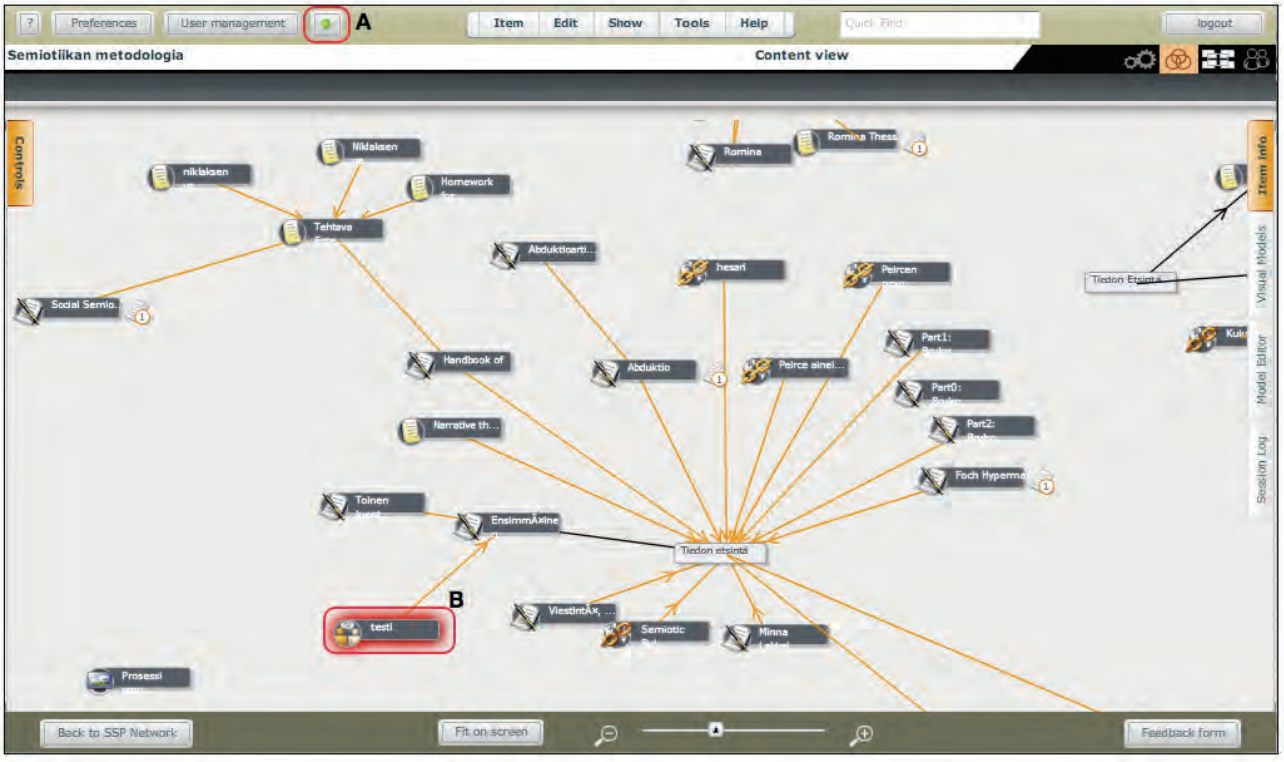


Fig.1 Visual arrangement of content items in a content view from a course held in Helsinki University

The *Note editor* supports quick brainstorming and production of ideas together and allows viewing, editing, versioning, comparing and printing of shared documents in flexible manner. With Note editor, users can directly write their ideas and thoughts as content items in a shared space, without the labor of creating and uploading an external text file. All members of a space can open and edit the created notes and view their previous versions. Furthermore, users can open many notes simultaneously for comparison and integration, and link notes to other content items in the Content view as shown in Fig. 2.

The Content view also includes a Sketch pad tool that is analogous to Note editor, but for creating knowledge through drawing instead of writing. The Sketch pad allows users to collaboratively view, create, compare and co-edit sketches synchronously or asynchronously in a collaborative manner. It provides the functionality for drawing simple graphics (squares, circles, triangles, lines, arrows, free drawing with colors and with simple editing commands) with texts added in between for sketching in white-boarding manner (Fig. 3). The Sketch pad enables easy in-context drawing to support brainstorming and externalizing ideas that are sometimes hard to explicate by only using verbal means. The drawn images can be used for semantic grouping of the rest of the material.

In addition, KPE affords groups the ability to write collaboratively in an integrated wiki. A wiki document can be created as a content item in the Content item view, which offers the possibility to access the same wiki document from a shared space. The progress and changes made to the document are visible to all group members.

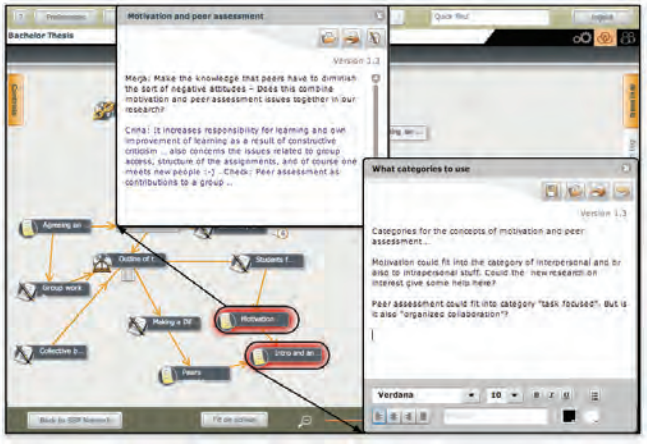


Fig. 2 Note editor GUI with two notes opened simultaneously

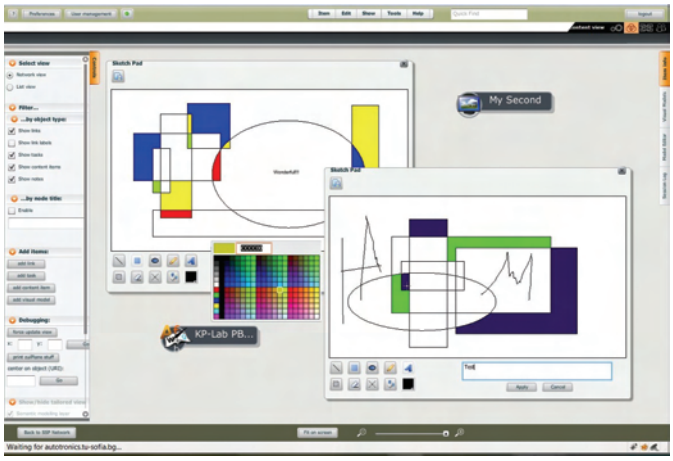


Fig. 3 Sketch pad GUI with opened sketches for editing

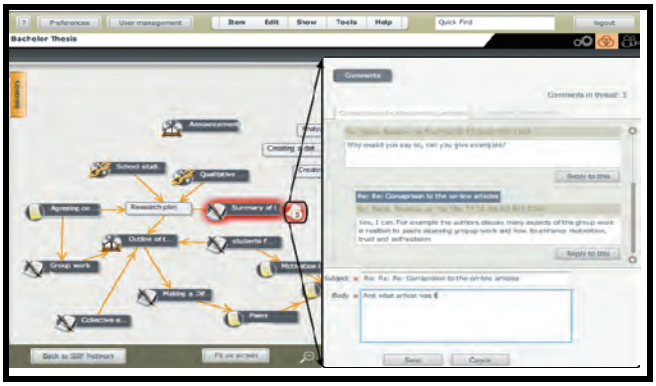


Fig. 4 Illustration of content-bound commenting

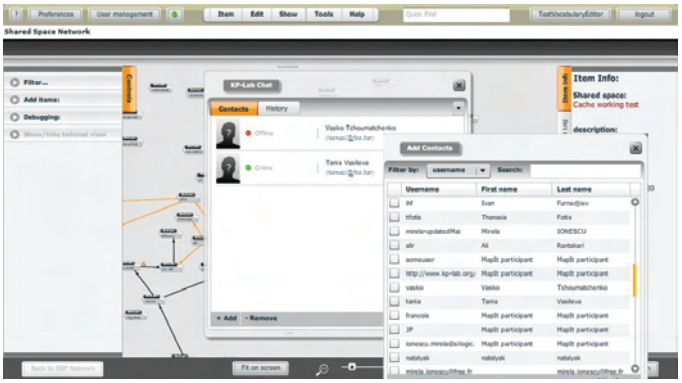


Fig.5. Chat tool integrated in KPE user interface

B. "Object-bound" interaction around knowledge artefacts

KPE answers the need to have individual contributions attached in collaborative work that is organized around shared knowledge objects embedded and embodied in a shared space. In the Content view, object-oriented collaboration is emphasized by content-bound commenting functionality that allows asynchronous, threaded discussions attached directly to knowledge objects (see Fig.4). Similarly, object-bound chat enables synchronous interchange attached directly in the content items at hand. Chat log is saved and linked to the targeted content item, therefore keeping the log attached to its object for possible re-use and continuation.

C. Process planning

In addition to content items, in KPE, users can explicitly define, modify and arrange *task items* to fit their process and domain, including, e.g., descriptors of title, responsible users, start and end dates, and status. In the Content view, task items can be presented, linked and arranged in the same visual representation together with the content items, which provides users with a holistic view of their knowledge creation processes, without separating tasks from contents (see the left screen shot in Fig. 6). The same tasks that are displayed in the Content view with their relations to content items can be viewed in *Process view*, presently in the form of a Gantt chart (see the right screen shot in Fig. 6).

Fig. 6 Right: Spatial arrangement of knowledge artefacts in the Content view including content items (black), tasks (grey) and labeled relations. Left: The Process view presenting the same tasks in a Gantt chart.

The Process view enables users to plan tasks and processes in a chronological manner as well as to monitor how the required tasks have been accomplished. For instance, in courses that teach collaborative design practices, where real design projects are executed, it is highly important (for flexible adjustment of the process) that participants be able to monitor the progression of the project and modify the tasks. Again, interdependencies and mutual connections between the tasks defined in the Process view are in turn automatically converted by the system into graphical constructions representing these connections in the Content view.

D. Social relations around shared objects and processes

In KPE, social mediation is envisioned in functionalities that support users in maintaining their contacts and keeping up with changing information about other participants, as well as their relations to the shared processes and content items. Social mediation provided by the tools allows users to lean on each others' competencies, expertise and experience and help them align their thoughts and actions with those of others.

KPE offers means to keep in contact with others, such are asynchronous commenting possibilities, or general chat (Fig.5) and context bound chat to enable synchronous discussions. Chat tool [4] aims to facilitate networking and community building through contextualized synchronous user communications when working on shared objects. Awareness features include clues and notifications of participants' status or past and present activities

E. Collaborative Semantic Modelling

One aspect of KPE that goes beyond current learning environments is the use of metadata and semantic features to support the usage and integration of knowledge artifacts in various ways. Tags and tag vocabularies can be created and edited by participants, or can be taken from ready-made vocabularies.

The KP-Lab system is ontology driven and provides a platform to develop advanced semantic tools for collaborative learning. The first prototypes of a semantic multimedia

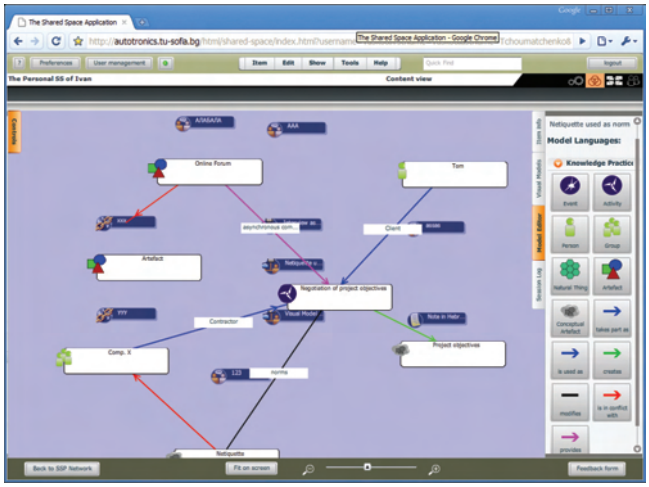


Fig. 7 Graphic User Interface of the Visual Model Editor

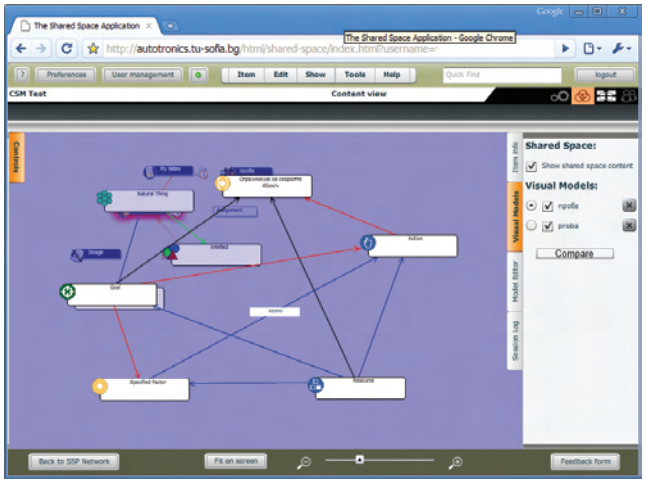


Fig.8. Comparing Visual Models in Visual Model Editor

annotation tool and an editor for visual models and visual modeling languages are released for field trials.

The Visual Modeling Language Editor [5] provides an extension to the basic functionalities offered by the KP-Environment and allows users to create, share, use, and update visual models as well as the underlying visual modelling languages as another type of shared artefacts.

The Visual Modeling Language Editor consists of two core components: the Visual Model Editor (VME), which allows users to create, compare and update diverse visual models and the Visual Modelling Language Editor (VMLE), which allows user to define and revise the underlying modelling languages and hence to specify the semantics of the models. Moreover the VM(L)E environment supports semantics at the Visual Model (VM) level, so that a model can be described and evolve as any other shared knowledge artifact. The semantics are accessible to the user by means of the respective visual modelling languages.

Fig. 7 depicts the user interface of the KPE Visual Model Editor. The Model Editor tab contains a library of the visual modeling language elements. Fig.8 illustrates VME ability to compare visual models. The direct integration into the KPE allows for an easy transition between modeling and other collaborative activities.

All KP-Lab tools are based on a foundational data model that provides common semantics for the tools and platform services. It is extended by the tool ontologies in order to describe the more specific semantics required.

III. Conclusion

The paper presents the KP-Lab System, a web-based flexible and extensible environment with a cluster of interoperable applications to foster collaborative knowledge creation and continuous work around shared objects. The system provides a coherent set of tools in order to initiate and organize collaborative processes, to support asynchronous and synchronous communication among participants, and to create, work on, share, and organize documents and artifacts collectively.

KP-Lab system is aimed to support both students as well as practitioners in their working and learning activities. The prototype and set of tools described in the paper are currently in use and under evaluation by several academic and professional partners throughout Europe and show promising results. Especially KPE flexibility, generic structure and integration of different tools for use, without loosing the semantics and metadata of the shared object under work, make it applicable to a broad range of pedagogical and professional scenarios.

Acknowledgement

The reported work is carried out within the KP-Lab project funded by the EU 6th R&D Framework program. We thank all our colleagues and partners involved in the design, implementation, and evaluation of the KP-Lab System and integrated tools.

References

[1] The KP-Lab website. ICT-27490: Knowledge Practices Laboratory (KP-Lab), Available:www.kp-lab.org/

[2] Markkanen, H., Holi, M., Benmergui, L., Bauters, M., & Richter, C. The Knowledge Practices Environment: a Virtual Environment for Collaborative Knowledge Creation and Work around Shared Artifacts. In J.Luca & E.R. Weippl (Eds.) Proceedings of ED-Media 2008, Vienna 2008, pp. 5035-5040.

[3] Vasileva, T., I. Furnadzhiev, V. Tchoumatchenko, Lakkala, M., Bauters M., Tools Facilitating Idea Generation in Knowledge Practices Environment, Proceedings of ICEST 2009,Veliko Tarnovo 2009, Bulgaria, v.2, pp. 421-444

[4] Furnadzhiev I., V. Tchoumatchenko, T.Vasileva, M.Lakkala & M. Bauters, Tools For Synchronous Communications in Collaborative Knowledge Practices Environment (KPE), In A.Mendez-Vilas, A. Soprano Martin, J. Mesa Gonalez (Eds) Research, Reflections and Innovations in Integrating ICT in Education, vol. 1, 2009 pp. 588-592

[5] Tchoumatchenko V., T. Vasileva, C. Richter, H. Allert ,Visual Model Editor Design and Implementation, IADIS International Conference on Cognition and Exploratory learning in Digital Age (CELDA 2009), Rome, Italy, 2009, pp. 307-310

Programming Environment for Management of Parallel Jobs with Concurrent Access to a Common Data Source

Ivaylo Penev[1]

***Abstract* – The paper presents a strategy and a programming environment for effective realization of parallel jobs with concurrent access to a common data source. Each job is started on a separate workstation in a local computer network. In order to avoid the conflicts, caused by the simultaneous access to the data source and to achieve better performance time, an approach with delay start of the jobs is proposed.**

***Keywords* – Parallel jobs, Concurrent jobs, Simulation problems.**

I. INTRODUCTION

In the area of banking information systems a class of problems, associated with simulation of many financial portfolios exists. In general each problem passes through reading data about portfolios from a data source (for example database), executing simulation calculations according to the performed analyses, and storing results in the same data source. The calculating procedures about each portfolio are completely independent each other. The whole simulation process is divided to jobs, each one simulating a separate portfolio. Prerequisites about realizations of such a class of problems in parallel computing environments exist. The efficiency of the realizations is significantly decreased due to the concurrent access of many parallel jobs to the common resource. The current work presents a strategy for preventing the concurrent access by proper shifting of jobs along the time axis of execution. The strategy is practically realized within a programming environment for management of the parallel jobs.

II. REALIZATION OF A PORTFOLIO MANAGEMENT SYSTEM IN A DISTRIBUTED COMPUTING ENVIRONMENT

The researches, presented in the current work, are based on the realization of a Portfolio management system (PMS) in a distributed computing environment [2]. The system simulates financial portfolios, calculating their positions. Information about each portfolio is stored in a database. When simulation about a specific portfolio is started, relevant data from the database are retrieved and suitable calculations are performed. The simulation finishes with storing report data in the database. Because of the great number of portfolios, positions and the various analyses performed, the simulation is heavy calculating process, taking long time to be finished on a single computer. Therefore the PMS executable is started on multiple connected computers, each of them performing simulation of a separate portfolio and saving the results in the database.

For managing the environment of computers (calculating resources, nodes) the system Condor is used, developed and spread for free by the University of Wisconsin-Madison, USA [3]. The distributed environment consists of a pool of computers and a server for database management (Oracle server), storing data about financial portfolios.

Each job has a description (in a submit description file) matching a portfolio simulation with a computer from the pool.

An execution of four jobs in series by a single computer and in parallel by the presented environment is tested. The time results differ from the expected. The increase of the count of the computers, participating in the simulation, decreases the parallel execution efficiency. The database management server is unable to process lots of simultaneous queries. Therefore, the concurrent access of the jobs to the common resource is a strong limiting condition over the execution time of the whole process. Applying a strategy for avoiding this limitation is necessary.

III. STRATEGY FOR MINIMIZING THE CONCURRENT ACCESS OF PARALLEL JOBS TO A COMMON DATA SOURCE

Each job could be decomposed to the following stages (simulation steps):

- Reading data about the positions of a financial portfolio from the database;
- Executing calculations over the data to perform analyses;
- Storing results from the simulation into the database;

The time of each job's stage is known in advance, according to the performed analyses.

[1]Ivaylo P. Penev is with the Faculty of Computer Science and Engineering, Technical University of Varna, 9000 Varna, Bulgaria, E-mail: ivailopenev@yahoo.com

The stages of reading and storing data from/into the database take time, which is significantly less than the time, necessary for performing calculations.

These observations give reason to apply a strategy, which introduces shifting of the jobs on the time axis. The aim is avoiding the recovering of the stages of different jobs, accessing the common resource simultaneously.

A. Formal description

The strategy is formalized by the help of theory of sets [1].

A set of simulation jobs is defined:

$$JOBS = \{job_1, job_2, ..., job_n\} \quad (1)$$

Each job has time for reading data τ_{READ}, time for calculations $\tau_{EXECUTE}$ and time for storing data τ_{STORE}. The times τ_{READ}, $\tau_{EXECUTE}$, τ_{STORE} are known in advance.

The execution of the whole process, including all the simulation jobs along the time axis is decomposed to a set of time intervals:

$$TIMES = \{t_1, t_2, ..., t_k\} \quad (2)$$

The stages, through which the execution of a job passes through, define a set of possible states, taken by the job in a time interval:

$$STATES = \{\mathrm{Re}\,ad, Execute, Store\} \quad (3)$$

The process of the execution of all jobs is described by a relation of the Cartesian product of the upper sets:

$Entire\,\mathrm{Pr}\,ocess = \{\langle t_i, j_j, s_l \rangle\}$, where:

$t_i \in TIMES, j_j \in JOBS, s_l \in STATES$.

When the jobs are performed in series, the total execution time of the whole process is given as a sum of the execution times of each job's stages:

$$t_{\sum} = \sum_{i=1}^{k} t_i$$

In the case of parallel performance the total time is theoretically equal to the maximum execution time of a parallel job:

$t_{\sum} = \max(\sum t_k)$, where t_k are all time intervals, during which a job from the set $JOBS$ is performed.

The practical increase of $t_{\sum}$ is caused by those time intervals, in which different jobs are set at $\mathrm{Re}\,ad$ or $Store$ state at the same time. These intervals form a subset of the $Entire\,\mathrm{Pr}\,ocess$ set.

$$ConflictJobs \subset Entire\,\mathrm{Pr}\,ocess$$

$$ConflictJobs = \left\{ \langle t_i, j_j, s_l \rangle, \langle t_i, j_k, s_l \rangle \right\} \quad (4)$$

, where

$$j \neq k \wedge (s_l = \mathrm{Re}\,ad \vee s_l = Store))$$

t_i - A time interval of the whole simulation process execution ($t_i \in TIMES$)

j_j, j_k - Jobs, executed in the t_i interval ($j_j, j_k \in JOBS$)

s_l - State of the jobs j_j, j_k in the interval, which is $\mathrm{Re}\,ad$ or $Store$ ($s_l \in STATES$)

The total time of the simulation process for the case of parallel execution is in inverse proportion to the count of the elements of the (4) set, i.e. to the set power:

$t_{\sum} \rightarrow f(|ConflictJobs|)$, where $|ConflictJobs|$ - power of the set.

The elements of the (4) set present those time intervals, in which different jobs are situated at the same state of access to the common resource at the current time interval. The following strategy is applied to decrease the number of these elements.

A complementary state $WAIT$ is added to the $STATES$ set:

$$STATES = \{Wait, \mathrm{Re}\,ad, Execute, Store\}$$

This state is added as an initial state at the beginning of each job, accessing the common resource. The job remains in this state until the other jobs occupy the $\mathrm{Re}\,ad$ state. After the common resource is released, the waiting job moves to the $\mathrm{Re}\,ad$ state. The job start is shifted along the time axis until the release of the data source.

As a result the $ConflictJobs$ set is shrunk in contrast to the $Entire\,\mathrm{Pr}\,ocess$ set, which is expanded with new elements. The results derive from the simulation process analysis, which shows that the concurrent access to the common resource is the strongest limiting condition over the efficiency of the whole simulation process.

B. Algorithm for realization of the proposed strategy

The strategy is realized using the following iterative algorithm for the execution of a simulation process, consisted of n parallel jobs:

1. Reading the execution times of the stages of each job.
2. Simulating a scenario with shifting the current job $i, 1 \leq i \leq n$:

2.1. Reading the times $\tau_{READ}, \tau_{EXECUTE}, \tau_{STORE}$ for job $k, 1 \le k \le n, k \ne i$.

2.2. Introducing delay time $\tau_{iWAIT} = \tau_{kREAD}$ for the job i.

2.3. Counting the intervals, in which the jobs access the common resource concurrently after delaying the job i.

2.4. Reading times for the next job k, transition to 2.1.

3. Simulating a scenario with shifting the next job i, transition to 2.

4. Final evaluation – estimating the scenario with minimum recovering of the common resource access.

IV. Programming environment, realizing the proposed strategy

The described strategy is presented by a programming environment for managing the execution of parallel jobs, accessing common data source.

The environment has two main parts:

- Visualizing the parallel execution of jobs, the recovering of states, the scenarios with shifting of jobs;
- Creating a description for each job for starting in the distributed computing environment (submit description file);

A. Part for visualization of the parallel jobs

This part performs the following main functions:

- Showing the times for each state of all jobs;
- Visualizing the parallel execution of the jobs by the help of diagrams;
- Showing the recovering of states of parallel jobs during the process of execution;

The times for the states of each job, participating in the simulation process, are defined in a data base. The parallel execution is visualized by diagrams, showing the development of the jobs by states. Furthermore, the whole simulation process time is divided to intervals. The state of each job in a time interval is specified, according to the known times for each job's states. The time intervals, in which different jobs occupy the data source simultaneously, are marked. These intervals are the limiting condition over the parallel execution efficiency (the elements of the set (4) from the formal description of the strategy).

By the help of the proposed interface scenarios with shifting of jobs are visualized. At the initial state the parallel jobs recover at the state of reading data from the data base (Fig. 1).

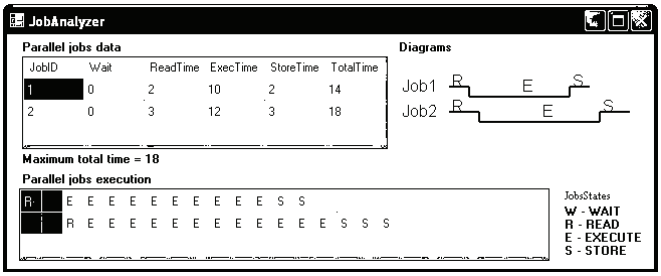


Fig. 1. Position of the jobs at the initial state

The described algorithm is applied. Scenarios with time shifts for each job are generated. The example presents managing of two parallel jobs.

The first job is shifted with two time intervals, until the second one finishes reading data and releases the common

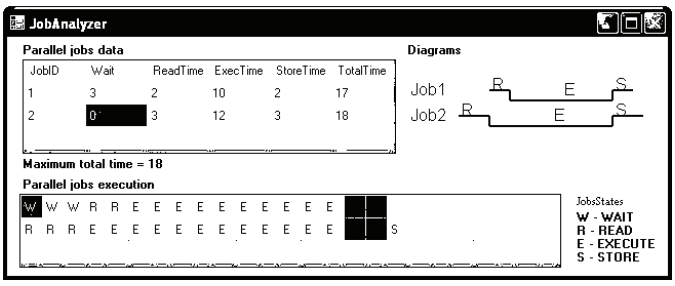


Fig. 2. Variant of job shifting

resource. The recovering of states shows, that after shifting of the first job, the parallel jobs become concurrent to the data source at the states of storing data. This means that the scenario would be ineffective for the whole simulation process and such ordering of jobs is improper (Fig. 2).

In the next scenario the second job is shifted with two intervals, until the first one ends reading data and releases the resource. The recovering of states shows, that the shift of the second job in relation to the first one prevents the concurrent access to the common data source. Although the maximal total time is greater than the same time in the previous scenario, this case is expected to avoid the limiting condition over the simulation efficiency. This should be the order of starting of the parallel execution in the distributed computing environment (Fig. 3).

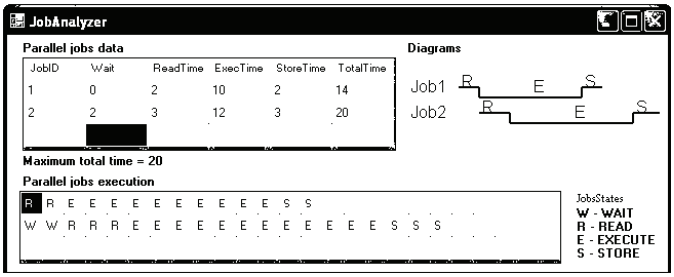


Fig. 3. Next variant of job shifting

B. Creating a description for each job for starting in the distributed computing environment (submit description file)

Each job must have a description to be started in the distributed environment. The description is stored in a file, called submit description file, which is created by the presented programming environment.

The file matches a computer from the pool with a portfolio from the database. The computer, which host name is specified in the description file, calculates the portfolio and stores the results into the database.

The description file has specified format, which includes information about the simulation execution and the work of the pool:

- Name of the executable, sent to the computers in the environment;
- Arguments – name of the portfolio to be calculated in background mode;
- Requirements – host name of the computer from the pool to calculate the portfolio;

V. CONCLUSIONS AND FUTURE WORK

Up to the current moment the presented strategy is in a process of development. The environment simulates small number of jobs. The purpose is building an effective model for researching the real execution of many parallel jobs with concurrent access to the common resource.

The formal description of the strategy will be complicated. After a job finishes, the computing resource is free for next job. Further analysis about the state of the other jobs is required. Complementary criteria will be defined, estimating the current results from the strategy appliance.

ACKNOWLEDGEMENT

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

REFERENCES

[1] L. Lovasz, K. Vesztergombi, “Discrete Mathematics”, Lecture Notes, Yale University, 1999.

[2] I. Penev, A. Antonov, "Realization of Portfolio Management System in a Distributed Computing Environment", International Scientific Conference Computer Science, Conference Proceedings, under print, Sofia, Bulgaria, 2009.

[3] http://cs.wisc.edu/condor/.

Creating Applications With PHP & AJAX

Ljupcho Apostolov[1]

***Abstract* – Using php scripting language and technology like AJAX make web interfaces more robust, faster and more user friendly.**

The project is about dropdown list that is part of web expert system, integrated system for e-medicine.

***Keywords* – php, ajax, javascript, XMLHttpRequest.**

I. INTRODUCTION

Ajax stands for **A**synchronous **J**avaScript **a**nd **X**ML. AJAX uses JavaScript to send and receive data between a web browser and a web server. It uses the JavaScript-based XMLHttpRequest object to send requests to the web server asynchronously, so the client won't have to refresh the page in order to get the results or upon clicking some button to be redirected to some other page.

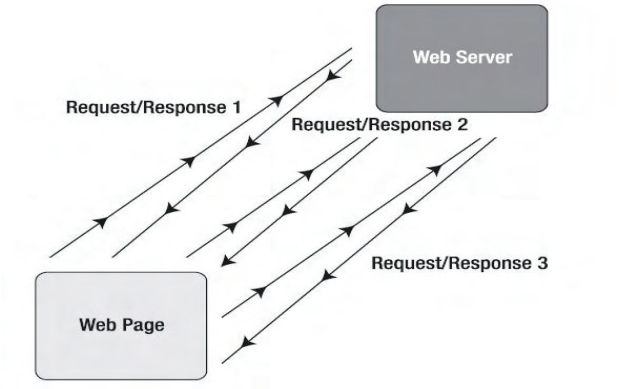


Fig. 1 - Request/response model using Ajax's asynchronous methodology

This way multiple server requests can be made from the page without need for a further page refresh.

[1] Ljupcho Apostolov is with the Faculty of Electrical Engineering and Information Technologies, Karpos 2 bb, 1000 Skopje Skopje,Macedonia,email: ljupcho.apostolov@gmail.com; web: www.ljupchoapostolov.com

II. FACING THE PROBLEM

The web expert system is a part of an integrated system for e-medicine that we are constructing for a developing information society. The main concepts that our system is based on include: creation of necessary Medical Information Systems (MISs) where hospitals had none; creating a framework and interfaces where various multiplatform MISs could interconnect in an integrated MIS; using modern telecommunication technologies for creating an integrated MIS and provision of advanced medical services (sharing knowledge, experience and expertise, enabling better remote patient-doctor communication).

The goal of the web medical expert subsystem presented here is to serve as a consultant to physicians when setting a diagnosis. Eventually the training is performed with the entire dataset so that the generated rules are based on the entire knowledge available. These final rules have significantly more impact in the deciding process when the system is in use.

The system has a modular design so that new classification forms can be added easily, when training data for the new disease is obtained. Modern web technologies (AJAX) are implemented to achieve fast and intuitive user interface.

I decided to make a form using AJAX and I was given a dataset of parameters for which I made the form with dropdown lists, so that the user can make a selection and regarding the condition made a decision for the final result, display it along with the parameters in a table.

In the following picture are presented two tabs, the first one without using ajax and the second (shown) with using ajax.

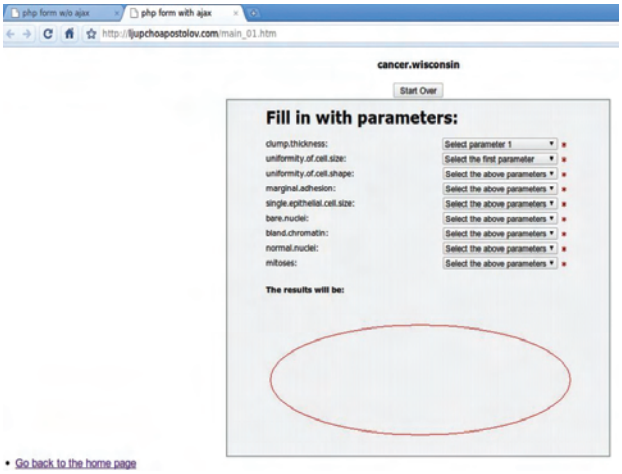


Fig. 2 – Web form with dropdown list based on AJAX

The marked red area is the place where the results will be displayed. That is a special div section in the html page where the response from the requested page is displayed, which was loaded into the innerHTML element of the passed-in object after request is sent.

```
document.getElementById("parametar10div").innerHTML=x
mlHttp.responseText;
```

The result is displayed after the last parameter is chosen. The result displayed looks like this:

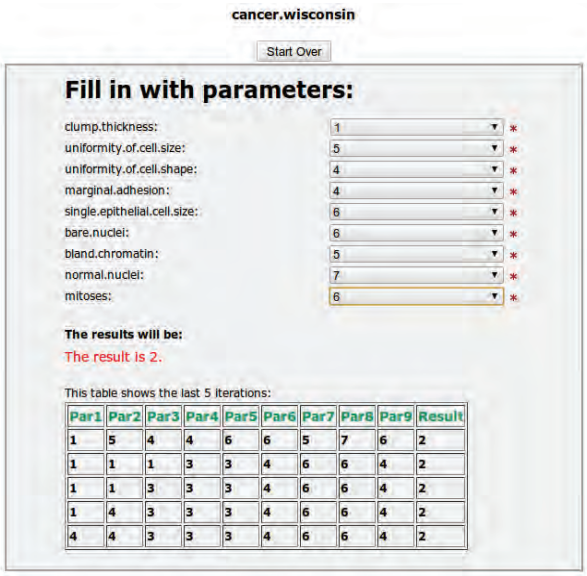


Fig. 3 – Displaying the final results

When I populate this form for the first time the results will be displayed after I choose the final dropdown, but after that if I change the 5th menu for example I will get the new updated table. Whatever I change, the table changes as well. Since I work with session variables foreach of the dropdown option values, I can kill the session, reset the form and populate it again from the beginning by pressing the "Start Over" button placed at the top of the form.

The difference between using dropdowns old fashion way and using AJAX is that with AJAX the choose of one dropdown will determine the list for the second menu and the choose of that list will determine the list of the next dropdown and so on. So, it changes dynamically and it depends regarding of what you are choosing.

Without using AJAX you can start choosing any dropdown you like and then go back to the first for example and implemented javascript will through you an error if a dropdown has not been selected. With AJAX we don't care about that since form validation is implemented differently with database and php as scripting language.

The in-deep explanation of the approach would be:

In my case I have dropdown lists that should have values from 1 to 10, so each of the selected should give me a menu (the next list) from 1 to 10. This example might not be the best to see the benefit of AJAX with dropdown lists or maybe it would be more useful for further development and maintenance when having different option values (in this case different number of option values to select by dropdown).

For better understanding let's presume that for the first dropdown list we have major cities in Macedonia, like Skopje, Bitola, Ohrid ect. Once I select Skopje I would get all the municipalities for example in the second menu, once I select one municipality I would get all the streets in that municipality in the third menu. This path would be totally different if I had chosen Bitola in the first menu or Ohrid, right! If I try to translate this into old fashion web form it would mean that I have Skopje, Bitola and Ohrid in the first dropdown, then all the municipalities for all the 3 cities in the second menu and all the streets for all the cities in the third menu - all of these lists would be in the html code as option values. That won't be practical from programming point of view as well as from user perspective, since it is a lot easier to choose the street from one municipality for you city.

III. Form Validation

One of the most important things about creating web forms is validation. We need to make sure that all the fields in the form are populated correctly and all the wrong scenarios are been considered, so we don't have wrong values inserted in the database.

Validation without AJAX means client and server side validation, so in case the client side has not succeeded to make the form valid, the server side will give an error at the top of the form saying that the user has to populate all the fields. The difference between the client side and server side is that the client side uses javascript and the server side is php code to take care of setting the session variables. Headers are used to make sure that the session has expired and that the cache is empty.

Validation using AJAX on the other hand has different approach and it's totally different then having the values in the html page. The values for the dropdown lists are pulled out from the database, or you won't be able to select any dropdown list unless you select the list before. This way we force the user to select the dropdown lists one by one starting from the top, so in the end we have all the values chosen and result is displayed. Selecting one by one is only mandatory when populating the form for the first time, after that any change of dropdown value displays the table with results.

Q: So, how do I make the values available in the rest of the lists?

A: I did an insert of those values in the database using other script, so for each selected value I can have a list of values from 1 to 10 for the next parameter.

```
for ($b=1;$b<11;$b++) {
    for ($i=1;$i<11;$i++) {
$sql="insert into `parameter2`(parameter1id,parameter2)
values ('$b','$i')";
$result = mysql_query($sql) or die("Invalid query: " .
mysql_error() . "<br><br>". $query);
    }
}
```

The first „for“ is for the previous menu and the second is for the next, that's actually the `where` clause in the sql statement. And I do this for parameter1 to parameter10. I made it like this so that selecting one parameter will trigger the javascript (named selectparametar.js) which will create XMLHttpRequest object; call the php page that will make the next dropdown list available for selection and it will display it in the div section specified on the main html page. So, I actually have only “div” in the html page and the option values menu has been drown with the php page, which is actually a mix of php code and html code.

IV. IMPLEMENTATION

The following project is uploaded on my page, so you can have a look : www.ljupchoapostolov.com

The new method, with AJAX consists of four ingredients:

- Database
- Java script file
- Html page
- Php page

If an item in the drop down box is selected the function in the javascript file executes the following:

1. Calls on the GetXmlHttpObject function to create an XMLHTTP object;
2. Defines the url (filename) to send to the server;
3. Adds a parameter (q) to the url with the content of the dropdown box;
4. Adds a random number to prevent the server from using a cached file;
5. Call stateChanged when a change is triggered;
6. Opens the XMLHTTP object with the given url;
7. Sends an HTTP request to the server;

The php page does:

1. PHP opens a connection to a MySQL server.
2. Sql query is made with the parameter selected in where clause.
3. The dropdown menu for the next parameter is created and the javascript function is called which will call the php page for next dropdown.

The method of onchange=”something(this.value)” for the selected dropdown will call the function in the javascript file and that one will create a new XMLHttpRequest to the server and make a select from the database for the next dropdown list.

Javascript controls the whole process here as it calls the php page to do the insert-ing values in the result table or selecting those values from the database and ones done with that, it specifies where that results will be outputted on the html page, in the specified div section, like

The php page has been called using the
xmlHttp.open("GET",url,true);
where url is
var url="./register_001.php";
url=url+"?parametar1="+parametar1id;

Here we process the parametar1 that we have selected from the dropdown menu and in the php page we use it in the where clause to make the sql query for selecting the next parameter.
The php page will have:
onchange="parametar03(this.value)">

This is when we have the list displayed and we select a value from the dropdown. What happens now is that this parametar03 is a function in the javascript file that calls the php page for making the options for the next menu available. (this.value) is what we choose from the list.

Making the options available, the actual drawing of the list is done with little html code within the php page:

<option value="<?php echo $row['parametar9'] ?>"><?php echo $row['parametar9'] ?></option> <?php } ?>

In this case its the 9th drowdown list, which is actually a column in the database table for the 9th parameter.

Selecting the last parameter will trigger the function from the js file that will call the php page that will display a result on the div section as specified in the html page and make us a table. The result will be loaded in the object and displayed in the html page and is actually “if clause” from the parameters. I had some dataset for those parameters and made a rule for the final result.

In this project the html page is only about css since we only have <option value=””></option> only for the first parameter and the rest will be made by the php files in the div sections we specify in the html.

I am operation with single php page which is executed (number_of_parameter) times through AJAX and I store the values in SESSION variables. That’s how I managed to make the results displayed after the form has been populated and a new change has occurred. Everything is reset after pressing the “Start Over” button which destroys the session.

V. CONCLUSIONS

This project briefly explains the use of AJAX for building web applications making the user interface more user friendly by staying on that exact page, not moving to another page with back links to get back to the form. Really cool features can be made using AJAX, not just the “staying on the same page”, but features like “auto complete” which is today widely used, like when you type some word in Google search or Youtube, it gives you suggestions. Another goodie would be photo gallery with creating dynamic thumbnails for previews.

VI. REFERENCES

[1] Beginning Ajax With PHP - From Novice To Professional (2006) – Lee Babin
[2] http://en.wikipedia.org/wiki/AJAX
[3] AJAX Open Directory Project

This page intentionally left blank.

Performance Assessment of Metrics for Video Quality Estimation

Zoran G. Kotevski[1] and Pece J. Mitrevski[2]

Abstract – **There are two types of metrics for measuring the quality of processed digital video: purely mathematically defined (DELTA, MSAD, MSE, SNR and PSNR) where the error is mathematically calculated as a difference between the original and processed pixel, and video quality metrics that have similar characteristics as the Human Visual System (SSIM, NQI, VQM), where the perceptual quality is also considered in the overall video quality estimation. In this paper, experimental comparison of the performance of PSNR, SSIM, NQI and VQM metrics is presented.**

Keywords – **Video quality, PSNR, SSIM, NQI, VQM**

I. INTRODUCTION

One of the simplest definitions of video quality is that the video quality is a state of perception by the Human Visual System (HVS) [1]. Hence, the best video quality estimation can be performed by trained human estimators, but in real world situations this is a huge problem and these video quality metrics are practical tool for fulfilling this complex task [9]. Basically there are two types of video quality metrics for measuring the quality of processed digital video. On one side there are purely mathematically defined video quality metrics like: DELTA, MSAD, MSE, SNR and PSNR [6], [11], [14], [15], where the error is mathematically calculated as a difference between the original and processed pixel. These metrics are more technical ones and because visual quality assessment is a task more complex than simple pixel error calculation, many consider their quality estimation to be deficiently accurate. On the other side there are video quality metrics that have similar characteristics as the Human Visual System like: SSIM, NQI, and VQM [3], [4], [5], where the perceptual quality is considered in the overall video quality estimation. This second group of metrics, beside the mathematical error calculation, also calculates the scene structure in the estimation of quality. The SSIM and NQI are known to have similar characteristics, because they calculate the quality by combining three components: luminance, contrast and structure or mutual characteristics. Differently from these two, VQM metrics uses DCT quantization to eliminate the spatial frequencies that are less visible to the human eye. In order to compare the performance of these four metrics, an experiment is conducted in which large number of differently processed video sequences are created and their quality is measured. The results are basic charts that present these metrics dependence to the most common changes in processed video i.e. changes in brightness, contrast, hue, saturation and noise. This paper pinpoints the key characteristics of each metric, gives the conclusion of the better performing one and gives some directions for improvement of objective video quality estimation.

[1]Zoran G. Kotevski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: zoran.kotevski@uklo.edu.mk

[2]Pece J. Mitrevski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: pece.mitrevski@uklo.edu.mk

II. BRIEF METRICS INTRODUCTION

A. Peak Signal to Noise Ratio (PSNR)

The PSNR parameter is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. PSNR is usually expressed in terms of the logarithmic decibel scale. Normally, higher PSNR indicates that the reconstruction is of higher quality. In ideal case the value of PSNR would be 100 dB, but in reality, in the field of image processing, typical values for PSNR are between 30 dB and 40 dB [9], [11].

$$PSNR = 10 \cdot \log_{10}\left(\frac{255^2}{MSE}\right) \text{ [dB]} \qquad (1)$$

$$\text{and } MSE = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\left[f(x,y) - g(x,y)\right]^2, \qquad (2)$$

According to the mathematical equations for calculating MSAD, MSE, SNR and PSNR, [6], [14], [15], can be inferred that they represent similar error values i.e. the calculated error is of the same degree. Because of this, PSNR can be considered as an unofficial representative of all the above mentioned video quality metrics and still the most widely used metric for video quality estimation in many video processing systems, especially in video compression systems.

B. Structural Similarity (SSIM)

The Human Visual system (HVS) is highly adapted to extracting the structural information from the area of viewing. SSIM metric uses this characteristic of the HVS in estimation of quality of the processed digital video [4].

Structural information of an image can be defined by those characteristics that represent the structure of the objects in the scene, independently of the mean brightness and contrast [3], [4]. These measurements are based on measurement of three

components: luminance comparison, contrast comparison and structure comparison.

$$S(x,y) = f(l(x,y), c(x,y), s(x,y)) \quad (3)$$

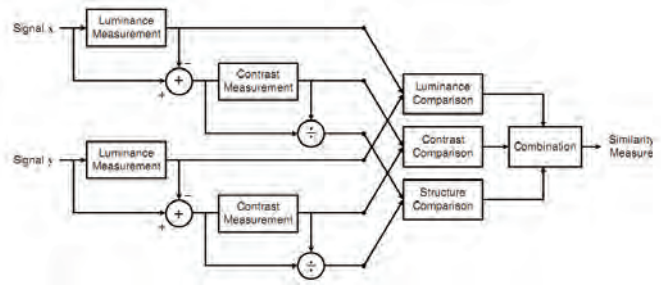


Fig. 1. Diagram of the SSIM measurement system

Structural similarity index is a combination of these separate components (3). The SSIM index can gain values from 0 to 1 where value of 1 represents maximum quality.

C. New Quality Index (NQI)

NQI works in a similar manner as SSIM index. NQI defines picture distortion as a combination of three factors: difference in mutual characteristics, difference in luminance and difference in contrast. Mathematical definition of the NQI can be found at [3].

The values of NQI span from 0 to 1, even though in some cases NQI can be lower than zero. Similar like SSIM the value of 1 represents maximum quality.

D. Video Quality Metrics (VQM)

Human eye sensitivity to spatial-temporal pattern decreases with high spatial and temporal frequency. Based on different sensitivity, high spatial or temporal information can be represented with less data and less precision, while human eyes are more or less insensitive to the loss of this information. This characteristic of HVS is exploited by DCT quantization, which is the base for VQM [5]. The values of VQM start from 0 and in real situations can reach around 12. VQM value of 0 represents minimum distortion and maximum quality. The system diagram of the VQM system is shown in figure 2.

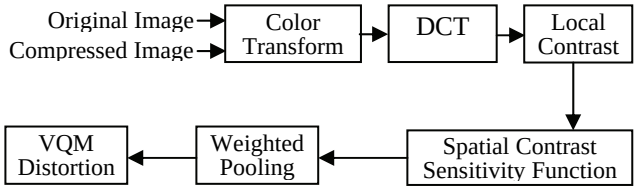


Fig. 2. Diagram of the VQM measurement system

III. Analysis of the Measurements of Modified Video Sequences

For the purpose of this analysis, at first three video sequences were created. In all three videos (Old boat, Sea view and Mountains), static picture in duration of five second is presented. All these three video sequences were made with different scene structure in order to determine if scene structure has some influence in these measurements.

Also for the purpose of creating the diagrams of metrics dependence to the changes in processed video sequences, more than 300 short video sequences were produced, each with different amount of introduced changes and effects, in order to illustrate the influence of more or less visible video deformation to the performance of these metrics. The most common changes that do not highly influence the viewer's quality of experience are changes in brightness, contrast, and saturation. In other video sequences, highly destructive video deformation like Gaussian noise is introduced.

All video sequences were produced with trial version of Sony Vegas Pro v8.0c [16], coded in Main Concept's MPEG-2 coder, main level and profile, with average bit rate of 4MBit/sec. Measurements were performed with the trial version of Elecard Stream Eye Tools v2.9.1 [13].

After the performed measurements, the charts of metrics dependence to different changes were drawn and some of them are presented below. These charts were visually evaluated by dozen estimators and the comments are summarization of their opinion.

Because of the limitations of this paper only the most relevant charts are presented. All the charts and images of differently processed video sequences are publicly available at http://vq.heliohost.org.

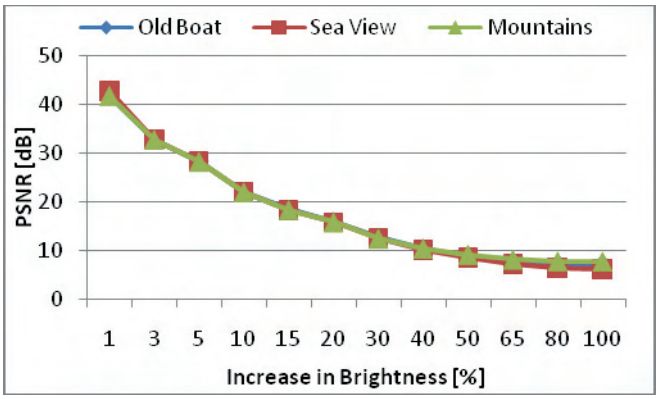


Fig. 3. PSNR decrease due to changes in brightness.

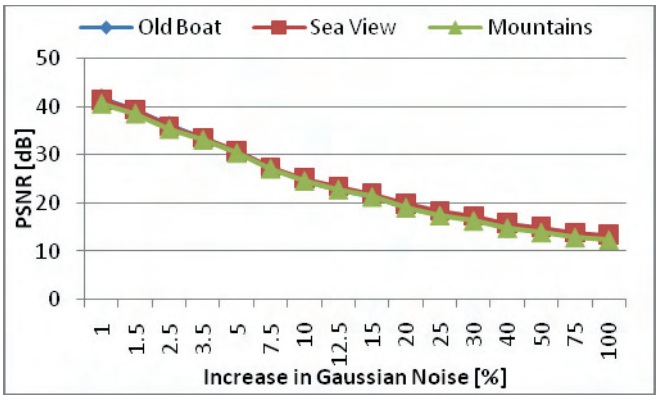


Fig. 4. PSNR decrease due to the amount of Gaussian noise.

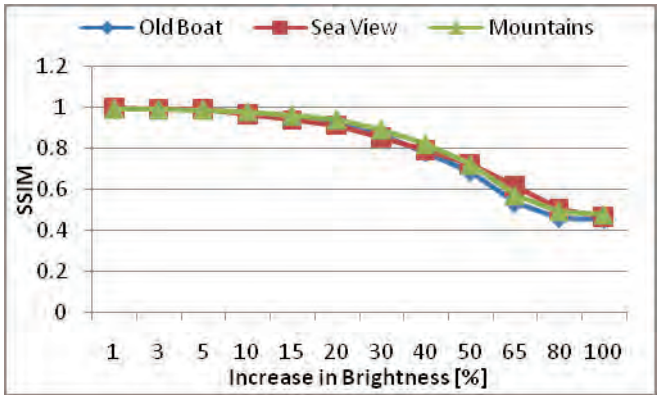


Fig. 5. SSIM decrease due to changes in brightness.

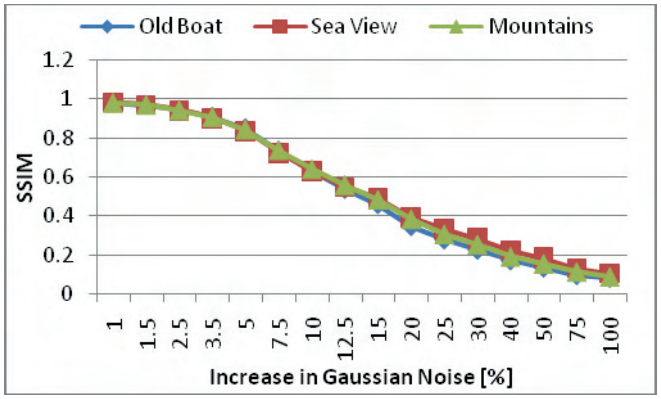


Fig. 6. SSIM decrease due to the amount of Gaussian noise.

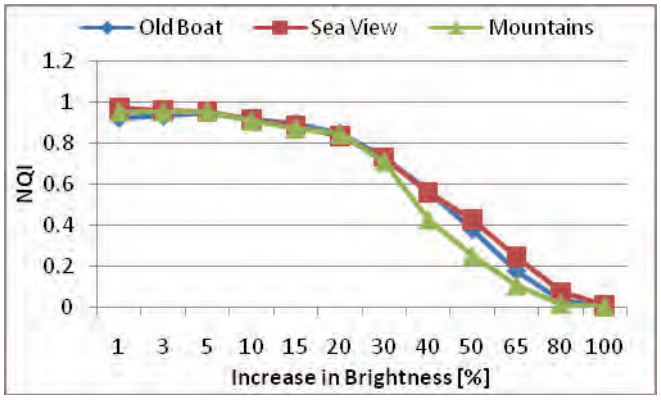


Fig. 7. NQI decrease due to changes in brightness.

After performed analysis of these charts, it can be concluded that the most drastic decreases in PSNR values are due to changes in brightness, as shown in Fig. 3, and combination of changes in brightness and contrast. The second influential factor is the introduced Gaussian noise, as shown in Fig. 4. The changes in hue and saturation have medium effect in decreasing PSNR value. These charts clearly describe the deficiencies of PSNR.

Considering the performance of SSIM, the most drastic decreases in SSIM values are due to the amount of introduced Gaussian noise (Fig. 6). Changes in brightness (Fig. 5), contrast and hue have only mild influence to the overall quality estimation. These characteristics of SSIM speak of certain similarities to the HVS and present solid background for more realistic platform for quality estimation of processed digital video, but it is also obvious that there are some imperfections present.

NQI index reacts most drastically to changes in Hue and increase in Gaussian Noise as shown in Fig. 8. Changes in brightness as shown in Fig. 7, or saturation have even less influence to the NQI index compared to the SSIM index. It is easily noticable that NQI reacts less than SSIM to changes in brightness or saturation, but reacts more drastically to changes in hue or Gaussian noise. These characteristic of NQI clearly speak of some deficiencies that NQI has compared to SSIM index.

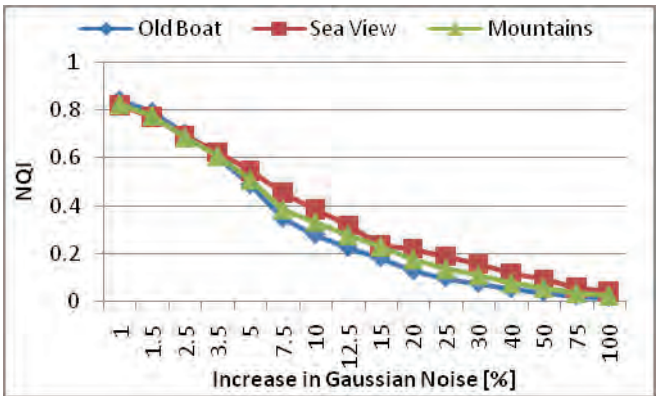


Fig. 8. NQI decrease due to the amount of Gaussian noise.

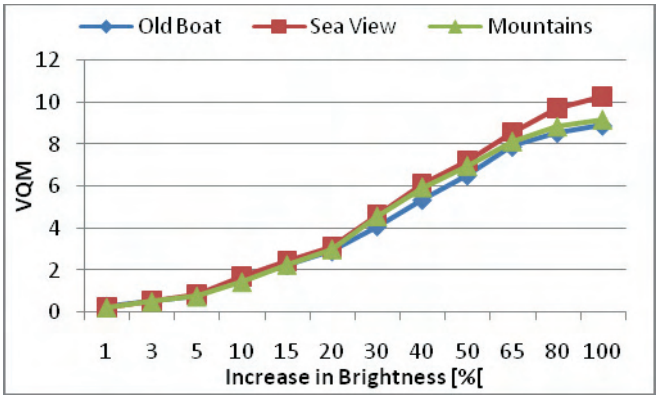


Fig. 9. VQM increase due to changes in brightness

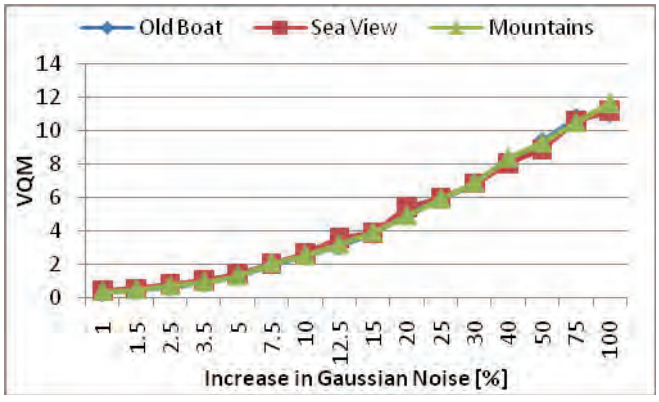


Fig. 10. VQM increase due to the amount of Gaussian noise

VQM index reacts the most to changes in the amount of Gaussian noise as shown in Fig. 10. But, the key concluding elements from these charts are that VQM is not so insensitive to changes in brightness (Fig. 9) or changes in contrast, and opposite to SSIM and NQI shows quite good sensitivity to changes in hue and saturation. SSIM's and NQI's issues of low sensitivity to changes in brightness/contrast and hue contribute to lower performance in quality estimation compared to VQM metric. This advantage of VQM over all other metrics can be observed in most of the examples.

IV. Conclusion

Given the examples, it can easily be concluded that PSNR metric is not valid enough to be used as objective measurement for video quality estimation. There are too many parameters that highly influence the PSNR value that are of minor visual influence to the viewer's perception of quality. Changes in brightness and contrast have high influence to the PSNR that in most cases causes decrease to the performance of this metric. To be more precise, PSNR metric can be taken as valid measurement in some cases if PSNR value is greater than 35 dB. Everything below this degree of PSNR cannot be considered valid because the origin of PSNR decrease is unknown in most cases and the results given by this metric can be misleading.

SSIM metric has quite better performance compared to PSNR and in most cases performs very similar to the Human Visual System. But, imperfections are also present. SSIM is almost insensitive to changes in brightness, contrast and hue that when these changes are bigger SSIM values can become largely inverted. However, many examples indicate good SSIM similarity to HVS and with some small improvements in mentioned areas SSIM performance can be enhanced.

If we compare SSIM to NQI, even though they calculate the quality in a similar manner, they differ in some points quite significantly. From all the examples can be concluded that SSIM metric performs quite better than NQI in most of the situations. Through the analysis of the presented charts it can be concluded that changes in brightness have mild influence to the SSIM index and even milder to the NQI index. Changes in Hue cause steep decrease in NQI but changes in saturation are almost invisible to it. These characteristics of NQI contribute to lower performance compared to SSIM and VQM.

VQM metric on the other hand performs quite better than all previous mentioned metrics in almost all situations. Through the analysis of the presented charts it can be concluded that VQM metric reacts to changes in brightness in a similar manner as HVS which can be considered as VQM's advantage. Changes in hue and saturation are rated in a good manner by VQM, compared to the SSIM and NQI where they are almost invisible to them. The analyzed charts and videos lead to a conclusion of certain imperfections of VQM metric also, but it can be concluded that VQM performs quite well. The overall conclusion is that VQM metric gives best performance of all of these analyzed metrics. The advantage that VQM has is that it exploits the DCT transformation and quantization technique, similar like coding techniques, where high spatial frequencies are omitted as being less visible to the human eye. This characteristic enables VQM to mostly consider the changes that are more noticeable to the human eye, which is the key to creating a better video compression system as well as creating better video quality estimation system.

If better video quality estimation metric is to be created, one must explore HVS behavior first. Drawing similar charts of HVS dependence to changes in brightness, contrast, hue, saturation and noise would be a challenging task, but such charts would be a great foundation for creating video quality metric that would resemble the HVS perception of quality.

Concerning the scene structure and its influence to these measurements it can be concluded that scene composition barely influences these measurements and can be considered as non influential factor.

V. References

[1] Z. Kotevski, *Analysis of Quality and Performance of MPEG-2 Video Compression Techniques*, Master Thesis, Faculty of technical sciences, Bitola, Macedonia, December 2007.

[2] Z. Kotevski, P. Mitrevski, "Experimental Comparison of PSNR and SSIM Metrics for Video Quality Estimation", Proc. of ICT Innovations-09 International Conference, Macedonia, September 2009.

[3] Z. Wang, Student Member, IEEE, A. C. Bovik, Fellow, IEEE, "A Universal Image Quality Index," IEEE Signal Processing Letters, March, 2002.

[4] Z. Wang, Member, IEEE, A.C. Bovik, Fellow, IEEE, H. R. Sheikh, Student Member, IEEE, E.P. Simoncelli, Senior Member, IEEE, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, April, 2004.

[5] F. Xiao, "DCT Based Video Quality Evaluation," Final Project for EE392J, 2000.

[6] Z. Wang, A.C. Bovik, "Mean Squared Error: Love It or Leave It?," IEEE Signal Processing Magazine, January, 2009.

[7] Z. Wang, X. Shang, "Spatial Pooling Strategies for Perceptual Image Quality Assesment," IEEE International Conference on Image Processing, Atlanta GA, October, 2006.

[8] Z. Wang, Q. Li, "Video Quality Assesment Using a Statistical Model of Human Visual Speed Perception," Journal of the Optical Society of America, 2007.

[9] P. Symes, *Digital Video Compression*, McGraw-Hill, ISBN 0-07-142487-3, 2004.

[10] C. A. Poynton, *A Technical Introduction to Digital Video*, John Wiley & Sons, ISBN 0-471-12253-X, 1996.

[11] A. Bovik, *Handbook of Image and Video Processing*, Academic Press, ISBN 0-12-119790-5, 2000.

[12] C. A. Poynton, *Digital Video and HDTV – Algorithms and Interfaces*, Morgan Kaufmann Publishers, ISBN 1-55860-792-7, 2003.

[13] Elecard, http://www.elecard.com/

[14] Compression Project, http://www.compression.ru/

[15] Compression Links Info, http://www.compression-links.info/

[16] Sony Creative Software, http://www.sonycreativesoftware.com/

PageRank Algorithm Overview

Ilija T Jolevski[1], Gjorgi K. Mikarovski[2] and Aleksandar N. Kotevski[3]

***Abstract* – This article introduces and discusses the concept PageRank link analysis algorithm, brief history, which factors do have an impact on PageRank, Based on research to several web portal, guidelines are provided on how to optimize web page context for search engines. In addition, we discuss for bad practice for over-optimizing articles posted to web page.**

***Keywords* – PageRank, optimize, SEO, algorithm**

I. INSTRUCTIONS FOR THE AUTHORS

Authors have two main goals: articles to be available for search engines users and where articles are displayed in the results list – articles which are displayed in top positions are more likely to be read. They should have an interest in ensuring that their articles are indexed by search engines like google, yahoo, ask, msn because they have two main goals: articles to be available for search engines users and where articles are displayed in the results list – articles which are displayed in top positions are more likely to be read. The most search engines use crawlers to find pages for their algorithmic search results. Pages that are linked from other search engine indexed pages do not need to be submitted - they are found automatically. Usually used link analysis algorithm is pagerank – used by google search engine. Exploiting the hyperlink structure of the web, pagerank surmises that each web page has a prestige score that ties to the number of in-links the page receives. It work based on assigns a numerical weights to each member of set of hyperlinked elements of documents. Assigned numerical weight that is assign to element e is also known as pagerang of e and denoted by pr(e). Pagerank can be calculated for any collections of documents of any size. With pagerang view, some page a has a higer pagerang than another page b, even though it has fewer links to it – the link it has is of a much higer value. A hyperlink to a page counts pagerang detect as a vote of support. So, if some page is linked to many pages with high pagerang, that it received a high rang itself. Contrary, of there are no links to a web page there is no support for that page.

[1]Ilija T Jolevski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: ilija.jolevski@uklo.edu.mk

[2]Gorgi K. Mikarovski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: gjorgi.mikarovski@uklo.edu.mk

[3]Aleksandar N. Kotevski is with the Faculty of Low Sciences, Prilepska bb, 7000 Bitola, Macedonia, E-mail: aleksandar.kotevski@uklo.edu.mk

II. HISTORY OF PAGERANG ALGORITHM

PageRank was invented at Standford University in 1995 year, by Larry Page and Sergey Brin, as a part of research project for developing a new kind of search engine. The first publication about this project, describing PageRank and the initial prototype of the Google search engine, was introduced in 1998 year. Shortly after, Google Inc. was founded to manage the Google search engine, which used PageRank. Currently the trademark PageRank belongs to Google Inc., but the original patent for the PageRank algorithm is assigned to Stanford University.

The PageRank algorithm is still one of the factors taken into account in Google's search engine result ranking, and it is constantly followed by interested search engine marketing specialists around the world.

III. WHAT IS PAGERANG ALGORITHM

PageRank is algorithm formulated by Sergey Brin and Larry Page, and today it is used by Google search engine. According to PageRank algorithm, importance of any web page can be define as number of links pointing to it from other web pages. Google uses PageRank to adjust results so that sites that are deemed more "important" will move up in the results page of a user' s search accordingly.

PageRank can be understand as number assessed solely the voting ability of all incoming links to a page, and also, how much they recommend that page. Every unique page of every site that is indexed in Google has own PageRank and it is one of the most important ranking techniques used in today's search engines. Not only is PageRank a simple, robust and reliable way to measure the importance of web pages but it is also computationally advantageous with respect to other ranking techniques in that it is query independent, and also content independent.

IV. PAGERANG CALCULATION

The basic PageRank's idea is if one web page A has a link to another web page B, then the author of A is implicitly conferring some importance to web page B. The first model for PageRank uses structure of links of the web pages to construct Markov chain with primitive transition probability matrix M. Irreducibility on chain is guarantees for existing of PageRank vector. If there is one set from x web pages and one matrix M. PageRank algorithm as first constructs new matrix N by renormalizing each row of matrix M to sum 1. If there is some web surfer who visit some web page, and decides for browsing to next web page. There are two possibilities: user

should click on hyperlink that is on current page, with probability 1-e, or, with probability e, user should enter new URL address at browser. In this case, e is parameter, set to 0.1-0.2. This process is Markov chain on the web pages, with transition matrix eU + (1-e)N. In formula, U is transitions matrix. R, the scores on vector of PageRank must to be defined to be the stationary distribution of this Markov chain. Also, R can be define as

$(eU + (1 – e)M)^T r = r$

After this, authoritativeness of page i is the asymptotic chance of visiting page.

Actually, PageRang algorithm is elegantly simple, and is calculation is following:

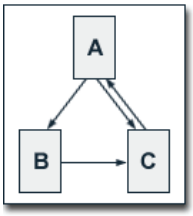
PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))

In this calculation, PR(A) is the PageRank of a page A, PR(T1) is the PageRank of a page T1, C(T1) is the number of outgoing links from the page T1, d is a damping factor in the range 0 < d < 1. D is usually set to 0.85.

It is natural to wonder what is the best value of the damping factor, if such a thing exists. In a way, when ® gets close to 1 the Markov process is closer to the "ideal" one, which would somehow suggest that ® should be chosen as close to 1 as possible. This observation is not new, but it has some naivety in it.

There are two ways in which this algorithm can affect to position of page on Google. The number of incoming links is important for web rating – the more of these links has positive effect. If some page hasn't incoming links, it can have a negative effect of web ratings. Contrary to incoming links, more outgoing links has negative effect of page rang.

Of there is some web consisting of three pages, A,B and C. Here, page A links to the pages B and C,page B links only C, and page C links only page A.



For this example, we have following calculation if we assume that d factor is 0.5:

PR(A) = 0.5 + 0.5 PR(C)

PR(B) = 0.5 + 0.5 (PR(A) / 2)

PR(C) = 0.5 + 0.5 (PR(A) / 2 + PR(B))

These equations can easily be solved. We get the following PageRank values for the single pages:

PR(A) = 14/13 = 1.07692308

PR(B) = 10/13 = 0.76923077

PR(C) = 15/13 = 1.15384615

From calculation, is obvious that sum of PageRanks of all pages is 3- that is equals the total number of web pages. Here is easy for calculation because of very small structure of web page. But, today web consist of a huge number of documents, and practically

It is obvious that the sum of all pages' PageRanks is 3 and thus equals the total number of web pages. As shown above this is not a specific result for our simple example.

For our simple three-page example it is easy to solve the according equation system to determine PageRank values. In practice, the web consists of billions of documents and it is not possible to find a solution by inspection. Therefore, Google search engine uses iterative computation of PageRank values. Iteratuve computation of PageRank means that each page is assigned an initial starting value. PageRanks of all pages are then calculated in several computation circles based on the equations determined by the PageRank algorithm. For the same example, the iterative calculation look like:

Iteration	PR(A)	PR(B)	PR(C)
0	1	1	1
1	1	0.75	1.125
2	1.0625	0.765625	1.1484375
3	1.07421875	0.76855469	1.15283203
4	1.07641602	0.76910400	1.15365601
5	1.07682800	0.76920700	1.15381050
6	1.07690525	0.76922631	1.15383947
7	1.07691973	0.76922993	1.15384490
8	1.07692245	0.76923061	1.15384592
9	1.07692296	0.76923074	1.15384611
10	1.07692305	0.76923076	1.15384615
11	1.07692307	0.76923077	1.15384615
12	1.07692308	0.76923077	1.15384615

From computation above, we can conclude that real PageRank values can be compute in only few iterations. Appropriation, and according of Lawrence Page and Sergey Brin, nearly 100 interactions will compute a good approximation of the Page Rang values of the whole web.

How is PageRang determinate?

In situation from 2 web page, A and B, if A links to page B, then Page A is saying

that Page B is an important page. Also, if a page has important links pointing to it, then its links

to other pages also become important, irrespective of the actual text of the link.

Simplified algorithm

In situation with four web pages, A,B,C and D. Each of these pages would begin with an estimated PageRank of ¼ (0.25).

If pages B,C and D link to A, then they would each confer 0.25 PageRank to A. All PageRank PR( ) in this simplistic system would thus gather to A because all links would be pointing to A.

PR(A) = PR(B) + PR(C) + PR(D).

This is 0.75.

If B also has a link to page C, and page D has links to A,B and C. The *value of the link-votes is divided among all the outbound links on a page*. Thus, page B gives a vote worth 0.125 to page A and a vote worth 0.125 to page C. Only one third of D's PageRank is counted for A's PageRank (approximately 0.083).

$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}.$$

PageRank can be calculated as document's own PageRank score divided by the normalized number of outbound links L( ).

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}.$$

In the general case, the PageRank value for any page u can be expressed as:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}.$$

An important aspect of PageRank calculation is the matrix inversion. Therefore, we'll discuss about few numerical inversion algorithms:

- Jacobi iteration

German mathematician Carl Gustav Jacob Jacobi has been developed simple method which is defined by

$x\{0\} = b$
$x\{i+1\} = (1 – M) * x\{i\} + b$

for $i \geq 0$, where {i} denotes the result after the i-th iteration.

Jacobi matrix is the Matrix (1 - M). In case of PageRank calculation the Jacobi matrix is given by d T (damping factor times transition matrix), a sparse matrix. The solution of the iteration is x, if the limit exists. The convergence is guaranteed, if the absolute value of the largest eigen value of (1 - M) is less than one. In case of PageRank calculation this is fulfilled for 0< d < 1.

There is a generalization of the algorithm. Using the decomposition (D - M), where D is an easy to invert matrix (e.g. the diagonal elements of M), leads to:

$x\{0\} = D^{-1} * b$
$x\{i+1\} = D^{-1} * ((D – M) * x\{i\} + b)$

Gauss-Seidel method

An improved algorithm is the Gauss-Seidel iteration. It based on the decomposition

$M = D + L + U$

where D, L and U are the diagonal, lower triangular and upper triangular parts of M. This yields

$x\{i+1\} = D^{-1} * (L * x\{i+1\} + U * x\{i\} + b)$

Introducing a relaxation parameter $\lambda \neq 0$ leads to a generalization of the Gauss-Seidel method:

$x\{i+1\} = (1 - \lambda) x\{i\} +$
$\lambda D^{-1} * (L * x\{i+1\} + U * x\{i\} + b)$

Minimal residue

Another iteration scheme is the minimal residue iteration. It is given by

$x\{i+1\} = x\{i\} + r\{i\} (r\{i\}_j M_{jk} r\{i\}_k) /$
$(M_{jk} r\{i\}_k M_{jl} r\{i\}_l)$

where the residue is defined by

$r\{i\} = b - M * x\{i\}$

The minimal residue iteration is never divergent.

In situation when some random surfer that starts from a random page, and at every time chooses the next page by clicking on one of the links in the current page. So, we could define the rank of a page as the fraction of time that the surfer spent on that page on the average. Clearly, important will be visited more often, which justifies the definition.

PageRank also can be define as the stationary distribution of a stochastic process whose states are the nodes of the web graph. The process itself is obtained by combining the normalised adjacency matrix of the web graph with a trivial uniform process that is needed to make the combination irreducible and aperiodic, so that the stationary distribution is well defined. The combination depends on a damping factor, which will play a major role in this paper. When damping factor is 0, the web-graph part of the process is annihilated, resulting in the trivial uniform process. As damping factor goes to 1, the web part becomes more and more important.
PageRank also can be defined as sum of the PageRanks of all incoming links, divided by the number of its outgoing links.

PageRank is Google's method of ranking individual web pages. Google looks at the pages which link to your page and how they rank in terms of importance. In a nutshell, pages that have links from important, high quality pages, receive a higher PageRank. Google combines PageRank with sophisticated text-matching techniques to find pages which are both important and relevant to your search

The PageRank system is a numerical grade from 0-10. There have been great articles written in the past on how to estimate and compute PageRank, and they can be confusing to some

V. FACTORS THAT IMPACT PAGERANK

- Valid Code – HTML/XHTML for the site should be valid
- Tags - using proper meta tags
- Sitemap – for sites with lots of folders and subfolders, the site should have sitemap with site tree structures linked
- CSS content
- Internal links and external links on the same page may not be splitting the Google PageRank vote equally

- Depending on the location of the link, Google PageRank may be weighted differently
- Multiple links to the same URL from the same page may not each get the same piece of the Page Rank vote
- "Run-of-site" external links, like Blogrolls, may now have diminished PageRank
- Links between domains that Google sees as "related" may have their PageRank significantly damped down. Possibly the same goes for sites that link to sub-domains
- Incoming Links from popular sites are important - If pages linking to you have a high PageRank then your page gains some part of their reputation
- Site can be banned if it links to banned sites – developer must be careful of any out-going links from their site because Google will penalize you for bad links
- Illegal activities like deceptive redirects, cloaking, automated link exchanges, or anything else against Google's quality guidelines will penalize your PageRank and possibly ban your site from Google
- Different pages from a site can have different Page Rank - Search engines crawl and index web pages not websites - therefore your page rank may vary from page to page within your website
- Content is not taken into account when PageRank is calculated - Content is taken into account just when you actually perform a search for specific search terms
- Bad incoming links don't have impact on Page Rank – from where the links come it doesn't matter. Sites are not penalized because of where the links come from

VI. CONCLUSION

As the web grows in size and value, search engines play an increasingly critical role, allowing users to find information of interest. Search engines used huge number of algorithm for ranking the sites. PageRank is a link analysis algorithm used by the Google Internet search engine.

VII. REFERENCES

[1] A. Arasu, J. Novak, A. Tomkins and J. Tomlin, "PageRank Computation and the Structure of the Web: Experiments and Algorithms", Technical Report, IBM Almaden Research Center, Nov. 2001

[2] Taher Haveliwala and Sepandar Kamvar. The condition number of the PageRank problem. Technical Report, Stanford University Technical Report, June 2003

[3] Taher Haveliwala. Efficient computation of PageRank. Technical report, Stanford University Technical Report, October 1999

[4] Chris Pan-Chi Lee, Gene H. Golub, and Stefanos A. Zenios - A fast two-stage algorithm for computing PageRank and its extensions. Technical report, Stanford University Technical Report, 2004

[5] Luca Pretto. A theoretical analysis of google's PageRank. In Proceedings of the Ninth Symposium on String Processing and Information Retrieval, 2002

[6] G. Jeh and J. Widom. Scaling personalized web search. In Proceedings of the Twelfth International World Wide Web Conference, 2003

[7] A. Arasu, J. Novak, A. Tomkins, and J. Tomlin. PageRank computation and the structure of the Web: experiments and algorithms. In Proceedings of the Eleventh International World Wide Web Conference, Poster Track, 2002.

Impact of the Number of Chromosomes on the Fitness Value Improvement in Standard GA Applications

Ivana Stojanovska[1], Agni Dika[2] and Blerta Prevalla[3]

***Abstract* – In this paper we make visual representation of the data we obtain from the applications that implement the GA, which clearly shows the fitness value minimization over the generations and the impact of the number of chromosomes on the fitness value and the time required to find the optimal solution.**

***Keywords* – Optimization, Chromosome, Fitness value, Generation, Visualization.**

I. INTRODUCTION

Genetic algorithms (GA) are search and optimization algorithms that use the theory of evolution as a tool to solve a problem in science and engineering. They incorporate the idea of survival of the strongest in a search algorithm that provides a searching method which does not necessarily needs to examine every possible solution in a given practice area to achieve good result. [18]

The main effect of GA is in the parallel nature of its search. They implement powerful form of "hill climbing" which supports multiple solutions, eliminate those who do not promise, and improve the best solution. [1] Fig. 1 shows several solutions that converge to the optimal points in the search space. Initially, the solutions are spread through the space of possible solutions. After several generations, they tend to group (cluster) around the areas with a higher quality solution.

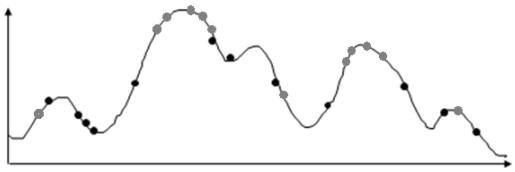


Fig. 1. Distribution of candidate solutions in Generation 1 (the black points on the curve) and Generation N (the gray points on the curve)

The process of GA generally consists of the following steps: Encoding, Evaluation, Crossover, Mutation and Decoding [18]. Once all of this is done, a new generation is evolved and the process repeats until they meet some stopping criterion. At this level the individual who is closest to the optimal solution is decoded and the process is completed.

[1]Ivana Stojanovska is with the Faculty of Information and Communication Technologies, Vojvodina bb, 1000 Skopje, Macedonia, E-mail: ivana.stojanovska@fon.edu.mk

[2]Agni Dika is with the Faculty of Contemporary Sciences and Technologies, Ilindenska bb, 1200 Tetovo, Macedonia, E-mail: a.dika@seeu.edu.mk

[3]Blerta Prevalla is with the Faculty of Information and Communication Technologies, Vojvodina bb, 1000 Skopje, Macedonia, E-mail: blerta.prevalla@fon.edu.mk

II. SOLVING THE TRAVELING SALESMAN PROBLEM

This problem is well known and is a standard problem for testing search algorithms of this type. The basic problem consists of the following: the traveling salesman is required to pass n given cities, but each city to visit only once. The applied algorithm has to find the minimum time for performing such a journey through the cities.

A. Problem Complexity

The traveling salesman problem is of particular importance because it is a classic example of NP-hard (non-deterministic polynomial time hard) problem, that so far can be solved only in exponential time. It is a classic problem with great computational complexity. If there are n cities, then the maximum number of possible plans to travel between towns is (n-1)!. You can create a simple algorithm which examines all possible paths and comes up with the shortest one. But the main problem is that the time required for algorithm execution grows with tremendous speed as the number of cities increases. If there are 25 cities, then the algorithm must look 24! possible routes. 24! is approximately 6.2• . Even if you use a computer that can investigate one million routes per second, it would take about 6.2• / = 6.2• seconds to solve the problem. This is approximately 1.96• years.

By using dynamic programming techniques the problem can be solved in $O(2^n)$ time. Although this solution increases exponentially, it is much better than O(n!).

B. Implementation

In our implementation we chose the encoding where each gene in the chromosome represents a city, and the chromosome represents the order in which the traveling salesman would move. We use an implementation with 25 cites and we must not forget that the salesman should visit each city only once.

8	2	3	4	7	6	1	5

Fig. 2. A chromosome which represents one of the possible solutions

The fitness function that characterizes each chromosome, represents the total length of the route from the first to the last

gene (city) moving according to the order of the genes in the chromosome. If the cities are represented with x and y coordinates in 2D coordinate system, then we calculate the distance between them according the Eq.1:

$$r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

The fitness value of each chromosome is the sum of all distances between the genes and the goal of this GA is to minimize this function.

When starting the application it is necessary to initialize a starting population with a given number of chromosomes (in our application this number is 150). Once created, this population should provide a method for crossing into the next generation where these chromosomes are replaced with new chromosomes by applying the GA operators. The best solution from the current generation is saved and added to the new generation, if it does not already exist there. All this is repeated a number of times, which equals the number of generations in our application. The latest generation of chromosomes should provide the best solutions.

C. Graphical Representation

Data obtained from the program that implements the GA are visually presented in the form of a graph from which we can see the progress of the GA's fitness value minimization over the generations. Also shown is the data for the smallest fitness value, generation of its occurrence and the solution itself (the chromosome with the smallest fitness value). This application is developed in Gambas under Linux Ubuntu 7.04 and its appearance is shown in Fig.3.

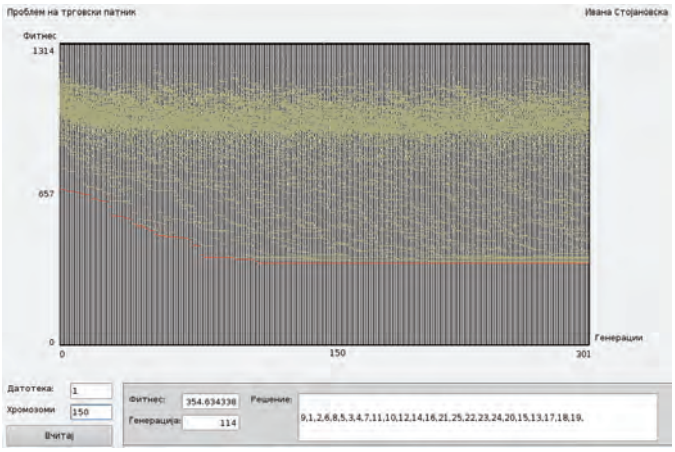


Fig. 3. Application for fitness value visualization over the generations

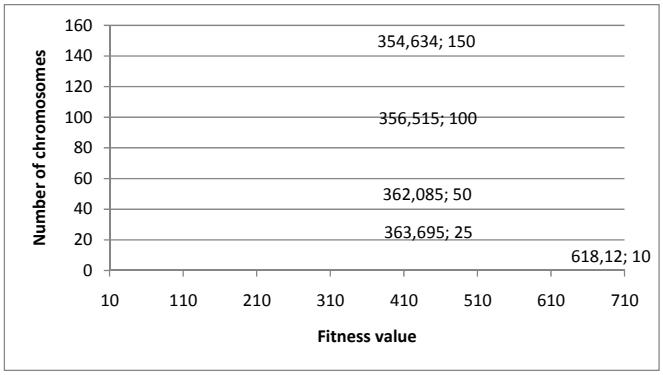
Initially, for the program execution we take population of 150 chromosomes, each composed of 25 genes (as the number of cities). For algorithm execution we use 300 generations to obtain the results shown in Fig. 3. From the figure we can see that the population is advancing toward a population with slightly better features. The gray cloud represents all the solutions in a generation, and in each of these generations the best solution is shown with a red line. As it can be seen in the figure, the best fitness value occurs in generation 114, with the best fitness value of 354.63 and the exact solution is: 9, 1, 2, 6, 8, 5, 3, 4, 7, 11, 10, 12, 14, 16, 21, 25, 22, 23, 24, 20, 15, 13, 17, 18, 19.

Within this visualization we made an analysis on the impact of the number of chromosomes on the fitness value and the time required to find the optimal solution. Moreover, we got the results shown in Table I.

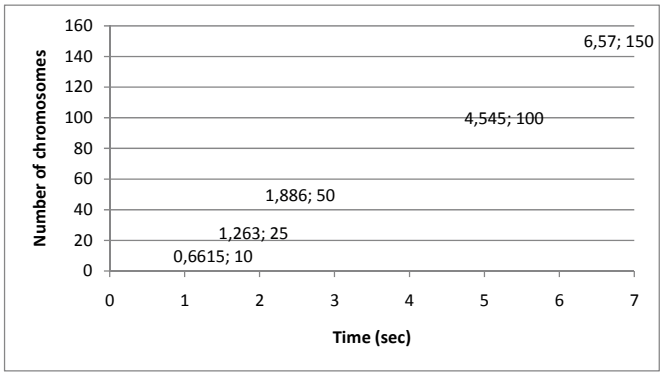
TABLE I
VARIABLE NUMBER OF CHROMOSOMES

No.of chromosomes	No.of generations	Fitness value	Time (sec)
10	300	618,120	0,661
25	300	363,695	1,263
50	300	362,085	1,886
100	300	356,515	4,545
150	300	354,634	6,57

Fig.4 (a), which illustrates the behavior of the fitness value when the number of chromosomes changes, shows that when the number of chromosomes is very small (10) the fitness value is significantly higher (618.120) according to the rest 4 examined cases (number of chromosomes = 25, 50, 100, 150) where it slightly improves. The graph in (b) illustrates that by increasing the number of chromosomes we get almost linear increase in time required for finding the optimal solution (the smallest distance between cities). The time increases slightly from 0.661, 1.263, 1.886, 4.545 to 6.57 sec, as the number of chromosomes gets bigger and bigger.



(a)



(b)

Fig. 4. Diagram of the influence of the number of chromosomes on the fitness value and the time

From the above analysis we can establish that when the number of chromosomes per generation is very small, it significantly affects the fitness value of the best solution found, but in further cases, there is almost no change in the fitness value of the best solution.

III. SOLVING THE 2D PACKING PROBLEM

The packing problem is actually a way of finding the optimal solution how to collect a given number of boxes (packages) in a large enclosed space. The problem is very easy to set and define, but it is quite difficult to solve. In mathematical terms speaking, as the traveling salesman problem, this problem is NP-hard, which means that when given a set of boxes and a space, it is very difficult to answer questions such as: Which is the best way to pack up the boxes, which would be the optimal solution and how good it would be. [19]

A. Problem Complexity

The packing problem is quite complex. When the number of packages is small the problem is relatively simple, but the complexity of the problem grows exponentially with the increase in the number of packages. For example, the number of possible combinations of 20 packages that can be oriented in 2 directions is: $(20 \cdot 2) \cdot (19 \cdot 2)...(1 \cdot 2) = 20! 2^{20} = 2{,}55108 \times 10^{24}$

First we have a choice of 20 packages that can be placed in 2 directions in the space, then we have 19 other packages to choose from, which can also be placed in 2 directions, etc.

If you perform a "brute force" test on this small problem of packing with 20 packages in 2D space, it will require a computer capable of checking millions of combinations of packings in a second, and still will need more than 10 years to complete the process. However, because GA use evolution to improve the solution, it relatively quickly starts to improve and becomes significantly better for a short time (several minutes), even when performing on relatively weak computer.

B. Implementation

In our implementation we selected the encoding where each gene in the chromosome is a package (number of packages equals 20), together with its orientation, and the chromosome represents the order in which they should be packed in the enclosed space (as shown in Fig. 5). The second field of each gene in the chromosome indicates whether the corresponding package should be rotated or not. If this field is set to true then the corresponding package is rotated from its initial position, and otherwise (false) it remains in the same orientation.

0	true	9	false	8	true	1	false	3	false	2	true	4	true	7	false	6	false	5	true

Fig.5. A chromosome which represents one of the possible solutions

The cost of each chromosome is calculated as the space wasted in the enclosed space, after all the packages are packed according to Eq.2:

$$\cos t = 1 - \frac{a}{b \cdot c} \tag{2}$$

where b is the highest point on the top package in the enclosed space, c is the width of the space, and a is the total area of the packages. So the optimal solution would have a cost of 0, if the area of the space is equal to the area of the boxes (packages).

At the start of the application we initialize a starting population with randomly generated chromosomes. As input, the application receives a .txt file that contains information about the height and the width of the packages. [19] Once created, this population provides a method for crossing into the next generation where these chromosomes are replaced with new chromosomes by applying the GA operators. All this is repeated a number of times and the latest generation of chromosomes should provide the best solutions.

C. Graphical Representation

As with the program for travelling salesman, data obtained from the program is visually represented in graph from which you can see the progress of the GA. Initially, for the execution of the program we take a population of 100 chromosomes, each composed of 20 genes (as the number of packages). During the execution it can be seen that the population gradually advances towards a population with better features. For algorithm execution we use 50000 generations to obtain the results shown in Fig.6. During the execution it can be seen that the population gradually advances towards a population with better features. It may be noted that after 50000 generations we obtain a fitness value of $\approx$8.9%, which is found in the 9069th generation and the exact solution is: 16, 0, 4, 12, 14, 15, 9, 1, 8, 17, 3, 13, 6, 7, 11, 5, 2, 10.

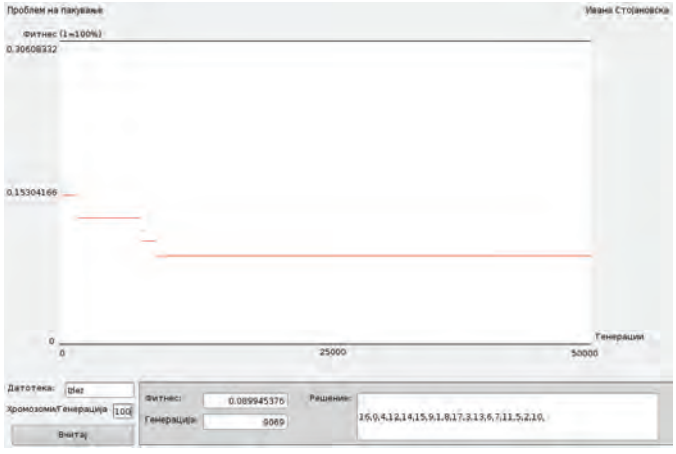


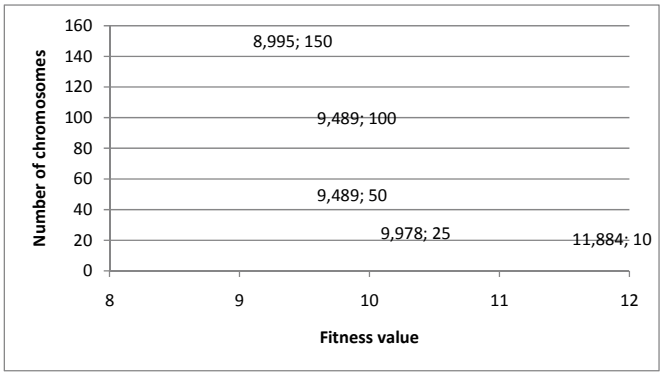
Fig. 6. Application for fitness value visualization over the generations

Fig.7 is made based on analysis conducted on the impact of the number of chromosomes to the fitness value and the time required to find the optimal solution. Within this analysis were obtained the results shown in Table II.

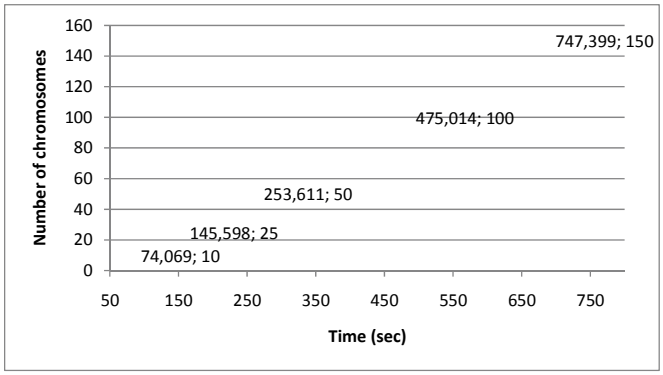
TABLE II
VARIABLE NUMBER OF CHROMOSOMES

No. of chromosomes	No. of generations	Fitness value	Time (sec)
10	50000	0,11884	74.069
25	50000	0,09978	145.598
50	50000	0,09489	253.611
100	50000	0,09489	475.014
150	50000	0,08995	747.399

Fig.7 illustrates the dependence of the fitness value and the time of the change in the number of chromosomes. At the graph in (a) it can be seen that when the number of chromosomes is very small (10) the fitness value is significantly higher (0.11884) in comparison to the rest 4 cases examined (25, 50, 100, 150) where it slightly improves as the number of chromosomes increases. The graphs in (b) shows that we have almost linear increase in time required for finding the optimal solution (packing with the smallest waist space). The time gradually increases from 74.069, 145.598, 253.611, 475.014,to 747.399 sec.



(a)



(b)

Fig. 7. Diagram of the influence of the number of chromosomes on the fitness value and the time

It may be concluded that increasing the number of chromosomes significantly affects the fitness value of the best solution found, meaning that as many generations we have as more likely that we will find a solution with minimal waist space on packing.

IV. CONCLUSION

With regard to what GA do best, which is an iterative improvement of the result, the conclusion has to be that the problems based on search, as the problem of packing and the traveling salesman problem, there are few other tools that are able to compete with the GA. One very interesting question in the analysis of this class of problems is whether it is worth spending many hours on expensive workstation to obtain a solution close to the optimum, or to work a few minutes on cheap personal computer (PC) to get "good enough" results for these applications.

REFERENCES

[1] George, F.L. (2005). Artificial Intelligence: Structures and Strategies for ComplexProblem Solving. Harlow, England: Addison Wesley.

[2] Mitchell, M. (1996). An introduction to Genetic Algorithms. London, England: MIT Press.

[3] Leardi, R. (2003). Nature-inspired methods in chemometrics: genetic algorithmsand artificial neural networks, Volume 23 (Data Handling in Science and Technology). Elsevier Science

[4] Mennon, A. (2004). Frontiers of Evolutionary Computation (Genetic Algorithms and Evolutionary Computation). Springer-Verlag

[5] Coppin. B. (2004). Artificial Intelligence Illuminated. Jones & Bartlett Publishers

[6] Steele, N.C. (2005). Adaptive and Natural Computing Algorithms: Proceedings of the International Conference in Coimbra, Portugal, 2005. Springer-Verlag

[7] Poli, Riccardo and Langdon, William B. and McPhee, Freitag N. (2006). A Field Guide to Genetic Programming. Lulu.com - Page 141

[8] Hauot, R.L., and Haupt, S.E. (2004). Practical Genetic Algorithms. Wiley-Interscience

[9] Zheng, Y., and Kiyooka, S. (1999). Genetic Algorithm Applications.

[10] Whitley, D. A genetic Algorithm Tutorial. Computer Science Department, Colorado State University.

[11] Holland, J.H. (1975). Adaptation in Natural and Artificial Systems. Press, University of Michigan.

[12] Goldberg, D.E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley.

[13] Pelikan, M., and Lobo, G.F. Parameter-less Genetic Algorithm: A worst-case Time and Space Complexity Analysis.

[14] Grefenstette, J.J. (1989). Genetic algorithms for changing environments.

[15] Grefenstette, J.J. (1989). How Genetic Algorithms work: A critical look at implicit parallelism.

[16] Herrmann, W.J. A Genetic Algorithm for Minimax Optimization Problems. Department of Mechanical Engineering and Institute for Systems Research, University of Maryland

[17] Sengoku, H., and Yoshihara, I. (1993). A Fast TSP Solution using Genetic Algorithm. Information Processing Society of Japan 46th Nat'l Conv.

[18] Bryant, K. (2000). Genetic Algorithm and the Traveling Salesman Problem.

[19] Sorensen, J.J., and Mortensen, H. The use of Genetic Algorithms on The Bin PackingProblem.

[20] Coley, D.A. (1999). An Introduction to Genetic Algorithms for Scientists and Engineers. World Scientific Publishing.

Performance Analysis of Sorting Algorithms Through Time Complexity

Blerta Prevalla[1], Ivana Stojanovska[2] and Agni Dika[3]

***Abstract* – Performance analysis through time complexity means analysing the time needed for execution of a program. It is very useful because it provides information of how to distribute efforts and resources in order to ensure greater efficiency and to keep developers focused on the essential goals of the program.**

In this paper will be analysed the performance of five sorting algorithms via analytical methods and experimental ones. Sorting algorithms are chosen to analyse because many scientist consider sorting as one of the most crucial problems in the study of algorithms and programs.

Sorting algorithms will be implemented in C++ and tested in 3 machines with different configurations and with different input values to see the changes of running time depending on many circumstances.

***Keywords* – Time Complexity, Sorting Algorithms, Performance Analysis.**

I. INTRODUCTION

By analysing the performance of a program we mean analysing the amount of computer time needed to run a program. To determine the performance of a program we use two approaches. One is analytical and the other experimental. During the performance analysis we use analytical methods and for performance measurement we conduct experiments.

Performance measurement indicates what a program is accomplishing and whether results are being achieved. It helps programers by providing them information on how resources and efforts should be allocated to ensure effectiveness. Performance measurement must often be coupled with evaluation data to increase our understanding of why results occur and what value a program adds.

The problem of sorting is one of the most widely studied practical problems in computer science, meaning using computer to put files in certain order. Many computer programs use sorting as an mediator step and that's why there are many sorting programs. is faced with the problem of determining which of the many available algorithms is best suited for his purpose.

This task is becoming less difficult than it once was for two reasons. First, sorting is an area in which the mathematical analysis of algorithms has been particularly successful: we can predict the performance of many sorting methods and compare them intelligently. Second, we have a great deal of experience using sorting algo- rithms, and we can learn from that experience to separate good algorithms from bad ones.

[1]Blerta Prevalla is with the Faculty of Information and Communication Technologies, Vojvodina bb, 1000 Skopje, Macedonia, E-mail: blerta.prevalla@fon.edu.mk

[2] Ivana Stojanovska is with the Faculty of Information and Communication Technologies, Vojvodina bb, 1000 Skopje, Macedonia, E-mail: ivana.stojanovska@fon.edu.mk

[2]Agni Dika is with the Faculty of Contemporary Sciences and Technologies, Ilindenska bb, 1200 Tetovo, Macedonia, E-mail: a.dika@seeu.edu.mk

II. PERFORMANCE ANALYSIS

To analyze the performance of an algorithm we must first identify the resourses of primary interest so that the detailed analysis may be properly focused. We describe the process in terms of studying the runing time since it is the resourse most relevant here. A complete analysis of the running time of an algorithm invloves the following steps:

- Implement the algorithm completely.
- Determine the time required for each basic operation.
- Identify unknown quantities that can be used to describe the frequency of execution of the basic operations.
- Develop a realistic model for the input to the program.
- Analyze the unknown quantities, assuming the modelled input.

A. Steps in analysing an algorithm

The first step in analysis is to carefully implement the algorithm on a particular computer. This implementation not only provides a concrete object to study, but also can give useful empirical data to aid in or to check the analysis. Presumably the implementation is designed to make efficient use of resources, but it is a mistake to overemphasize efficiency too early in the process.

The next step is to model the input to the program, to form a basis for the mathematical analysis of the instruction frequencies. The values fo the unknown frequencies are dependent on the input to the algorithm: the input size is normally the primary parameter used to express our results, but the order or value of input data items also ordinarly affect the running time, as well. For these sorting algorithms it is normally convenient to assume that the inputs are randomly ordered and distinct. Another possibility for sorting algorithm is to assume that the inputs are random numbers taken from a relatively large range.

Several different models can be used for the same algorithm: one model might be chosen to make the analysis as simple as possible; another model might better reflect the actual situation in which the program is to be used.

The average case results can be compared with empirical data to verify the implementation, the model, and the analysis. The end goal is to gain enough confidence in these that they can be used to predict how the algorithm will perform under whatever circumstances present themselves in particular applications. For example, we may wish to evaluate the possible impact of a new machine architecture on the performance of an important algorithm. [1]

Often it is possible to do so through analysis, perhaps before the new architecture comes into existence. Another important example is when an algorithm itself has a paramenter that can be adjusted: analysis can show what value is best.

III. TIME COMPLEXITY OF SORTING ALGORITHMS

A. Time Complexity

The time complexity of a program is the amount of computer time it needs to run to completion. [2]

We are mainly interested in that how long does the sorting programs run. It possibly takes a very long time on large inputs until the program has completed its work and gives a sign of life again. Sometimes it makes sense to be able to estimate the running time *before* starting a program. Nobody wants to wait for a sorted phone book for years! Obviously, the running time depends on the number n of the strings to be sorted.

We are interested in the time complexity of a program because some computer systems require the user to provide an upper limit on the amount of time the program will run. Once this upper limit is reached the program is aborted. An easy way out is to simply specify a time limit of a few thousand years. However, this solution could result in serious fiscal problems if the program runs into an infinite loop caused by some discrepancy in the data and you acgtually get billed for the computer time used. We would like to provide a time limit that is just slightly above the expected run time. Also, the program we are developing might need to provide a satisfactory real-time response. For example, all interactive programs must provide such a response.

B. Conducted experiments for sorting algorithms

During the experimental study we:

- Write a program to implement the current algorithm.
- Run the program for different input values.
- Get exact measurements from the actual execution time.
- Compare results.

The analysis of the average-case performance depends on the input being randomly ordered. This assumption is not likely to be strictly valid in many practical situations. In general, this reflects one of the mos serious challenges in the analysis of algorithms: the need to properly formulate models of inputs that might appear in practice. Fortunately there is often a way to circumvent this difficulty: “randomize” the inputs before using the algorithm. This simply amounts to randomly permuting the input file before sort. If this is done, then probabilistic statements about performance such as those made above are completely valid and will accurately predict performance in practise, no matter what the input.

Limitations of experiments:

- It is necessary to implement the algorithm, which may be difficult
- Results may not be indicative of the running time on other inputs not included in the experiment.
- In order to compare two algorithms, the same hardware and software environments must be used

In this paper will be shown the experiments done with most famous programs for sorting: Merge Sort, Insertion Sort, Selection Sort, Quick Sort and Bubble Sort.

We will see the following case:

- When we have as input random 10000, 15000, 25000, 30000, 45000, 50000, 65000, 75000, 90000, 100000 numbers. With these entries we will see the running time of algorithms when we execute them in machines with different performances like:

  - Computers with normal performance (Toshiba Satellite with processor: Intel Core 2 Duo, 1:50 GHz, 500 MHz and 1GB of RAM memory)
  - Faster Computer (T555 Dell Studio, the processor: Intel Core 2 Duo P7450 1033 MHz, 2.13 GHz, 1033 MHz and 4GB RAM memory)
  - Slower one (Dell Latitude D800, Intel (R) Pentium (R) M processor 1.70 GHz, 209 MHz and 512 RAM memory).

TABLE I
RUNNING TIME OF MERGE SORT ON THREE DIFFERENT MACHINES

N	Dell studio 1555	Toshiba Satellite A200	Dell Latitude D800
10 000	0.3750	0.6050	1.7520
15 000	0.5620	0.7810	2.6340
25 000	0.9220	1.2930	4.3860
30 000	1.0940	1.6250	5.2480
45 000	1.6560	2.6310	7.9210
50 000	1.8590	2.7630	8.7430
65 000	2.3600	3.6440	11.3460
75 000	2.7820	3.8930	13.1190
90 000	3.2810	4.6780	15.7330
100 000	3.6720	5.1820	17.5350

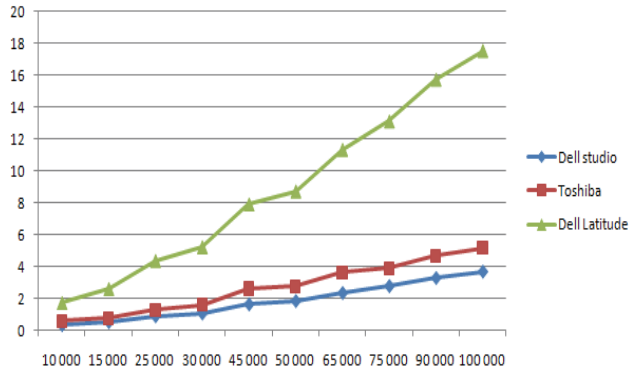


Fig. 1. The comparison of running time for Merge Sort

TABLE II
RUNNING TIME OF INSERTION SORT ON THREE DIFFERENT MACHINES

N	Dell studio 1555	Toshiba Satellite A 200	Dell Latitude D800
10 000	0.1410	0.2440	0.5110
15 000	0.3280	0.4610	1.1810
25 000	0.9060	1.2810	3.8560
30 000	1.3130	1.8420	5.0770
45 000	2.9370	4.6250	10.8760
50 000	3.6560	5.1280	13.4690
65 000	6.0940	8.6220	24.1850
75 000	8.2810	12.5440	30.7340
90 000	11.7500	17.5940	43.8930
100 000	14.6250	20.4590	55.1300

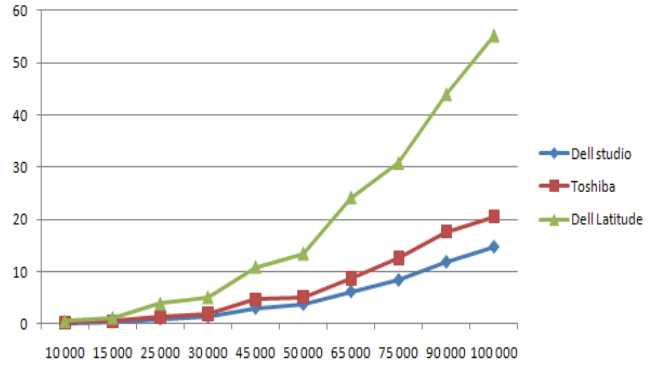


Fig. 2. The comparison of running time for Insertion Sort

TABLE III
RUNNING TIME OF QUICK SORT ON THREE DIFFERENT MACHINES

N	Dell studio 1555	Toshiba Satellite A 200	Dell Latitude D800
10 000	0	0.014	0.030
15 000	0	0.019	0.120
25 000	0.016	0.029	0.120
30 000	0.015	0.039	0.160
45 000	0.032	0.054	0.251
50 000	0.031	0.062	0.320
65 000	0.047	0.080	0.381
75 000	0.047	0.091	0.421
90 000	0.046	0.113	0.541
100 000	0.047	0.128	0.671

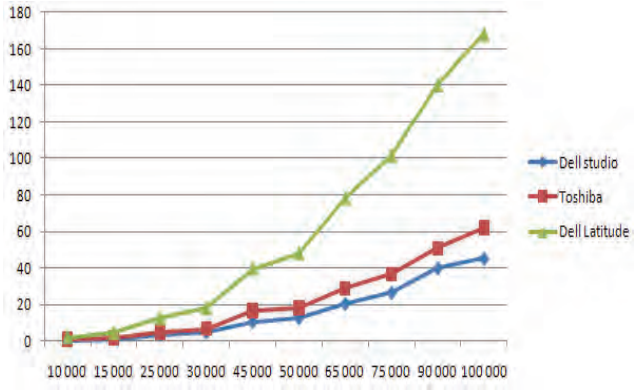


Fig. 3. The comparison of running time for Quick Sort

TABLE IV
RUNNING TIME OF SELECTION SORT ON THREE DIFFERENT MACHINES

N	Dell studio 1555	Toshiba Satellite A 200	Dell Latitude D800
10 000	0.5320	0.8470	2.0720
15 000	1.1720	1.6680	4.6160
25 000	3.2190	4.7890	12.6180
30 000	4.6410	6.6740	18.0960
45 000	10.5150	16.6840	39.5670
50 000	12.5940	18.2250	48.1900
65 000	20.3590	28.8850	78.3920
75 000	26.4380	36.6340	101.5860
90 000	39.8900	50.7140	140.2720
100 000	45.2340	62.0270	168.1320

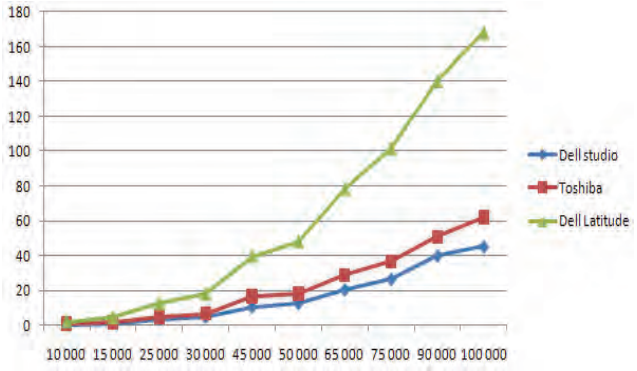


Fig. 4. The comparison of running time for Selection Sort

TABLE V
RUNNING TIME OF BUBBLE SORT ON THREE DIFFERENT MACHINES

N	Dell studio 1555	Toshiba Satellite A 200	Dell Latitude D800
10 000	0.5310	0.8410	2.1630
15 000	1.2040	1.6780	4.8170
25 000	3.3430	4.8870	13.4490
30 000	4.8280	7.7300	19.2380
45 000	10.9850	16.0330	43.3820
50 000	13.3440	19.8430	53.5470
65 000	22.7030	32.6130	90.2900
75 000	30.1560	50.1970	126.4020
90 000	43.4690	67.4900	173.6290
100 000	53.8750	85.1890	216.6920

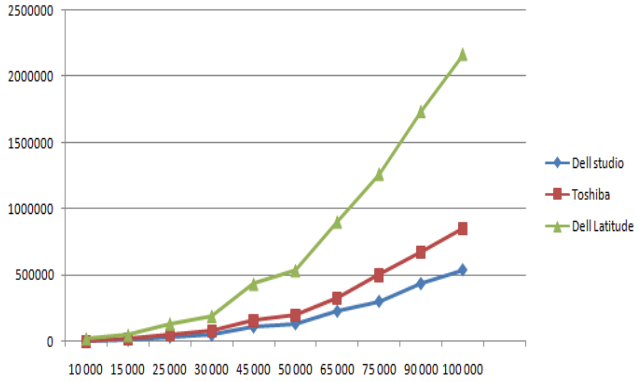


Fig. 5. The comparison of running time for Bubble Sort

From the results presented in the tables and the graphs above, we can see that the performance and the running time of a program depends directly from the machine we are running it.

If we see the Merge Sort program for sorting 100000 random numbers we see that on the faster computer it took 3.672 sekonds to sort them, and in the slower computer it took 17.535 almost 14 seconds more. The faster computer for sorting 100000 elements with Insertion sort program needs 14.625 seconds and the slower computer needs 55.13 seconds to sort those numbers.

We can see that the difference here is bigger than in the Merge Sort, its almost 41 seconds. In contrary, the difference in running time of sorting 100000 numbers with Quick Sort is much smaller, it takes only 0.047 seconds for sorting the elements with the computer with better performances and 0.671 seconds for the slower computer, 0.624 seconds more.

Bubble Sort and Selection Sort in principal are programs that need more time to do sorting, and they are not so propriate to use because they need much more time. Selection Sort running on the faster computer needs 45.234 seconds to sort 100000 numbers and almost 123 seconds more for sorting them with the slower computer.

Even though Bubble Sort its famous for sorting because it is easier to program it, still it takes more time to sort than any other sorting algorithms. It takes 53.875 to sort 100000 elements on the computer with high performances and 216.692 seconds for sorting with the slower computer.

We can conclude that the slower a program is, the bigger is the difference when we execute it on different machines.

IV. CONCLUSION

A full performance analysis like that above requires a fair amount of effort that should be reserved only for our most important algorithms. Fortunately, there are many fundametal methods that do share the basic ingredients that make analysis worthwhile:

- Realistic input models can be specified.
- Mathematical descriptions of performance can be derived.
- Concise, accurate solutions can be developed.
- Results can be used to compare variants and compare with other algorithms, and help adjust values of algorithms parameters.

We conclude performance analysis is very important because we can predict the time needed by a program to solve a problem and also we'll know how to distribute efforts and resources in order to ensure greater efficiency.

These are the areas involving the most significant intellectual challenge, and deserve the attention that they get.

REFERENCES

[1] Sedgewick, R., Flajolet, P. *An Introduction to the Analysis of Algorithms* English Reprint Edition (2006) Pearson Education Asia Limited and China Machine Press

[2] Sahni, S. (2002)*Structures, Algorithms, and Applications in C++*, University of FloridaData McGraw-Hill

[3] Cormen, T.H., & Leiserson, C.E., & Rivest, R., & Stein, C.,(2001) *Introduction to Algorithms*, Second Edition. MIT Press and McGraw-Hill

[4] Flamig, B. *Practical algorithms in C++*, John Ëiley & Sons Inc., Canada, 1995

[5] Knuth,D. *The art of Computer Programming, Volume 3: Searching and Sorting* (2nd ed.), Addison-Wesley, 1998. Longman Publishing Co., Inc., RedWood City, CA, 1998

[6] Mehta, D., Sahni, S. *Handbook of Data Structures and Applications*, (2005) Chapman & Hall / CRC Press Company

[7] Preiss, R.B, Data Structures and Algorithms With Object-Oriented Design Patterns in C++, (1997) Waterloo, Canada

DBpedia as Entry Point in the Web of Data (Semantic Web)

Mile Gjorgjioski[1], Dijana Capeska-Bogatinoska[2] and Mirjana Trompeska[3]

***Abstract* – Linked Data lies at the heart of what Semantic Web is all about: large scale integration of data on the Web. The importance of DBPedia is that it extracts structured information from Wikipedia and incorporates links to other datasets on the Web, e.g., to Geonames. By providing those extra links, in term of RDF triples, applications exploit the extra knowledge from other datasets when developing an application.**

***Keywords* – semantic web, web of data, linked data, DBPedia.**

I. INTRODUCTION

Essentiality of the Semantic Web isn't just about putting data on the web, but in making links among them, so that a person or machine can explore the Web of data. With Linked Data, when you ask for some data, you can also find other related data, besides the one you were searching for. The goal of Linked Data is to enable people to share structured data on the Web, as easily as they can share documents today.

The term Linked Data, coined by Tim Berners-Lee, refers to a style of publishing interlinking structured data on the web. The value and usefulness of data increases the more if it is interlinked with other data. In summary, Linked Data is about using the web to create typed links between data from different sources.

So, to get Linked Data, we have to:

- Use the RDF data model to publish data on the web
- Use RDF links to interlink data from different data sources

Applying those principles, we can create a data community on the web, a space where people and organizations can post and consume data about anything. This community of data is called Web of data or Semantic Web.

The Web of data can be accessed using Linked Data browsers, such as Tabulator, Marbles, OpenLink RDF Browser or Disco, just as the traditional Web of documents is accessed using HTML browsers. Linked Data browsers enable users to navigate between different data sources by following RDF links, instead of following links between HTML pages.

The user can start with one data source, and then move through a potentially endless Web of data sources connected by RDF links. For example, a user can start with data about one person in one source, and it might be interested in information about person's home town. By following an RDF link, the user can navigate to that information contained in another dataset.

Just as the traditional document Web can be crawled by following hypertext links, the Web of data can be crawled by following RDF links. Search engines can provide sophisticated query capabilities, similar to those provided by conventional relational databases. Because the query results themselves are structured data, not just links to HTML pages, they can be immediately processed, thus enabling a new class of applications based on the Web of data.

A typical case of a large Linked Data is DBPedia, which makes the content of Wikipedia available in RDF. The importance of DBPedia is not only that it extract structured information from Wikipedia, but also that it incorporates links to other datasets on the Web, e.g., to Geonames. By providing those extra links, in term of RDF triples, applications may exploit the extra knowledge from other datasets when developing an application. DBPedia allows us to ask sophisticated queries against Wikipedia, and to link other datasets on the Web to Wikipedia data. This makes it easier for the huge amount of information in Wikipedia to be used in new and interesting ways, and also, this might inspire new mechanisms for navigating, linking and improving the encyclopedia itself.

II. RESOURCES AND RESOURCE IDENTIFIERS

If we want to publish data on the web, we first have to identify the item of interest in our domain. Resources can be anything: books, people, organizations, etc. They are things whose properties and relationships we want to describe. In web terminology, all items of interest are called *resources* [2].

Resources are named by Uniform Resource Identifiers (URIs), and this provides globally unique, distributed naming system that we need for distributed knowledge [3]. For anything that has URI, we say that is on the web. For example: you, the book which you have bought last week etc. URI have fundamental role in the Semantic Web; they keep the Web of data together. URI's work is not just as a name, but also as a mean of accessing information about a resource over the web.

III. RDF DATA MODEL AND RDF LINKS

RDF is an acronym for Resource Description Framework. It is a standard data and modeling specification used to encode metadata and digital information. The Semantic Web vision is

[1]Mile Gjorgjioski is with Omnia Computers, Pitu Guli No10, 7500 Prilep, Macedonia, E-mail: Mile.Gjorgjioski@omniacomputers.com

[2]Dijana Capeska-Bogatinosta is with Eurokompozit 11 Oktomvri AD, A.Makedonski 2/42, 7500 Prilep, Macedonia, E-mail: finance@eurokompozit.com.mk

[3]Mirjana Trompeska is with OU Kiril i Metodij, s.Kanatlarci, 7500 Prilep, Macedonia, E-mail: mirjanatrompeska@yahoo.com

predominantly based on the fundamental power of the RDF language. In RDF, a description of a resource is represented as a number of triples. The three parts of each triple are called: *subject*, *predicate* and *object*. For example:

Dijana	has the **email address**	**dijanac@excite.com**
(subject)	(predicate)	(object)

The subject of a triple is the URI identifying the described resource. The predicate indicates what kind of relation exists between subject and object; its URI too. The object can either be a simple literal value, or the URI of another resource that is somehow related to the subject.

Collection of RDF triples is called **RDF model**: *RDF triple is a labeled connection between two resources* [1]. **RDF links** represent links between two resources. An RDF link simply states that one piece of data has some kind of relationship to another piece of data. These relationships can have different types. For example, an RDF link that connects data about people can state that two people know each other. RDF links are the foundation for the Web of data.

IV. WHAT IS DBPEDIA

DBpedia is a community effort to extract structured information from Wikipedia and make them available on the Web. DBpedia allows us to ask sophisticated queries against Wikipedia and to link other data sets on the Web to Wikipedia data. DBpedia makes it easier for the huge amount of information in Wikipedia to be used in new and interesting ways [4].

Knowledge bases are playing an increasingly important role in enhancing the intelligence of Web and enterprise search and in supporting information integration. Today, most knowledge bases cover only specific domains, and are created by relatively small groups of knowledge engineers, also, they are very cost intensive to keep up-to-date as domains change. At the same time, Wikipedia has grown into one of the central knowledge sources of mankind, maintained by thousands of contributors. The DBpedia project leverages this gigantic source of knowledge by extracting structured information from Wikipedia and by making this information accessible on the Web.

The DBpedia knowledge base currently describes more than 3.4 million things, including 312,000 persons, 413,000 places, 94,000 music albums, 49,000 films, 15,000 video games, 140,000 organizations, 146,000 species and 4,600 diseases. The DBpedia data set features labels and abstracts for these 3.2 million things in up to 92 different languages. The DBpedia knowledge base altogether consists of over 1 billion pieces of information (RDF triples) out of which 257 million were extracted from the English edition of Wikipedia and 766 million were extracted from other language editions.
The DBpedia knowledge base has several advantages over existing knowledge bases: it covers many domains; it represents real community agreement; it automatically evolves as Wikipedia changes and it is truly multilingual [4].

The DBpedia knowledge base allows you to ask quite surprising queries against Wikipedia, for instance "Give me all famous persons which were born in Skopje". The SPARQL query language is used to query this data.

The DBpedia knowledge base is served as Linked Data on the Web. DBpedia defines Linked Data URI for millions of concepts, and various data providers have started to set RDF links from their data sets to DBpedia, making DBpedia one of the central interlinking-hubs of the emerging Web of data. DBpedia uses RDF as a flexible data model for representing extracted information and for publishing it on the Web.

DBpedia uses RDF as flexible data model, for presenting the extracted data and their announcement on the web. Each "thing" in DBpedia set of data, is identified through URI reference in the format: http://dbpedia.org/resource/Name where "Name" is taken from the source article from Wikipedia, which has the following format:
http://en.wikipedia.org/wiki/Name**.** Each DBpedia resource is described with different characteristics. Resources are described with their label, short and long abstract on English language, link to the appropriate Wikipedia article and link to the picture that describes the "thing" (if it exists at all). If the "thing" has multilanguage version in Wikipedia, then a short and long abstracts on the appropriate language are added to the description, and links to the Wikipedia pages in that language.

DBpedia contains HTML links to outside web pages and RDF links to external sources of data. There are two kinds of HTML links: *dbpedia:reference* - which lead to couple of web pages for the "thing"; and *foaf:homepage* - links which lead to web pages which are considered to be official web pages (homepages) for the "thing".

For example, *dbpedia:reference* for Skopje is http://www.skopje.gov.mk/en, while the *foaf:homepage* is http://en.wikipedia.org/wiki/Skopje.

Wikipedia articles consist mostly of free text, but also contain different types of structured information, such as infobox templates, categorization information, images, geo-coordinates and links to external Web pages. For example, Fig. 1 illustrates structured infobox data from Wikipedia for the City of Skopje. Some of the extracted infobox information for the City of Skopje in DBpedia, are presented below:

```
@prefix dbpedia <http://dbpedia.org/resource/>
@prefix dbterm <http://dbpedia.org/property/>

dbpedia: Skopje
    dbterm:officialName Скопје
    dbterm:longd 21
    dbterm:longm 26
    dbterm:areaTotalKm 1854
    …….
    dbterm:LeaderName dbpedia:Koce Trajanovski
    ……..
    dbterm:location Skopje
    ……..
    dbterm:populationTotal 506926
    ……..
```

With DBpedia search engines, you can find all available information about the City of Skopje, in different languages. DBpedia data can automatically link with other data sets with

owl:sameAs, which mean that two or more resource are identical (it is the same resource, in this case – Skopje):
<http://dbpedia.org/resource/Skopje>
Owl:sameAs opencyc:Mx4rvVj-GZwpEbGdrcN5Y29ycA
Owl:sameAs http://sws.geonames.org/785842/
Owl:sameAs http://umbel.org/umbel/ne/wikipedia/Skopje

Fig. 1. Structured information from Wikipedia

V. Access to DBpedia through the web

DBpedia enables making queries which give us answers, based on the data from Wikipedia. It is made with the help of the query language SPARQL (http://dbpedia.org/sparql). As we have already said, Linked Data is method for announcement of RDF data on the web, and connecting the data between different sources of data. We can access them with Linked Data browser, same as the traditional web document is accessed with HTML browser.

So, instead following the links between HTML web pages, the Linked Data browser enables the user to move through different sources of data, while following the RDF links. This enables the user to start from certain source of data and after that to move through potentially endless web of sources, interweaved between them with RDF links.

This also helps the robots from the Linked Data browsers to follow these links, so they can collect data from the Semantic Web. DBpedia set of data, is represented as Linked Data. This gives us the opportunity to display DBpedia data with Linked Data browsers like: DISCO, Marbles, OpenLink Data Explorer, Tabulator or Zitgist Data Viewer.

In addition we have a few Linked Data URI, from the DBpedia set of data. In order to start surfing through the Semantic Web, enter some of these URI in some of the browsers named above:

- http://dbpedia.org/resource/The_Lord_of_the_Rings
- http://dbpedia.org/resource/Berlin
- http://dbpedia.org/resource/Semantic_Web

VI. SPARQL query interfaces

DBpedia provides a public SPARQL endpoint at http://dbpedia.org/sparql, which enables users to query the RDF data source with SPARQL queries such as:

```
SELECT ?abstract
WHERE {
{ <http://dbpedia.org/resource/Skopje>
<http://dbpedia.org/property/abstract> ?abstract.
}
```

The query returns all the abstracts for the City of Skopje, given in each of the available languages.

The next query refines the abstracts returned to just the language specified, in this case ‘en’ (English):

```
SELECT ?abstract
WHERE {
{ <http://dbpedia.org/resource/Skopje>
<http://dbpedia.org/property/abstract> ?abstract.
FILTER langMatches (lang(?abstract), 'en')}
}
```

The SNORQL query explorer shown in the Fig. 2, provides a simpler interface to the DBpedia SPARQL endpoint. Fig. 2 shows both the query and the result returned for all famous persons who are born in Skopje.

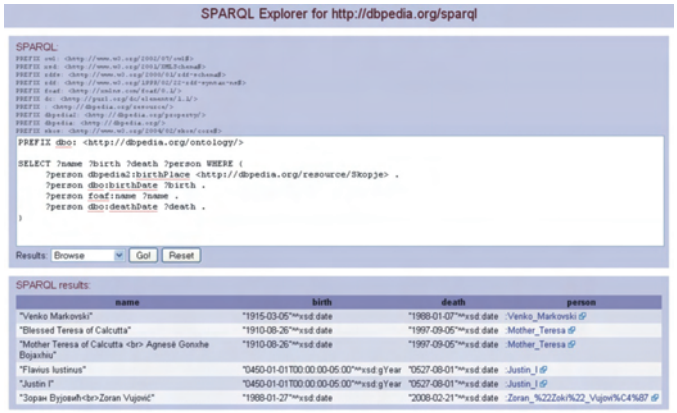


Fig. 2. SNORQL query and result

VII. DBpedia Applications

This is the list of applications (in no particular order) to get us started using DBpedia [5]:

Faceted Browsers

- **Faceted Wikipedia Search** – allows us to explore Wikipedia via a faceted browsing interface;
- **OpenLink Virtuoso built-in Faceted Browser, and Search & Find Service**, on the DBpedia host instance – offers several paths of DBpedia data exploration, starting from Keyword or URI or Label

User Applications

- **DBpedia Mobile** – provides a map view annotated with DBpedia entities and information from other knowledge bases
- **DBpedia Relation finder** – enter objects and find out what connects them (Shockwave / Flash-based)
- **DBpedia Navigator** – navigate through DBpedia data

URI Lookup Services

- **DBpedia Lookup** – find DBpedia URI for keywords

Query Builders

- **OpenLink iSPARQL Visual Query Builder** – drag & drop visual interface for construction of SPARQL queries against DBpedia and/or other data sets
- **DBpedia query Builder** – build your own DBpedia queries

SPARQL query interfaces

- **SNORQL Query Builder** – query DBpedia using the SPARQL query language
- **OpenLink Virtuoso built-in SPARQL endpoint**, on the DBpedia host instance

Browser enhancements

- **DBpedia UserScript** – enhances Wikipedia pages with links to their corresponding DBpedia page

VIII. Conclusion

The Semantic Web represents evolutionary development of the World Wide Web, in which the meaning (semantics) of the information and the web services are defined, which helps the web to "understand" the user requirements. It appears as a result of the Tim Berners-Lee's vision, as universal media for exchange of data, information and knowledge across the web. Tim Berners-Lee calls the resultant web of Linked Data, the Giant Global Graph, in contrast to the HTML-based World Wide Web.

The DBpedia knowledge base is served as Linked Data on the web. As DBpedia defines Linked Data URI for millions of concepts, various data providers have started to set RDF links from their data sets to DBpedia, making DBpedia one of the central interlinking-hubs of the emerging Web of data. Already today, the resulting Web of data around DBpedia, forms an exciting test-bed, to develop, compare and evaluate data integration, and to deploy operational Semantic Web applications.

References

[1] Ivan Herman, "Tutorial on the Semantic Web" – W3C: http://www.w3.org/People/Ivan/CorePresentations/SWTutorial/Slides.pdf , pp 42-45 and pp 131-140, 2010.

[2] Chris Bizer, Richard Cyganiak, Tom Heath, "How to Publish Linked Data on the Web" - http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/, 2008

[3] Joshua Tauberer, "What is RDF" – XML.com: http://www.xml.com/pub/a/2001/01/24/rdf.html?page=2

[4] DBpedia homepage, http://dbpedia.org/About

[5] DBpedia applications, http://wiki.dbpedia.org/Applications

[6] W3C Recommendation, "SPARQL Query Language for RDF" - http://www.w3.org/TR/rdf-sparql-query/ , 2008

POSTER SESSION PO IX

PO IX - Computer Systems and Internet Technologies III

System Architecture for Capital Budgeting in Non-Financial Institution

Galina T. Naydenova[1]

***Abstract* – This article will be provided universal structure of management system of the modeling business transaction and risk analysis for a long- term horizons and possible links between the modules. The system will allow managers of corporations to make better investments and/or financial decisions. It will propose methods for interpolation and extrapolation with the ability to minimize errors.**

***Keywords* – Capital budgeting, Monte Carlo Simulation, Interpolation, Extrapolation, Financial model, System structure**

I. INTRODUCTION

Capital budgeting [8,9] is a process of generation, estimation and risk indexing and choosing of the best investment project of point of view of maximize firm's value.

Capital budgeting is one of the most important areas of decision-making by financial manager. It is from this perspective is increasingly necessary methodology for calculation and forecasting the financial agreements or risk analysis of investment project in conditions of market uncertainty. Market uncertainty arising from market conditions, by participants, on misinformation, competence and experience of the financial manager. From the perspective of financial managers, the most important aspect in risk analysis of investment project is the impact of project over interrelation risk-profit of the firm.

This document is occupied with the task to describe the architecture of the system that will help the corporate manager for project evaluator.

The purpose of financial manager is to select those projects that have greater importance to maximize the wealth of corporation. This will be described in the next point of this paper.

II. SYSTEM STRUCTURE

The aim is to create a flexible system with which to predict of estimation values of the financial model or agreement.

The system consists of several modules interconnected with each other.

The first of them is the ***Market variables***, which consists of a set of values of market variables. These market variables may be interest rate, exchange rate (all possible), cost of product manufactured and obtained from reliable world-famous sources. All in one degree or another is correlated. This in turn leads to a correlation matrix with a huge amount and this leads to the cover of many system resources and time for calculation.

In making prediction and taking the right decision for future period financial managers should conform with the phases of the cycle. The economical cycle [15] is divided into three stages of economic expansion - early, middle and late, and conditional of two stages of economic contraction (recession) - early and late. There are sectors that are covered completely with this cycle (Fig.1). For example, automobile meets the phase of "early economic recovery", electrical engineering of the "average expansion", production of ferrous metals' – "latest economic expansion", the food industry – "early economic contraction" and with the insurance business – "recession later".

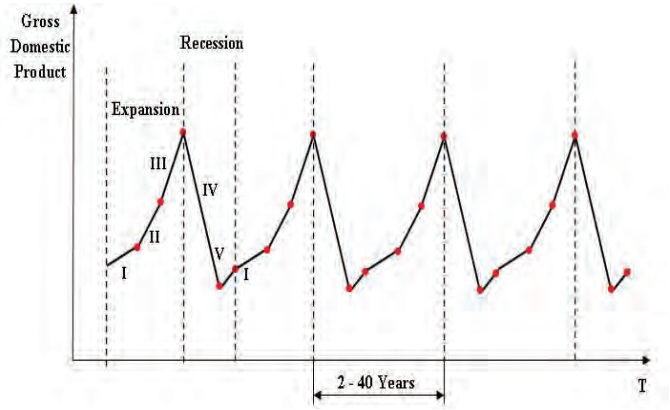


Fig. 1. Phases of economic cycle

Financial manager makes projection for future projects in two ways: intuitive or based on the economic cycle. Through the second manner is made selection (according to the perspective of the manager) of the best horizon from previous years, which meets in such conduct of future expectation.

The proposed system (Figs.2, 5 and 6) for planning and forecasting projects, those spans of time is used to minimize the correlation matrix calculations and increase performance. This information retrieval is proposed in the second module of the system – ***Define Market***. Not only, here are included necessary part of the history of market variables but their volatility and correlation between them and the estimated value of the financial model (Fig. 2). Volatility and correlation matrix are input values for Monte Carlo simulation. Forecast values can be depicted graphically. Unlike the spreadsheet software, in which by changing the values in the table change

[1]Galina T. Naydenova is with the Faculty of Computer Since and Technology of TU-Varna, Studentska 1 str., 9000 Varna, Bulgaria, E-mail: galina_m@abv.bg

chart type, in this system will be possible here also not only this but and return option to change the values in the table by manipulating the ***2D graphics***. This gives greater freedom of financial managers to determine the width of the volatility of the forecast values according to his expectations.

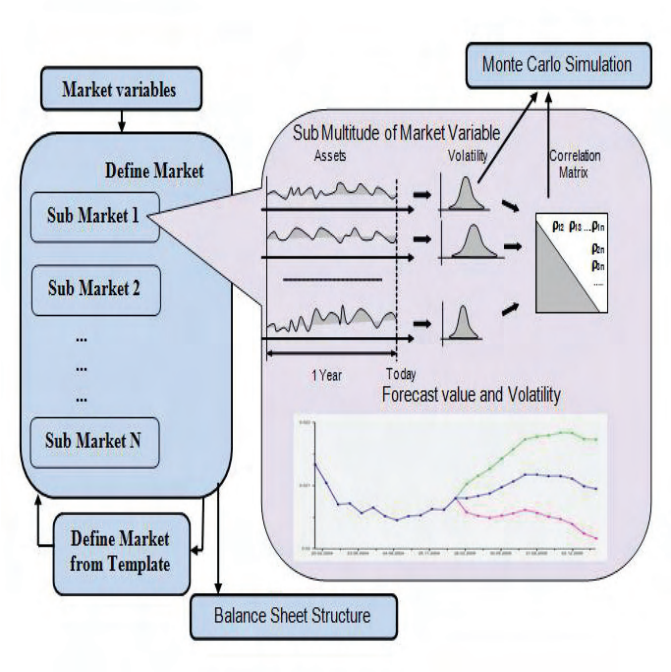


Fig. 2. Creation and Define Markets

It should be mentioned that taken up by the time span of time of historical data is possible for one reason or another to missing data.

This will lead to problems in the calculations that must be performed by the proposed system for risk management. Solving the problem is by using a linear interpolation method [3].

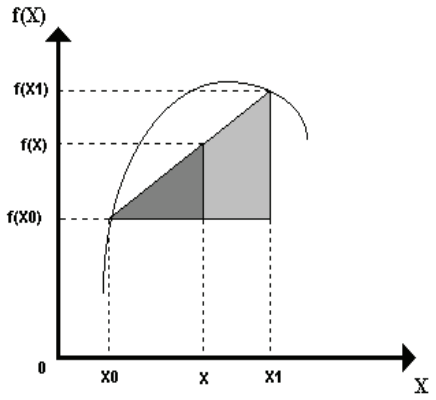


Fig. 3. Linear interpolation method

The simple method of interpolation, shown in the Fig.3, is to connect two points in a straight line. And this can be described with Eg.1:

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \quad (1)$$

Interpolation is the analytical method of estimating output values inside the range of tabulated, known or measured data points. This method gives relatively accurate results only in case where the distance between two points is very small or close to each other. It will be use in building the system.

Extrapolation is another method by which performs detection of missing or estimated future points. Extrapolation is the process of taking data values at points x_1, ..., x_n, and approximating a value outside the range of the given points. This is most commonly experienced when has periodic repetition in the case of economic cycles and that data is used to approximate the next data point. For example, forecast values predictions take historic data and extrapolate a future forecast scenarios. Extrapolation can be used as a tool for predicting economic phenomena.

It will consider four variants of extrapolation:

First variant

By extrapolation method finding a point which lies outside the scope of the sample can be done only with approximate accuracy in the vicinity of the last point of the range. The formula for calculating the extrapolation Eq.1 is the same as interpolation. Depending on the values that have starting and ending point of the sample is assumed and the direction (up or down in Fig.4 – blue line), which is expected to find a future point for a long time. The disadvantage of this method is that a large interval of time finding the point or value will be a big mistake.

Second variant

To avoid the disadvantage of the first variant can be assumed that all points before and after the interval adopts a value of two end points thereof (Fig.4- orange lines). This will minimize the error in the setting of a large space. It uses interpolation formula for finding points in the near surroundings and the proposed way to remove error in larger intervals.

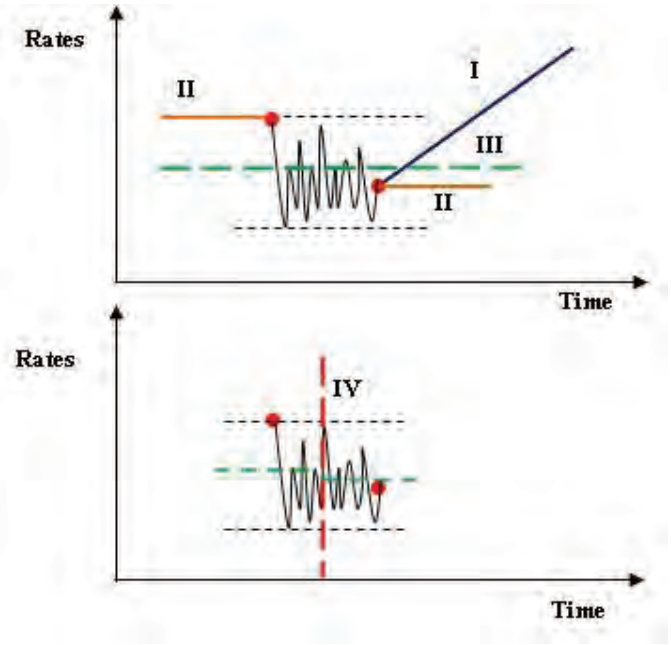


Fig. 4. Extrapolation methods

Third variant

Since the second variant also has a shortcoming here is offered the opportunity to find the average of the sample (Fig.4 – green line). Thus the error becomes even smaller. However, this also is insufficient to reach as close as possible to the finding value of market variable.

Fourth variant

The above three variants not offered sufficient reduction of error. For this is fourth version offers.

In the third variant the two endpoints of the range are equal value after averaging it. This will lead to a problem in cases when extreme situation occur different from that of the normal behavior of market variables.

For example, in cases where have unusual trend of the price distribution for two consecutive months. For instance, lowering the price of a product together with a large user activity before a specific event and then increase the price and lowering the interest of the consumer to that product. In such situation, it will get an extreme point, different from the average of the range produced by a third variant. This will lead again to an error.

Therefore it is proposed to be made conditional split the sample of market values. Each half is to find the average value (Fig.4).

Thus, finding the expected value for a long time interval is suppose to be much more accurately than the proposed three variant above.

Thus, by methods of interpolation and extrapolation is proposed to solve one problem that will arise in the system.

One advantage of this system in comparisons with spreadsheet software is that is possible to quickly and easily create a new market from existing. The goal here is to change some values to the perspective of the manager without much effort and waste of time [6]. This is the third module of the system called ***Create Market from Template*** (Fig. 2)**.**

Simulation is a statistical – based behavior approach that applies predetermined probability distribution and random number to estimate risky outcomes. By typing the various cash flow components together in a mathematical model and repeating the process numerous times, the financial managers can develop a probability distribution of project returns.

Monte-Carlo simulation is another module (Fig.5) of the system. Monte-Carlo simulations are popular in financial applications, as they are able to value multi-dimensional financial instruments or models. The simulations are easy to specify mathematically, with an obvious translation to sequential code, and so can be easily implemented in software.

As mentioned above, the input values for Monte Carlo simulation is the correlation matrix and the volatility of market variable. After receipt and multiplication in the simulation is obtained covariance matrix is undergoing transformation as a result is obtained Cholesky matrix.

Through a random generator to generate random numbers normally distributed. These normally distributed numbers are presented with a normal distribution bell-shaped with range from 0 to 1. After generating their numbers are presented in Deviation vector, which is obtained by multiplying Cholesky matrix. The result of this multiplication is co variation vector. He in turn is summed with a series of forecast values, provided by a variety of established markets. Steps listed here to be repeated multiple times from 1000 – 5000 runs. As a result of this entire scenario is derived vector, which applied to financial instruments or portfolio calculations and from here the presentation of cost allocation and generate the results of Monte Carlo simulation. These results may be: Value at Risk, Expected Value, Expected Loss, Confidence Value, CFaR/EaR, etc., which are presented in normal distribution form.

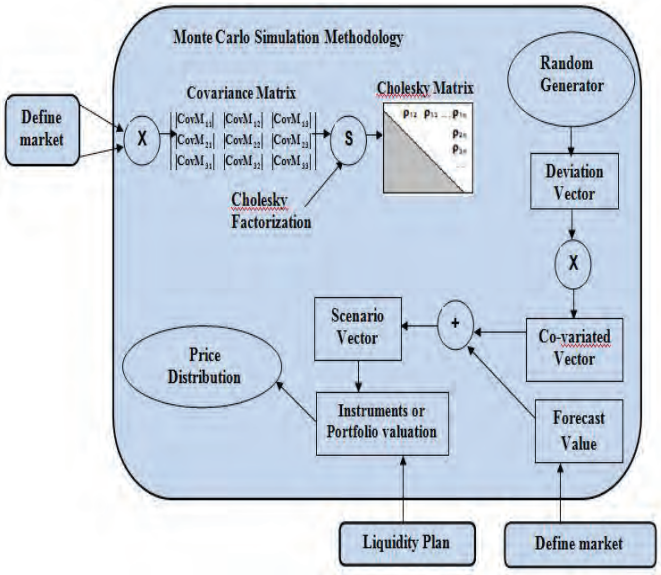


Fig. 5. Monte Carlo Simulation

Generated results from Monte Carlo simulation are recorded in the ***Liquidity Plan module***. Then here can be compared with planned values and this forecast results generated by Monte Carlo simulation.

Ones created several submarkets in the *Define Market* module, and be made Monte Carlo simulation of the same financial model is provided a comparative and assessment of different forecast scenario by ***2D or 3D graphics***.

Liquidity Plan module and *Balance Sheet Structure module* have hierarchical structure but Liquidity Plan module is develop in the time. This means a description of the hierarchical structure of all cash flows and market variables distributed in time periods.

The structure of the ***Balance Sheet Structure module*** provides for a description of financial models. They may include derivative instruments, estimation project model of capital budgeting (NPV, IRR, etc.) or other complex models. After describing the module structure, each of the nodes and sub nodes must be connected with market variables provided for this model. To the Balance Structure includes three modules that transmit the flexibility of the system. This are *Mechanism for determine signs of + and - module, Mechanism for assign of arbitrary time screen module* and *Formula Editor module.*

It is provided in the Balance structure a ***Mechanism for determine signs of + and -***. That gives the possibility of unification and subtraction of sub nodes in the main node.

A ***Mechanism for assign of arbitrary time screen*** is necessary for modeling of different balance structures with possibility for given time intervals, different from initially defined, for example if market variables are defined with value through 3 months, and if necessary the calculation to be

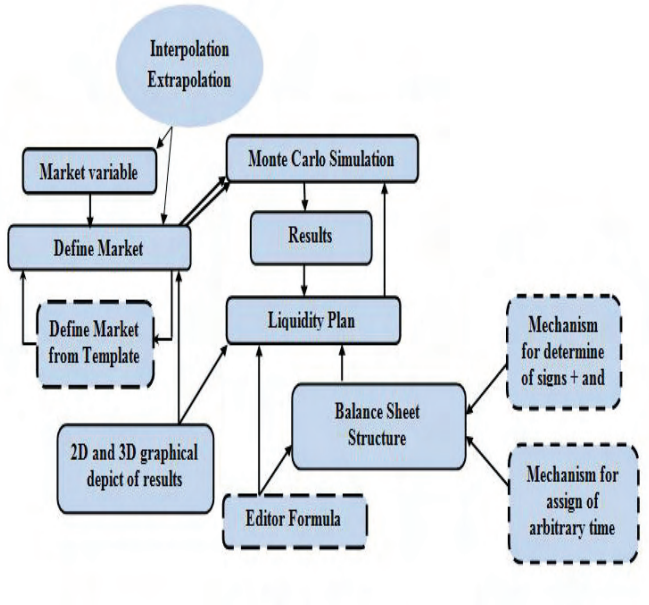


Fig. 6. Management System Architecture

changed through 1 month. That gives a possibility for modeling of different contract and gives fast computation of different at given time intervals and without need of help from firm supporting the software for calculation of price model.

Formula Editor module is projected for calculation of more complex expression consisting of values from nodes and sub nodes. Since in the *Balance Sheet Structure* can only add and subtract nodes or inheritor that module gives the right of supporting operations as addition, subtraction, multiplication, division, logical if, average, natural logarithm, exponent and grade. *Formula Editor* module is in help of Balance Sheet Structure module and Liquidity Plan module.

By calculation of financial model for every next Monte Carlo run each of nodes of hierarchical structure are typed set of values of the temporary memory. All of this sub notes are collected in the main node and result for complete financial model can be view with ***2D or 3D Graphic*** **module**.

III. Conclusion

In document has proposed a management system structure. It will be in help of financial manager who must be able to decide whether an investment is worth undertaking and be able to choose intelligently between two or more alternatives. To do this, a sound procedure to evaluate, compare, and select projects is needed. Each of this principles lied of capital budgeting conception. The system structure depicts each of these principles by flexible way and it is suggested a calculation of more sophisticated financial model. All modules shows in Fig. 6 with a broken line, transmit flexibility and uniqueness of the system compared with spreadsheet software.

Presented are linear interpolation and extrapolation methods to calculate the market values of samples in which there is missing data or to predict futures point based on historical occurrences. Extrapolation approach is proposed to minimize the error. There is no absolutely best method but only the optimal choice under certain circumstances.

Future work will be focusing on building of algorithm of the system. This system architecture is foreseen to work under Windows operation system and Oracle database.

References

[1] Eugene Stern, “Risk management for non-financial corporations”, RiskMetrics Group JORNAL vol 5, no 1, pp 42-70, Winter 2004

[2] Glasserman Paul, *Monte Carlo method in financial engineering*, New York, United States of America: Springer Science + Business Media, Inc., 2004

[3] Hagan, Patrick S.,Graeme West, “Interpolation Methods for Curve Construction”, Aplied Mathematical Finance JORNAL vol.13, no 2, pp 82-129, June 2006

[4] Haugh, M., *The Monte Carlo Framework, Examples from Finance and Generating Correlated Random Variables*. Course Notes, 2004

[5] Kim, J., Allan, M. Malz, Jorge, M., *Long Run Technical Document, 1st ed*; New York: RiskMetrics Group, 1999

[6] Lee, A., Kim, J. Malz, A. M., *CorporateMetrics Technical Ddocument*, New York: RiskMetrics Group, 1999

[7] Lui, Jun S., *Monte Carlo strategies in scientific computing*, New York, United States of America: Springer, 2001

[8] Petersons, Plamena P., Fabozzi, Frank J., *Capital Budgeting: Theory and Practice*, Canada, Wiley&Sons, Inc., 2002

[9] Ross, Marc, “Capital Budgeting Practices of Twelve Large Manufacturers”, Financial Management JORNAL, University of Michigan, vol.15, no 4, pp 15-22, Winter 1986

[10] Rubinstein, Reuven Y., Kroese, Dirk P., *Simulation and the Monte Carlo Method*, 2nd ed; Hoboken, New Jersey, United States of America: John Wiley & Sons, Inc., 2008

[11] Savvides, Savvakis C., *Risk Analysis in Investment Appraisal*, Beech Tree Publishing, Project Appraisal, Volume 9, Number 1, pp 3-18, March 1994

[12] Tjia, John S. *Building financial models. A guide to creating and interpreting Financial Statements*; New York, United States of America: McGraw-Hill, 2004

[13] Vause Bob, *Guide to Analysing Companies*, 2en ed, Sofia, Bulgaria: Classic &Style, 2006

[14] Дочев, Дочо, *Теория на риска. Управление на инвестиционния риск*, Университетско Издателство ИУ – Варна, 2001

[15] Цанков, Теодор, *Финансов Инженеринг*, Сиела – Софт енд паблишинг, София, 2005

Graph-Based Analytical Approach to Testing Programs

Mitko M. Mitev and Pavlina Vladimirova[2]

Abstract: **In the paper is presented an approach based on theoretical properties of graphs for structural interpretation of programs and opportunities are provided for planning and structural testing programs. The approach includes a number of developed tools, providing specialized programming environment for testing software modules.**

Keywords: **graph theory, generators of input streams, roads and contours, programming controls**

I. INTRODUCTION

The process of testing is one of the milestones in software technologies for industrial production of programs. It is characterized by the following major parameters:

- *Type of the test system:* ordinary software systems, systems with distributed databases, network systems, real-time systems, Internet applications etc. The technology and used methods of testing directly dependent on the type of tested system.
- *Object of test:* from separate programming procedure, through testing the functionality of the developed classes and ends with a series of tests of the software system.
- *Goals of tests:* detection and localization of errors at different stages of used software technology, examining the various operating parameters, determined in the technical assignment, such as productivity, response time, a priority system, behavior in extreme conditions, data protection and opportunities for recovery etc.
  - *Used methods*: different types of internal and external tests, real-time tests etc.
  - *Tools:* input stream generators, programs for simulating phenomena and processes, fictitious modules etc.
  - *Compatibility with the design process:* preliminary - based on preliminary project, accompaniment software system development and final – after the final development of separate parts or the whole system.

The proposed graph-based analitical approach for testing refers primarily to local executed program modules for detection and correction of errors at the design stage. It is in the class of internal, structural tests, called "white box" method.

[1]Mitko M. Mitev, Technical University - Varna, Department of Computer Sciences and Engineering, Studentska 1, 9010 Varna, Bulgaria, E-mail: mitevmm@abv.bg

[2]Pavlina St. Vladimirova, Technical University - Varna, Department of Computer Sciences and Engineering, Studentska 1, 9010 Varna, Bulgaria, E-mail: pav_varna@yahoo.com

II. FORMALIZATION OF THE PROBLEM FOR TEST

For internal test it is necessary to represent the programming module as a structure [3]. Therefore, in every program module the structure determined operators have to be found. Without violation the algorithm of the program, in accordance with the semantics of the operators, they are transformed into so-called IF structures. This allows the control structure of the program to be presented as a finite oriented graph *G (X, U, P),* where:

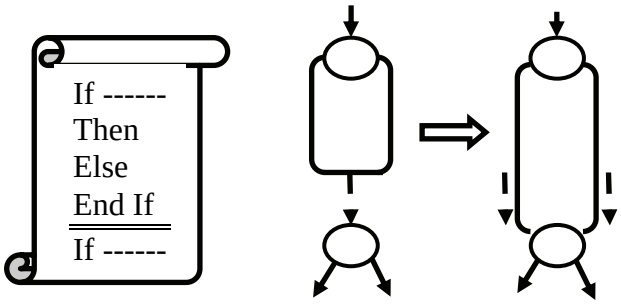
X is the set of vertices. Each vertice represents IF operator of the modified structure.

U – Set of edges of the graph. Each edge represents a linear section of the program. The edge must be oriented and must connect two vertices.

P – Incidentor, which is three-place predicate and introduces relationship between the vertices and arcs of the graph. The predicate takes value true, if two specific vertices and edge are incident each other. Otherwise, the predicate takes value false.

In this transformation there are two special cases:

- Sequence of two IF operator with linear section between them (the example on Figure 1 is in *Visual Basic. Net*). By analogy, the same is true for other programming languages.



a b c

Fig.1. Transformation of the IF structure with a conecting linear section

a) code of the program
b) performance with impaired connectivity
c) conversion with duplication of the common linear section

- Cycle with post condition (Fig. 2)

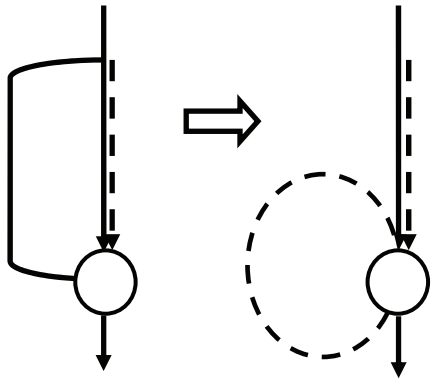


Fig.2. Transformation of cycle with post condition
a) the presentation of impaired connections
b) conversion with duplication of cyclic section

Such presented program can be tested with known methods[1,2].

Let's is given the set of data $D = \{d_k\},\ k \in \{1, K\}$, which is usually classified as input, output or intermediate. On the other hand input data can be divided into data for calculations, control and diagnosis. By analogy, output data are numerical results, diagnostic or data used for control (structuring) of the next software modules.

For the purposes of testing data can be classified as:

- data involved in the composition of the operands, but does not change its value,
- data involved in a decision making,
- data, which during the execution change their value,
- double or triple mixed data from the previous three groups.

On this basis, for each data in the structure of the program may be specified path (paths) of its movement. Each path starts from the place in the structure of the program, where the data appears for the first time and ends with its last use. The path of the input data, represented by its values, starts from the moment of calling the module for execution. The path may end in the logical end of the module if they appeared as results included in the interface.

The path of each data can be described as a sequence of vertices and edges. For example, let $X = \{x_j\}, i \in \{1, N\}$, where $N = |X|$ and $d_k \in D$, with initial vertex ***i*** and vertex *m (m> i)*, one of the possible paths $s_{i,m}^k \in S$ of testing is described as

$$x_i, u_{i,i+l} x_l, u_{l,l+1} \text{K K K } u_{l+n,m} x_m \qquad (1)$$

In particular, it is possible *i = 1* and *m = N*. This is a case, when data is both input and output. Otherwise, it is an input and participates in the formation of intermediate results or it is output, but caused of the execution of the program module. Let's are given two vertices $x_i x_j \in X$ and $i \neq j$ and $P(x_i, u_{i,j}, x_j)$ is true, i.e. $u_{i,j} \in U$. Therefore, can be determined following, generally crossed subsets of data, i.e. the conditions in the two vertices is determined by data included in them

$$\{d_k^i\} \subset D;\ \{d_k^j\} \subset D \qquad (2)$$

Similarly,

$$\{d_k^{i,j}\} \subset D \qquad (3)$$

are data included in the calculation of the linear section of the program.

Paths of two data are *directly independent* unless they include common elements such as vertices and edges. Therefore, these paths can be differentiated as two independent tests. Otherwise, they are *directly dependent,* because they have common data, which in accordance with the algorithm can determine changes in values or paths of execution.

It is recommended directly dependent paths to be tested independently of each other in an exchange sequence. Implementation of the second test will depend on the results of the first test and contrariwise. This sequence does not exhaust all possible cases of directly dependent paths, but decided the main ones.

Two routes are *indirectly dependent*, if there is a third path, which is directly dependent on the previous two paths, because its implementation could lead to termination of one (or both) path or to include other paths. The concept of *indirect dependence* can be extended to two sequences of directly dependent paths, but obligatory finish with a path, which is directly dependent on paths in the last two sequences.

Testing at this situation can be done in reverse order, starting with the path connecting the two sequences and alternatively the rest directly dependent paths of both sequences are changed.

This approach is convenient for a description of a linear section of the programming module. In the case of cycles, calling the other modules for execution, implementation of early or late binding or creation instances of classes, the approach is not applicable due to its low efficiency.

III. METHODOLOGY FOR TESTING

Testing of the program module includes the following steps:

1. *Preparation of the module.*
   - Determination of the programming module for testing and placing it in the program environment.
   - Analysis of program text and identification control statement.
   - Select the linear sections (if any), which are out of a control statement and its subsequent inclusion in the linear sections of the block structure of the previous control statement.
   - Finding of cycles with post-condition (if any) and duplication of the cycle body, according to method, suitable to the occasion.

- According to the text sequence, numbering of all real and additionally included IF operators, resulting from the transformation of the control statements.
- Establishment of the related linear sections in the obtained IF structure and respectively their dual indexing.
- Tabular or list representing of the structure as a finite oriented graph ***G (X, U, P)*** with a completely determined sets and values of the three-place predicate.
- For each vertex or edge is determined the data, included in their composition, whish are a subset of all data.
- An unique link is determined between the vertices (2), edges (3) and subsets of data.

2. *Formation of paths and marking of test sequences.*
- For each of the specified data is determined the location of its first appearance and respectively the final point of its existence.
- According to (1) are determined the paths for the existence of the data in the program. For this purpose in the graph G is looked for routes from the specified initial and final vertex.
- The routes are tested in pairs in order to establish their *direct independence*. Firstly these routes are numbered in increasing sequence (counter and identifier of the test path in the program are formed).
- If there are any *direct dependence* paths, tests are numbered with the next counter values by alternative change of execution of the two roads.
- If there are proved *indirectly dependence* on the third or other paths then have to analyze the sequence of roads with *directly dependence* until be found a direct dependent path.
- If such path exists, then the test counter begins with it and alternatively increases in both branches till reach the first two routes.
- Otherwise directly dependent paths are marked by incrementing counter.

3. *Preparation of data and testing.*

Tests are carried out according to their identification number. For that purpose have to prepare the necessary data:
- The input data are determined, also their range of variation, accuracy of presentation, etc.
- The data are classified into specified four groups.
- For each data are defined two classes: correct and incorrect values.
- All data involved in testing path obtained value of the class of correct values and next value of the incorrect class.
- Tests are carried out in a test environment with consistent monitoring of the executed path and temporary and final values.

4. *Analysis of the results.*

The analysis of the results is performed by well known methods of direct comparison or mathematical statistics when a comparison is impossible.

III. Conclusion

In the report is presented a way for transformation programming text into so-called IF structure and its subsequent interpretation as a finite oriented graph. An analysis of the used data is introduced and their classification is proposed, according to their actual participation in the calculation process and the decision making, depending on the conditions in control statements. On this basis, subsets are defined for every linear section or control statement. As a result, is proposed methodological sequence for determining test routes according to their mutual dependence (or independence).

The proposed theoretical algorithms are developed and implemented as a software library of tools. They can be used alone or be included in the composition of a software system for testing.

References

[1] J. Zhao, "Data-Flow-Based Unit Testing of Aspect-Oriented Programs", Department of Computer Science and Engineering Fukuoka Institute of Technology, Japan,2001
[2] P. Jorgensen, *Software Testing. A Craftsman's Approach*,2nd ed.-,CRC Press,2002 , 359
[3] S. M. Kauffman, "Graph Theory and Linear Algebra" , 2007,105 http://home.comcast.net/~smka2436/Graph_Theory_and_Linear_Algebra.pdf

This page intentionally left blank.

A Study of Information Technology Use Among Students at South East European University

Marika A. Trpkovska[1]

***Abstract* – The main intend of this research was to study and point important issues in the human computer interaction through existing positives and drawbacks of students' information technology (IT) use at South East European University. The study was conducted to offer better education for the students by proposing information technology in learning and teaching process using optimal resources, minimal costs and maximal results.**

***Keywords* – Human computer interaction, information technology, teaching and learning process, enhanced education.**

I. INTRODUCTION

The interaction between human and computer supports the idea of designing, evaluating and implementing the interactive computer systems for human use [2].

For most students information technologies (IT) are essential for both the place of work and activities in everyday life's. Information technologies are part of how the students achieve knowledge, how they communicate and understand with each other.

The main requirement for students to benefit from this novelty depends on the level of achievement and understanding of these fundamental technologies and achieving essential technological skills.

Followed by the previous research studies made by other educational institutions like ECAR project [1] which focuses on positive and negative concerns about information technology use among students in higher education, the main point of this research was to investigate what students are doing with the existing IT technology in the SEE-University and how this technology affects on their experience. Several suppositions were given in this context:

1. Students insist greater use of IT in teaching and learning process.
2. Students prefer taking courses that use IT in the curricula.
3. Students are more engaged in courses that require using technology.
4. Students think that the technology has positive effect on their learning experience.
5. Students want more advanced technology to be used in teaching and learning process.
6. Students need more training or education in the use of IT.
7. Students can only benefit if IT is included in courses.

The general purpose of this research is to find evidence that supports these suppositions which will supply on proving that IT contributes to the educational processes.

In this connotation I was investigating the student's use and perceptions which focuses on their experiences and opinions about SEEU information technology background. For this reason an online questionnaire was e-mailed to undergraduate students from South East European University and it was completed by 250 participants from several departments including Computer Science direction, Public Administration, Business Administration, Law and Teacher Training Faculty. Some of the previous suppositions of the research were confirmed and some bring a level of surprise. The questionnaire included 17 questions sufficient to point to the student's expectations, preferences and opinions about various aspects of information technology use in South East European University (SEEU).

As it was mentioned previously, respondents were students familiar with information technology, its uses, and its unique characteristics. The modern, interactive and collaborative qualities of IT equipment use at SEEU were regarded as important reasons for using them. The results of the questionnaire indicated a high point of student's satisfaction with most aspects of the information technology use. In the following we will take a brief look of the results.

II. OUTCOMES OF THE QUESTIONNAIRE

In order to support the above suppositions I divide them into several categories based on area of correspondence.

A. Influence of IT on learning Student's preferences regarding the use of IT in courses

In order to extract the student's preferences, the levels of IT skills presented in ECAR study [1], helped me to confirm student's needs related to Information Technology use. In this context I came out with information that undergraduate students prefer taking courses that use information technology rather than not using it which is confirmed with the below results. As it's shown in Fig. 1, there is positivity among students on the use of IT. 31 percent of the students prefer technology to be used at moderate level (e.g., e-mail, several PowerPoint presentations, some online activities or content), 47 percent respond extensively (e.g., class lecture notes online, computer simulations, PowerPoint presentations, streaming video or audio, etc.) and 11 percent exclusively

[1]Marika A. Trpkovska is with the Faculty of Computer Sciences at South East European University, Ilindenska bb, 1200 Tetovo, Macedonia, E-mail: m.apostolova@seeu.edu.mk

(i.e., are entirely online with no required face-to-face interactions) in the courses.

For that reason, first and second supposition where students would like greater use of IT technology in teaching and learning and students' aspiration that IT should be included in course curricula are proved. Therefore this information should encourage teacher to use more technology in the teaching process.

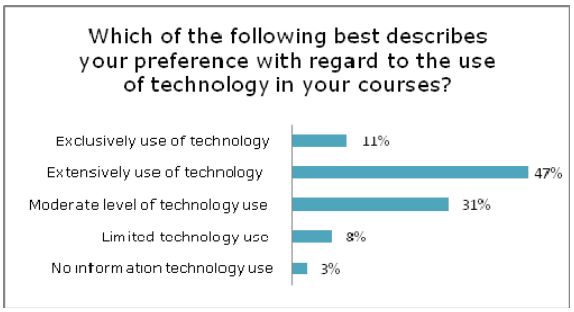


Fig. 1. Student's preferences regarding the use of technology in courses

B. Training or education in the use of IT

Summarizing both levels moderate and extensively technology use in Fig. 1, we get that exactly 78 percent of students are more comfortable using fundamental information technologies such as e-mail, online messaging, Power Point presentations, Word processing and some advanced features like audio and video streaming but they rate themselves as not skilled in their use. The big percentages of the respondents give an inflection on not having many of the necessary skills to use IT as a support of academic work.

This information plus their significant need for additional training in the use of IT supporting learning and problem-solving skills, illustrated in Fig. 2, can confirm that besides of the crucial need and positive impact of the IT in the teaching and learning process, priority for the university should be to offer training or education in the use of IT. With this conclusion the sixth suppositions is confirmed.

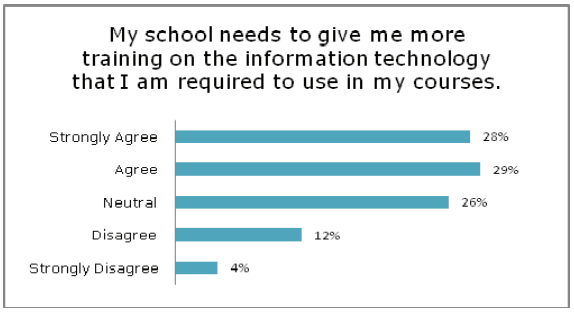


Fig. 2. Necessitate of IT training for students

C. Engagement of students in courses with IT surroundings

The questionnaire results indicate that the students are more engaged in courses that require using technology rather than not using it. This is result from the fact that students want to involve themselves in a rapidly changing world in which every day work and other activities are more and more transformed by contact to various information technologies.

In the Fig. 3 we can see clearly that 42 percent of the respondent's agree and 26 percent strongly agree with this (third) supposition.

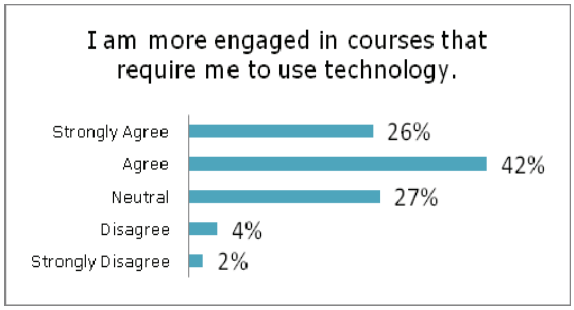


Fig. 3. Engagement of students in courses with IT surroundings

D. Benefit of IT in learning

The seventh supposition which specifies that the students can only benefit if IT is included in courses, is confirmed as well. Students' responses illustrated in Fig. 4, show that the most valuable benefit of using technology in courses is Management of course activities like downloading course materials, sharing free time (36 percent) and learning (30 percent). Better communication with classmates and instructors (22 percent), followed by convenience (9 percent) are next. Only 3 percent of the students identify no benefit at all from using technology in courses.

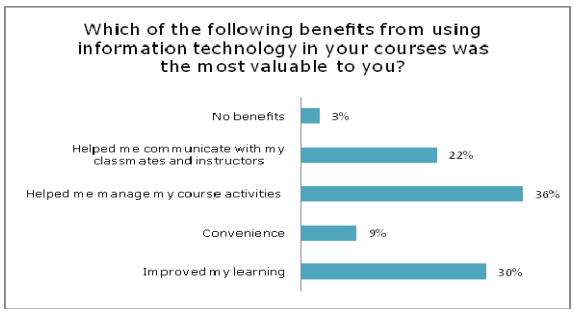


Fig. 4. Benefits from IT in the classroom

E. Impact of IT in learning

The student perspective that the technology is improving their learning experience is also confirmed. While using technology to enhance their learning is the second highest benefit point out when students are required to choose just one benefit, students agree that IT in courses improves learning. 50 percent of respondents agree and 26 percent strongly agree (total of 76 percent) that IT in courses improves their learning illustrated in Fig. 5. This results support the forth supposition.

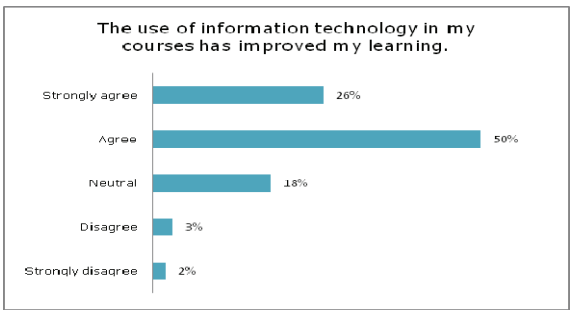


Fig. 5. Using IT in classroom improves student's learning

Consequently with above conclusion I can add that student's comments in the open-ended question also indicate on the importance of the instructor's skill on the learning outcome, despite of their negative experiences with the existing technology. Students also think that IT, used in appropriate way, will improve the qualifications of the instructor as well.

F. Advanced technology in teaching and learning process.

Besides the modern IT infrastructure [4] that SEEU offers to the students, 68 percent of them are not very satisfied with the existing technology. Students believe that instructors should use some advanced IT facilities which will contribute on enhanced education.

Leaded by the world trends and the human psychology, at first place students ask for advanced technology not wondering whether they need more advanced technology or whether the whole capacity of the existing facilities is exploited. Everyone wants a fast, stable, and more relevant technologies intended, but if we look back, past generations of SEEU students had outdated technology [3] and still were able to program, present, communicate, learn and work on satisfied level for today's conditions.

What is said above is not in the context that there is no need for newer and more modern technology, but that there is a need of technology that will offer significantly more than the preceding, but of course for a reasonable price, greater utilization and greater compatibility with the educational directions outlined in the program of the University.

Modernization of the education requires upgrading of IT technology, and vice versa. Optimal changes are needed in both fields in order to achieve the goal and that is the modern education.

III. STUDENTS' EXPECTATIONS OF INFORMATION TECHNOLOGY.

SEEU has spent considerable resources on technologies aimed to satisfy students' preferences, expectations and facilitations. Therefore I used questionnaire data, both quantitative and qualitative, to describe students' expectations of IT in three different districts: convenience, connection and learning.

Students' expectation in area of convenience is to have up-to-date modern technology, online resources and services which will satisfy their education needs. Moreover these facilities should be always available, accessible anytime and from anywhere, supported by stable network and professional technical support team.

From the other side when the connection is taken into account, students' stress the need of mobile electronic connections, combined with personal, customizable and portable IT devices like laptops with a possibility to be networked for online communication in order to be able to access existing resources and services at SEEU and to utilize them in their daily work.

And finally, concerning the expectation in learning district, students' advocate class with rich IT surroundings, combined with more practical, professional, experimental and inclusive participation.

Perceiving the students' experiences and expectations regarding the use of IT, of what was said above, I can conclude that the positive constraint of above suppositions depends on the real ability of faculty to use IT technology effectively in the teaching process and the negative depends on the existing up-to-date technology offered to students.

IV. RECOMMENDATIONS

Summarizing the student experience and expectations in area of IT, I can point out the following three concerns that have to be taken into a consideration.

First concern is related to SEEU IT support. Guided by the qualitative comments of the students and gathered information on their likes, and overall satisfaction with the existing technology at SEE-University, I have to express student's dissatisfaction when networks are down, technical support isn't available when needed, or existing technology will prevent the completion of daily duties which are indented to students. Knowing this concern, IT support team is the one who is charged for the implementation of the technology environments in which students are learning, and without a core set of reliable IT systems and services, students will not entirely accept technologies to enhance the learning and also teaching environment.

Second concern is related to the University representatives. In this context, if we take into a consideration the research results, I believe that University has to pay attention to the following areas:

- incorporation of IT into the course curricula,
- available IT training for students and instructors (Students need to know how to study with the new technologies and instructors how to use them)
- IT service and support

Finally the third concern refers to the teachers and points on the following. The interactive and collaborative qualities of diverse technologies were viewed as significant reasons for using it. The research results point out a high student's satisfaction with most aspects of the information technology use. Furthermore, it is true that many new applications and tools exist and will continue to advance. If we take into consideration students' preferences conveyed on more information technology use in teaching process, this should

convinced teachers to value these innovations, to use them and to be trained in their use.

V. CONCLUSION

Humans used to develop tools for making their life's easier. Automation of processes made people to be released of physical work with goal to have more time to advance psychologically. IT and especially HCI are developed for that goal. The question is: Is that goal achieved, and is the use of IT justified?

It is assumable that the use of IT contributes to educational processes. Today's generations are accustomed to technology. Education is process that must evolve, changing the methods and resources, to adapt to current needs of society. This research proved that students are more comfortable in learning, communicating and relating to educational processes when they are using the wealth of IT. The use of IT makes achieving knowledge for students easier, interesting, interactive, and modern.

Assuming this I can say: Yes, everyone must use IT for all purposes. But I will be wrong. IT has own price, and in most cases it's not so cheap. Present technologies are expensive. On other side growth of IT, and fast aging, makes decision for using IT difficult. To use IT, user must to be trained, which also costs. Maintains makes price higher. Student's preposition to IT support team has proved that in this research.

Taking a brief look into the university curriculum I can summarize that the part of implementation of IT in educational process has proved this conclusion which emphasize that the University must take into consideration the current needs of students and their likes and dislikes. IT must be used, but, the goal must justify the means.

It is of interest for the University to make a long-term investments and development in this area. There is a need of enduring strategy for the use of IT supported with evaluation.

REFERENCES

[1] Caruso J-B, Kvavik R-B. ECAR Study of Students and Information Technology 2005: Convenience, Connection, Control, and Learning. http://www.educause.edu/ers0506

[2] Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, Strong, Verplank. ACM SIGCHI Curricula for Human-Computer Interaction. 1992,1996 ACM SIGCHI. http://old.sigchi.org/cdg/cdg2.html

[3] Kadriu A, Abazi L. ICT at South East European University. Information Technology Interfaces; 2006 June 19-22; Cavtat, Croatia, 28th International Conference on Volume; 2006. p. 271-276.

[4] South East European University (SEEU) official website. http://www.seeu.edu.mk/en/the-university/presentation/infrastructure

Virtual Infrastructures in Education

Hristo G. Valchanov[1], Nadezhda S. Ruskova[2] and Trifon I. Ruskov[3]

***Abstract* – The traditional teaching method requires the students to attend classes. The classes are scheduled at certain time and in certain laboratory or seminar hall. All these conditions bring limitations in the teaching process. It appears not to be flexible and unable to react quickly enough to the rapidly changing areas of study like IT technologies. The usage of virtual machines and virtualization technologies will increase the capacity and usage of the existing infrastructure. Due to, the teaching will become more adequate to the rapidly changing world of IT industry. In this paper we present our approach for implementing a virtual infrastructure in Computer Science Department at Technical University of Varna.**

***Keywords* – Virtual infrastructures, Virtual computers, E-learning.**

I. INTRODUCTION

The traditional way of learning is based on the laboratory facilities used by students for practical utilization of the material. Such laboratory facilities in most cases consist of a number of personal computers, cabling and related network devices located in a specific laboratory. Each computer has a separate operating system and installed software according to the curriculum. In carrying out the laboratory exercises the students need to be physically present in the room at specific times. The quality of the conduct of classes requires the maximum number of students does not exceed the capacity of the laboratory.

All these conditions bring limitations in the teaching process. It appears not to be flexible and unable to react quickly enough to the rapidly changing areas of study like IT technologies. Changing the curriculum, both in theoretical and practical aspect, will bring the need of changing the existing laboratory infrastructure, and the software already installed.

Problems can be summarized in two categories. First, these are problems associated with maintenance of hardware and need to reinstall software and update the various courses. The large number of computers requires a laboratory and staff time to support this. By increasing the number of applications also grow and the need to increase the number of servers on which these applications run. This means that the cost of building and maintaining an IT environment will increase significantly.

On the other hand, it is possible to provide a PC for each student. Students will be able to fully carry out laboratory exercises, which reflects in the quality of their training. The limited number of school laboratories and equipment in them are reducing the opportunities for full deployment of training.

Solve many of these problems can be done by applying a modern technology – virtualization.

In this paper we present our approach of implementing and applying a virtual infrastructure in Computer Science Department at Technical University of Varna.

II. VIRTUAL INFRASTRUCTURES

Virtual computer is a logical representation of a computer using software [1]. By separating the physical hardware from the operating system virtualization provides greater operating flexibility and increased utilization of available hardware. Although virtualization is implemented mainly by software, many modern microprocessors include components specifically designed to increase the effectiveness of virtualization. In traditional physical computer there is an instance of the operating system serving one or more applications - Fig. 1.

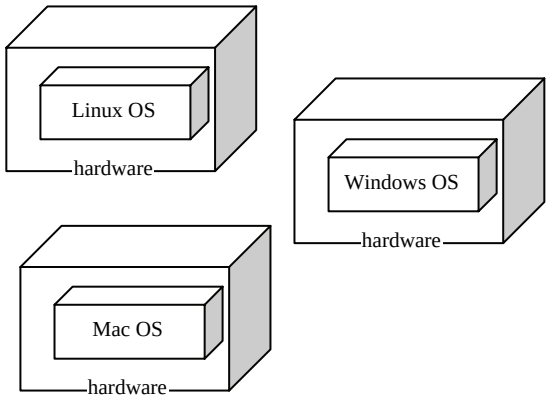


Fig. 1 Traditional computers

In virtual environments one physical computer is running software that represents the physical components of the computer as "abstract" and they can be shared between multiple "virtual computers" – Fig. 2. Each virtual computer can run a separate operating system (Guest OS) independently of other virtual machines within a physical machine. Failure or bugs in one virtual machine do not affect other virtual machines.

[1]Hristo G. Valchanov is with the Computer Science and Engineering Department at Technical University of Varna, Bulgaria, E-mail: hristo@tu-varna.bg

[2]Nadezhda S. Ruskova is with the Computer Science and Engineering Department at Technical University of Varna, Bulgaria, E-mail: ruskova@tu-varna.bg

[2]Trifon I. Ruskov is with the Computer Science and Engineering Department at Technical University of Varna, Bulgaria, E-mail: ruskov@tu-varna.bg

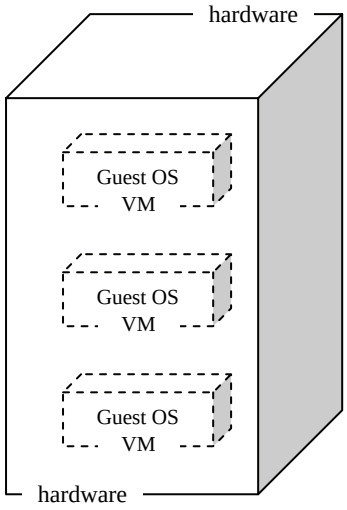


Fig. 2 Virtual server model

Virtualization can be implemented in two directions: desktop and server virtualization.

Desktop virtualization transfers the operating system and it supported applications from student computer to a server. Each user can set his/her desktop environment. But unlike the work of a single computer when moving to another student in another lab, its desktop environment is the same regardless of the currently used workstation. If the used computer is damaged, the student's desktop environment is available at any workstation on the network.

Server virtualization, also known as *virtual infrastructure*, allowing on one physical server to run multiple virtual servers [1],[2]. The result of the virtual infrastructure is a significant reduction in maintenance costs of equipment, balancing the load of computers, flexibility in the management and adding new applications. If the resource requirements of applications running on the virtual server grow, moving the virtual server to another more powerful physical machine is extremely simple form of copying files.

One of the most popular software for building virtual infrastructures is VMware ESX Server 3 [3]. The choice of VMware Infrastructure 3 is dictated by its capacity for multi-processor support, dynamic balancing and allocation of resources between virtual machines and migrating virtual machines between servers without interrupting their work.

III. THE GOALS

The main goal is by virtualization of the infrastructure, combined with mobility of the learners, to increase the efficiency of the teaching process and the research capabilities of students and lecturers. With the achieved flexibility, the teaching will become more adequate to the rapidly changing world of IT industry. The usage of virtual machines and virtualization technologies will increase the capacity and usage of the existing infrastructure, and it will ease its administration.

The virtualization will avoid the need for redesigning and reconfiguring the hardware and software in the laboratories, when changes in the curriculum occur. With the virtual infrastructures it would be possible to use simulation tools to analyze and investigate real systems and processes. To build this, without the mentioned technology, a significant amount of expenses will be needed. Changes in the curriculum will lead to logical configuration of the virtual infrastructure, in form and structure corresponding to the new requirements.

Using the mobility by notebooks, the students will not need to attend the university halls to complete their practical classes. This will give them the possibility to plan their students' activities in their own way and style. Also, as a result, the students would be more interested in the learning process. On the other hand, this system will let the lecturer to have a complete view on the students work, and their achievements.

As a result, this would give the opportunity to implement *Anyplace Anytime Learning* (AAL) in higher education.

IV. IMPLEMENTATION AND RESULTS

A. Implementation

The virtual infrastructure (Fig. 3) is built based on two server machines Sun Fire V20Z, connected with 1G Ethernet network and running Linux operating system. Over this operating system the VMware ESX Server 3 are installed.

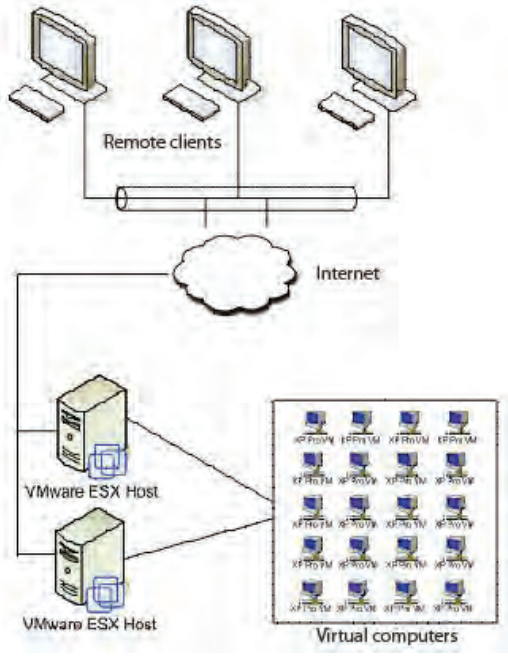


Fig. 3 Virtual infrastructure

B. Parameters

Planning the virtual infrastructure is made based on analysis of requirements for laboratory facilities to conduct laboratory exercises on the different courses such as "Operating Systems", "Compilers", "Distributed Programming" and "Network Infrastructures". Characteristic

of these courses is that classes are conducted in the same laboratory. Different operating systems and application software are used. In Table 1 the data associated with these requirements are shown:

TABLE I
PARAMETERS OF THE LEARNING PROCESS

Course	Number groups	Number students	Work places	OS
Operating Systems	8	15	8	Linux
Compilers	8	15	8	Windows XP
Distributed Programming	8	17	8	Linux
Network Infrastructures	3	12	8	Windows Server 2003

As the table shows the number of desktop computers is insufficient for practical exercises for all students. This is a problem that inevitably leads to incomplete mastering of the educational material.

The second problem is that there are required different operating systems for different courses. When using stationary desktop computers these operating systems could to be installed on separate partitions on hard disks, but experience has shown that this leads to serious difficulties in their reinstallation. This is situation because of frequent failures in the system caused by the work of students with administrative rights, viruses, etc.

C. Measure approach

To measure results we introduce the following approach. For those students, participating in Anyplace Anytime Learning (AAL), a baseline at the beginning of an academic year was defined. Various data and information about the learning process were kept, such as: consumed time for preparation of classes, number of projects assigned to students, number of practical tasks prepared by students, number of practical classes, results and marks achieved by the students.

This information was gathered as feedback, by the lecturers and tutors during the semester. Based on this data, at the end of each academic year a statistical analysis was created.

D. Results

After deployment of the virtual infrastructure of each student is given the opportunity to work on a personal virtual host. Using X-windows terminals (thin clients) allows reducing the cost of laboratory equipment. There are many benefits to use thin clients.

With no hard drive, fan or other moving parts, thin clients have a much longer lifespan than standard computers and use significantly less power. Unlike a traditional desktop or notebook computer, no applications or data are stored locally on the thin client. This makes them easy to replace if lost, stolen or damaged. Unlike a traditional desktop or notebook computer, no applications or data are stored locally on the thin client. This makes them easy to replace if lost, stolen or damaged.

Deployment costs are also reduced as thin clients can be remotely configured and do not need to be set up individually. Significantly the reconfiguration of work on computers is reduced.

In addition we give the opportunity for the students to have remote access to virtual machines.

As a major positive impact from the implementation of virtualization, however, has made increasing the completion of tasks for the realization of students. The ability to remotely access virtual infrastructure resources allows students to work not only in laboratories but also from other places.

On Fig. 4 the evaluated results for the experimentally involved courses are shown.

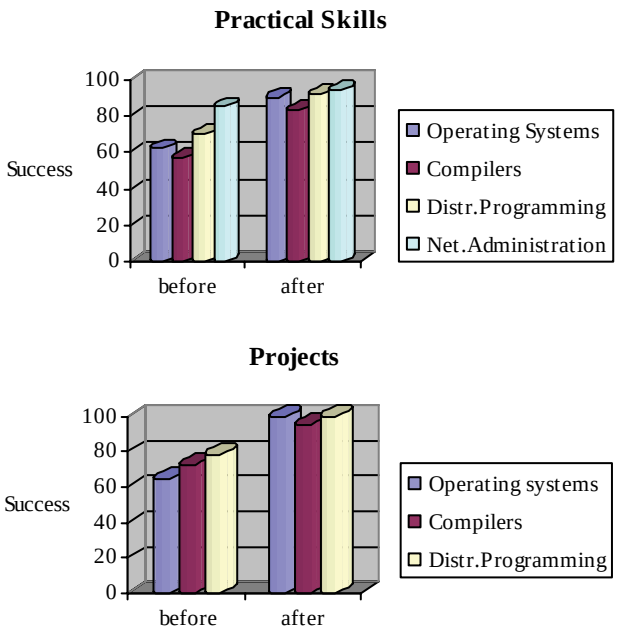


Fig. 4 Summary results

There is represented the percentage of success in terms of "practice skills" and "coursework". The courses have different improvement, which is a consequence of their specific matter. The improvement is more significant in courses mainly related to programming such as “Compilers” and “Distributed Programming”.

As a result, the utilization of laboratory exercises increased to 95% on average. In comparison, before applying the new approach, success due to the delay of the students work was on average about 65%.

V. CONCLUSION

In this paper the virtual infrastructure for student education is given. The main goal of our work is to combine the possibilities of the mobile communications and virtual infrastructure in order to improve the teaching quality of the

higher education. The integration of the virtualization and the mobility will provide unified methodology for teaching different subjects in the field of the computer science [4].

Using the mobile computers will make the classes flexible. The classes will depend less from specific equipment, location, and time schedule. Beside the mobility of learning, other benefits for the students from this new way of teaching will be: ease in exchange of ideas and materials among students and lecturers, quick access to teaching materials, effective usage of the network infrastructure. This approach will also promote the research activity to the students.

Objective of future work is the building of virtual laboratories. Virtual laboratories allow build a separate computer environments for each student. This enables students to access applications that can not install or have on their PCs. Virtual laboratories allow access to laboratory facilities at any time and from any place, something that would enable all students to perform practical sessions remotely. With the help of Web interface, the students will be able to access the real equipment in the laboratories. Also, the simulation software will let them make their research, without attending the classrooms. This will give the opportunity to achieve better usage of the existing university network.

By thus achieved mobility training is expected to raise interest of students to learning and to research. The project will improve the university teaching methodology, will bring new learning techniques and will enrich the experience of both students and lecturers.

ACKNOWLEDGEMENT

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

REFERENCES

[1] I. Rozenfeld, *Server Virtualization,* NW City Technology Forum, 2007.

[2] D. Marshall, W. Reynolds. *Advanced Server Virtualization,* Taylor & Francis Group, LLC, 2006.

[3] VMware Virtual Infrastructure 3. http://www.vmware.com/technology/virtual-infrastructure.html

[4] HP Innovations in Education Worldwide Summit, San Francisco Bay Area,CA, 2010 http://www.istevision.org/postergallery.php

An Approach to Factorization and Attack Against the Asymmetric Cryptographic Algorithm RSA

Petar Ts. Antonov[1] and Valentina R. Antonova[2]

***Abstract* – Problems of the security of the most popular asymmetric cryptographic algorithm RSA are considered. A new approach to factorization of large numbers and attack against RSA is proposed. A recommendation about secure generation of public and secret keys of RSA is formulated.**

***Keywords* – RSA, factorization, crypto-attack.**

I. INTRODUCTION

RSA (Rivest-Shamir-Adleman) is the most often used asymmetric cryptographic algorithm in security schemes of computing and communication systems. It is one of the few known cryptographic algorithms of this class that are distinguished for their universal application – in ciphering of messages, digital signatures and exchange of session secret keys in hybrid ciphering schemes. It has been perceived by ISO, ITU-T, ANSI, banking and financial spheres, military units, etc. It is used in crypto-systems to protect e-mail, in electronic trade, credit card systems, Internet-browsers, etc. Actually RSA is de facto world–wide standard for asymmetric cryptography, well known also as cryptography with public keys.

The description of RSA is very simple [1, 3, 4, etc.]. Initially every user chooses two sufficiently large prime numbers a and b, whereupon is calculated

$$n = a.b \quad \text{and} \tag{1}$$

$$\Phi(n) = (a-1).(b-1) \tag{2}$$

The public key Kp and the secret key Ks of the pair "public/secret" keys of that user are determined from the correlations:

$$\text{НОД}(K_p, \Phi(n)) = 1 \quad \text{and} \tag{3}$$

$$(K_s.K_p) \bmod \Phi(n) = 1, \tag{4}$$

where НОД is the greatest common divisor of the numbers in square brackets [the number Kp must be accidentally chosen and mutually simple to the number Φ(n)].

Further on in the procedures of ciphering and deciphering respectively the pairs of numbers (Kp,n) and (Ks,n) are used, as the first pair is made public attribute, Ks is kept in secret, and the initial numbers a and b are destroyed.

Any accidental source can address a protected message to that user, ciphering the open text M of the message with the public key K_P and send the resultant ciphered text E. Only that user is able to restore M after deciphering E with his own secret key K_S.

Any eventual offender, received somehow the ciphered text E, will know K_P and n. To restore the open text M however, this offender will have to factorize n to its unknown prime factors a and b (this procedure is known as factorization of n), whereupon to calculate Φ(n) and K_S. But this factorization is very labour-consuming and with sufficiently large values of a and b becomes practically beyond the potentialities of contemporary level of technologies.

It is possible that the offender makes a try to find out Φ(n) directly, without factorization of n, but this is not simpler than the factorization itself.

The third possibility for the offender is to immediately calculate K_S without factorization of n and determination of Φ(n), but with sufficiently large K_S, this possibility is not easier than the factorization itself either.

In essence, the above three crypto-attacks to RSA can be considered as attack of the brute force. In the end, however, the real practical attack to RSA comes to solving the problem with the factorization of n.

II. AN APPROACH TO FACTORIZATION

A number of methods to solve this problem are known (see [2, 3, 4, etc.]), which differ in various labour-consumption. A new method of factorization is presented below, which in a number of cases requires less steps, compared to the other similar methods known. The idea of this method is based on the popular Theorem of Euler, saying that any even number, greater than 2, can be represented as a sum of two prime numbers.

If a successful factorization of n is made and the prime factors **a** and **b** are found, then further on Φ(n) will be calculated and after that the secret key K_S from the correlation:

$$(K_s.K_p) \bmod \Phi(n) = 1. \tag{5}$$

To expose the main point of the proposed method for factorization, we shall use the following example:

[1]Peter Ts. Antonov is with the Department of Computer Science and Engineering, Technical University of Varna, 1, Studentska Str., 9010 Varna, Bulgaria, E-mail: peter.antonov@ieee.org

[2]Valentina R. Antonova is with the Department of Computer Science and Engineering, Technical University of Varna, 1, Studentska Str., 9010 Varna, Bulgaria,E-mail:valyvarna@yahoo.com

a=13, b=29, n=377, $\lfloor n^{0.5} \rfloor$=19.

Further on we shall write down the two-row sequence of the numbers (1÷$\lfloor n^{0.5} \rfloor$) and ($\lfloor n^{0.5} \rfloor$÷37) so that the sum of the numbers situated one under another is equal to 2.$\lfloor n^{0.5} \rfloor$=2.19=38:

1 2 3 4 5 6 7 8 9 10 11 12 **13** 14 15 16 17 18 19
37 36 35 34 33 32 31 30 **29** 28 27 26 25 24 23 22 21 20 19.

As a<b, then in the first row of the sequence above, the number a=13 is situated closer to 19=$\lfloor n^{0.5} \rfloor$, than b=29 in the second row. The distance from a to $\lfloor n^{0.5} \rfloor$ in this case is La=$\lfloor n^{0.5} \rfloor$-a=19-13=6, and from b to $\lfloor n^{0.5} \rfloor$ - Lb=b-$\lfloor n^{0.5} \rfloor$=29-19=10. If we increase $\lfloor n^{0.5} \rfloor$=19 consecutively by one and two, then the new similar two-row sequences of numbers will look like the ones below:

1 2 3 . . . . . 11 12 **13** 14 . . . . . 19 20
39 38 37 . . . . . . **29** 28 27 26 . . . . . 21 20

1 2 3 . . . . . 11 12 **13** 14 . . . . . 19 20 21
41 40 39 . . . . . 31 30 **29** 28 . . . . . 23 22 21

In the second sequence the numbers a=13 and b=29 are already situated one under another and their sum is equal to 42=2.21, and their product - to n=377. To obtain this correspondence in this case only two iterations came out to be enough.

If in the general case, the necessary number of iterations for factorization of n is denoted by V, it is easy to prove the validity of the following correlation:

$$V = \frac{1}{2}\left(L_a - L_b\right) = \frac{a + b - 2\lfloor \sqrt{n} \rfloor}{2}. \quad (6)$$

It can be seen that the prime factors a and b of the number n we have been after, can be determined from the correlations:

$$a = \lfloor \sqrt{n} \rfloor + V - x \quad (7)$$

$$b = \lfloor \sqrt{n} \rfloor + V + x, \quad (8)$$

which lead to

$$a + b = 2\left(V + \lfloor \sqrt{n} \rfloor\right). \quad (9)$$

In the example, considered above, V=2 and x=8, while a=19+2-8=13 and b=19+2+8=29. Essentially x is the distance between a and b, and the right-hand beginning of the two-row sequence of numbers at the last iteration, which in this case is the second in order. After the second iteration, the numbers a and b proved to be one under another and at the same distance from the beginning of this sequence, which in this example is the number ($\lfloor n^{0.5} \rfloor$+V)=21.

As a.b=n, then the second factor b can be expressed simply by the first factor a, i. e. b=n/a and then:

$$a + n/a = 2\left(V + \lfloor \sqrt{n} \rfloor\right) \quad (10)$$

$$a^2 - 2\left(V + \lfloor \sqrt{n} \rfloor\right)a + n = 0. \quad (11)$$

Solving the quadratic equation above, finally we get:

$$a = V + \lfloor \sqrt{n} \rfloor - \sqrt{\left(V + \lfloor \sqrt{n} \rfloor\right)^2 - n}, \quad (12)$$

$$b = V + \lfloor \sqrt{n} \rfloor + \sqrt{\left(V + \lfloor \sqrt{n} \rfloor\right)^2 - n} \quad (13)$$

These two correlations can be used for factorizing n. For that purpose V is given consecutive integer values, starting with 1. For each consecutive value of V the expression under the radical is calculated

$$\left(V + \lfloor \sqrt{n} \rfloor\right)^2 - n \quad (14)$$

and is checked if it is equal to the square of an integer. If it is not, V is given the next consecutive value and so on. The solution for a and b is reached at that V, for which the expression

$$\sqrt{\left(V + \lfloor \sqrt{n} \rfloor\right)^2 - n} \quad (15)$$

accepts integer value, denoted in this case by x. Then from the expressions above for a and b we calculate their specific needed values.

If in the expression for V we replace b by α.a, i.e. b=αa, then

$$V = \frac{a + \alpha a - 2\lfloor \sqrt{\alpha a^2} \rfloor}{2} = \left\lfloor a\frac{1 + \alpha - 2\sqrt{\alpha}}{2} + 1 \right\rfloor. \quad (16)$$

For approximate calculation of the necessary number of iterations the formula

$$V^* = a\frac{1 + \alpha - 2\sqrt{\alpha}}{2}, \quad (17)$$

can be used, which for the sake of convenience can be presented in the form:

$$\left(\frac{V^*}{a}\right) = 0.5\left(1 + \alpha - 2\sqrt{\alpha}\right) \quad (18)$$

In this formula, for the purpose of simplification of the expression, the requirement for separation the integer part of the result of the square root calculation of αa^2 is removed.

The graph of the changes in (V*/a) depending on α is shown below (fig.1).

If the prime factors a and b have the same decimal rank, i.e. the same number of decimal classes, then

$$a_{\min} = 1000..... \quad (19)$$

$$b_{\max} = 9999..... \text{ and} \quad (20)$$

$$\alpha_{\max} = 9{,}999..... \approx 10\,. \quad (21)$$

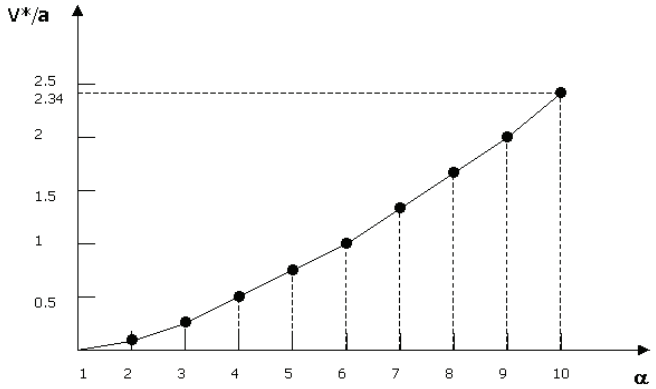


Fig. 1.

In this most unfavourable case of factorization of n, the number of iterations $V^*=2.34a=2.34\cdot 10^h$, where h is the decimal rank of a and b.

III. Conclusion

Considering what has been said above, we can draw the following conclusion: the exposed method for factorization requires smaller number of iterations at smaller difference between the values of a and b. Furthermore, the method is applicable even when a and b are not prime numbers, but odd or even numbers.

The comparative estimations made for the necessary number of iterations between the method presented and the methods for factorization, known from literature, show that in some cases it is with less labour consumption, and in other cases – not. For example, for factorization of n=22317 using the method proposed, only two iterations will be necessary. For comparison, according to the estimations given in [4] for the necessary number of iterations with different methods, for the algorithm of Dixon in this case will be needed:

$$e^{\sqrt{\log n.\log n(\log n)}} \cong 6 \text{ iterations,} \quad (22)$$

And for the Quadratic Sieve Algorithm:

$$e^{\sqrt{\ln n.\ln(\ln n)}} \cong 122 \text{ iterations.} \quad (23)$$

In conclusion we shall annotate that the method proposed can be considered as development of the problem for factorization that has more general nature. Using this method, a quicker crypto-attack can be accomplished to the cryptographic algorithm RSA, provided the numbers a and b have close values.

References

[1] Антонов, П. Ц., С. В. Малчев, *Криптография в компютърните комуникации*, Варна, 2000.

[2] Кнут, Д. *Искусство программирования для ЭВМ*. Т. 2. Получисленные алгоритмы, Пер. с англ. – Москва, Мир, 1977.

[3] Schneier, B. *Applied Cryptography. Protocols, Algorithms and Source Code in C*, John Wiley & Sons Inc., 1996.

[4] Goldwasser, S., M. Belare. *Lecture Notes on Cryptography. – Cambridge*, Massachusetts Institute of Technology, June 1997.

This page intentionally left blank.

Secure Authentication Protocol for Distant Access

Petar Ts. Antonov[1] and Valentina R. Antonova[2]

***Abstract* – In this report we propose a secure dialogue authentication protocol with passwords, under which password images can be preserved on the server, and every distant access is carried out by different key-word, transmitted through the channels in type.**

***Keywords* – Authentication protocol, cryptographic algorithms with public key.**

I. INTRODUCTION

As well known, the measures for authentication aim at provision of sufficiently secure procedures to establish the legality of the access to the resources. The users are legalized by introducing some authentic information, intended for the protection scheme of the respective resource. In the capacity of authentication information passwords are used most often. At the same time, the majority of attempts for illegal access to the resources are aimed at theft and destruction of precisely these passwords [1, etc.]. This imposes serious approach to the choice of the passwords themselves as well as to the development of specific protocols for authentication.

II. SECURE AUTHENTICATION PROTOCOL

For determination we consider information – communication system, including server and distant clients, using for connection some kind of WAN-technology. The server must operate the orders of many clients that have legal registration and reject the attempts for illegal access.

To the server S is assigned a pair of public K_{PS} and secret K_{SS} keys, in accordance with some of the known and secure asymmetric cryptographic algorithms (cryptographic algorithms with public key). This applies to every legal client X, who must also have available his own public K_{PX} and secret K_{SX} keys.

The public keys of the registered clients are preserved on the server in the first row of a two-row authentication matrix, whose second row is empty at fist (the file with these keys is not necessary to be protected as the keys themselves are accessible).

Using the mentioned above, the procedure of the considered protocol for authentication can be described in the following way [2]:

1. Client X establishes communication with the server S and originally legitimizes himself by his own public key K_{PX}.
2. The server S finds the field with K_{PX} in the first row of its authentication matrix and in the corresponding field in the second row of the same column of the matrix writes down generated by it accidental binary word T_X, that is to be used as password and is unique for each new séance for authentication of client X.
3. The server ciphers T_X with the public key K_{PX} of X, i.e. $E = F_E (T_X, K_{PX})$ and sends the cipher message E to X.
4. The client X deciphers the received cipher message E with his own secret key and restores T_X from the correlation $F_D(E, K_{SX}) = T_X$.
5. Further X ciphers T_X with the known public key K_{PS} of the server and sends to it the cipher message $E^* = F_E (T_X, K_{PS})$, together with the key K_{PX}.
6. The server deciphers E* with its own secret key, i.e. $F_D(E^*, K_{SS}) = T_X$, whereupon it compares the calculated value of T_X with the one, initially written down in step 2 and at coincidence it deletes T_X from the matrix and allows the access, and at lack of coincidence – deletes T_X and breaks off the connection.

The clients and the server can use only one asymmetric cryptographic algorithm, but it is possible that separate clients work with different algorithms of that kind. In the last case, in the authentication matrix of the server should be kept also information about the type of the asymmetric algorithm of every registered client, and on the server itself should be juxtaposed corresponding pair of public and secret keys for every algorithm used. At his original registration the client should choose one of the asymmetric algorithms that the server offers, whereupon he should generate his own pair of keys and submit his public key to the server.

In this case, the generated by the server accidental words T_X are preserved only for the time of the procedure for authentication. As this time is very short, then for T_X the danger of being revealed and meanwhile used for illegal access by an eventual offender is very little, too. But for still greater security, the words T_X can be preserved in ciphered type with the help of some known symmetric cryptographic algorithm, like DES, for example. In this situation, it is not necessary for the length of T_X to be greater than 64 bits, as ciphering and deciphering with DES is fulfilled upon binary blocks of that size. But because of the short time of existence of the words T_X, the ciphering can be done with weaker, but quicker algorithm than DES, whereupon the length of T_X will be decreased. This will lead to increase in productivity without considerable offence in security.

The main advantages of the authentication protocol, stated above, come to the following:

- No secret information about the client is preserved constantly on the server, but only their popular public keys.

[1]Peter Ts. Antonov is with the Department of Computer Science and Engineering, Technical University of Varna, 1, Studentska Str., 9010 Varna, Bulgaria, E-mail: peter.antonov@ieee.org

[2]Valentina R. Antonova is with the Department of Computer Science and Engineering, Technical University of Varna, 1, Studentska Str., 9010 Varna, Bulgaria,E-mail:valyvarna@yahoo.com

- New clients are registered simply, only by their public keys.
- The procedure is dialogue and at every new attempt for access different accidental word T_X is exchanged, which means that the offender can not use a heard and recorded legal current séance for establishing a connection and getting afterwards illegal access.
- During the time of dialogue through the communication channels only the open public key of the client is transmitted and the generated by the server accidental word T_X, which is in ciphered by the server or the client type, can not be deciphered by eventual offender.
- The accidental words T_X, used for passwords, are preserved by the server for a very short interval of time, which brings to minimum the danger of being caught.
- There could be additionally provided in the protocol periodic control of the access during legally established séances of connection and at every new control will be generated a new value of T_X.
- Modification is possible, introducing interruption of initially established connection and second invitation by the client or reverse invitation by the user.
- In essence the protocol represents a complete authentication scheme as it includes also the method for generation and preservation of passwords themselves.

Bellow we propose and give convincing argumentation on an effective variety of the authentication protocol considered, where the use of cryptographic algorithms with public keys is avoided, and the operations ciphering and deciphering are replaced by operations in modular arithmetic.

To the server is juxtaposed a sufficiently large prime number Q and an accidentally chosen number K in the interval [0, Q-1]. The number Q is known to all of the clients, and K is going to be used as a secret key of the server. A client X chooses two accidental numbers A and B in the same interval [0, Q-1]. The number A will perform the role of identifier of the client X, and B – his password for access.

The client calculates $H=A^B \text{mod} Q$ and registers himself in the server with the pair A and H. The server maintains a 3-row authentication matrix with number of columns equal to the number of clients registered. In the first field of the column for client X stays the number A, in the second field - $D=H^K \text{mod} Q$, and the third field remains empty.

At every attempt of client X to obtain access to resources of the server the following successive steps of the protocol will be realized:

1. X establishes communication with the server and inputs his own identification number A.
2. The server generates an accidental number T in the interval [0, Q-1], writes down this number in the empty field of the column for X in the authentication matrix and calculates $C=(A^K \text{mod} Q)^T \text{mod} Q$, whereupon sends C to the client.
3. The client calculates $F=C^B \text{mod} Q$ and returns to the server the pair A and F.
4. On receiving A and F, the server calculates $F^*=D^T \text{mod} Q$, compares F and F* and if F=F*, permits the access.

The veracity of final assertion of step 4 is based on the fact, that:

$$F = \left(\left(A^K \bmod Q\right)^T \bmod Q\right)^B \bmod Q \quad \text{and} \tag{1}$$

$$F^* = \left(\left(A^B \bmod Q\right)^K \bmod Q\right)^T \bmod Q, \tag{2}$$

so the equation F=F* can be proved easily using the known in modular arithmetic correlation

$$XY \bmod Z = \left(X \bmod Z . Y \bmod Z\right) \bmod Z . \tag{3}$$

The scheme of realization of the presented authentication protocol is shown visually in fig. 1, where the separate consecutive steps, fulfilled by the server and the client, are marked.

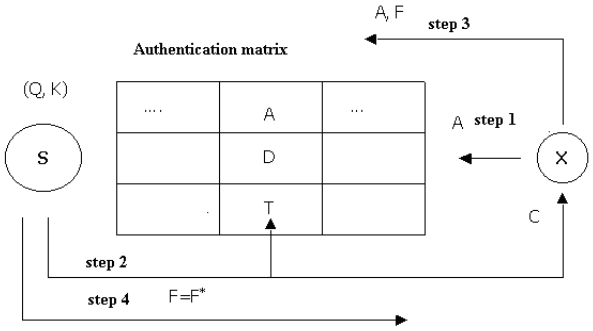


Fig. 1.

III. Conclusion

It can be seen that with this procedure in the authentication matrix of the server are not preserved any passwords, but their images, calculated on the basis of clients' identifiers, the actual passwords and the secret key of the server. Besides, during the dialogue different numbers are exchanged, depending on the numbers T, which are accidentally generated for every new attempt for access. In this way the security of the authentication protocol is guaranteed against evil-minded access to the authentication matrix as well as against eventual hearing by an offender in the communication channel.

References

[1] Кландер, Л. *Защита от хакери и най-добрите хакерски трикове и техники*, Прев. от англ. – София, СофтПрес-ООД, 1999.

[2] Антонов, П. Ц., В. Р. Антонова. "Протокол за защитено използване на пароли", ТЕЛЕКОМ'1999, Сб. докласди, Варна, България, 1999.

[3] Антонов, П. Ц., С. В. Малчев, *Криптография в компютърните комуникации*, Варна, 2000.

[4] Schneier, B. *Applied Cryptography. Protocols, Algorithms and Source Code in C*, John Wiley & Sons Inc., 1996.

Java-Based Systems for Analysis and Estimation, with Application in Diagnostics of Complex Objects

Hristo Nenov

***Abstract* – In this paper is described full process of design, realisation and testing of system for analysis and estimation. Like a tool for realisation of the system are used the possibilities of Java technologies, and advantages granted by it. Based on powerful mathematic apparatus, using supervised learning this system is very strong tool for observing of complex objects.**

***Keywords* – System, Analysis, Estimation, Java, Object, Pattern recognition.**

I. INTRODUCTION

The concept – complex object or complex system is connected not only with their physical characteristic but with complexity of their condition recognition. Normally this is a fuzzy multitude of data and crossing classes of conditions. The systems used for solving of this type of tasks have different realizations, but common principal of using. The system described here has variants of realizations and different ways of using.

II. SYSTEM FOR ANALYSIS AND ESTIMATION

A. General Principals

Pattern in pattern recognition theory, we call the multitude of effects, united in common characteristic. Image is – every concrete effect which must be definite to one of the patterns, based on his characteristics. Follow this, it can be said, that pattern is formatted group like etalon, image is hypothetic member of a priory definite patterns [1].

After deep analysis over probabilistic and geometrical algorithms for pattern recognition, the conclusions are, that the main calculating procedure for pattern recognition described with normal distribution, using geometrical and probabilistic algorithms are common:

- calculating of vectors of mathematical meaning and co-variation matrix;
- transpose, multiplication and other manipulation over matrix;
- searching of own values and singular values;
- defining of border values, depending of chosen strategy.[1]

Like a common form for divide function can be accepted equation 1.1. First two members express square mahalanobis distance and the free member is the specific of strategy. (maximum of a posterior probability, maximum likelihood estimation MLE)(1).

$$g_i(X) = -\frac{1}{2} X^T \sum_i^{-1} X + (\sum_i^{-1} \mu_i)^T X - \frac{1}{2} \mu_i^T \sum_i^{-1} \mu_i - \\ -\frac{1}{2} \ln\left|\sum_i\right| + \ln P(w_i) + \ln|c_i| = X^T A_i X + a_i^T X + a_{i0} = \max \quad (1)$$

With m classes of condition we searching maximum of dividing function for different classes (2):

$$\bar{y}_i \in X_1, \_if\_g_j(\bar{y}) = -\frac{1}{2} y^T \sum_j y + (\sum_j^{-1} \mu_j)^T - \\ -\frac{1}{2} \mu_j^T \sum_j^{-1} \mu_j - \frac{1}{2} \ln|\sum_j \mu_j| + \ln P(x_j) + \ln|c_j| = \quad (2) \\ = Y^T A_j Y + a_j^T y + a_{j0} \Rightarrow \max$$

For visualization of classes is used principal component analyze (PCA).

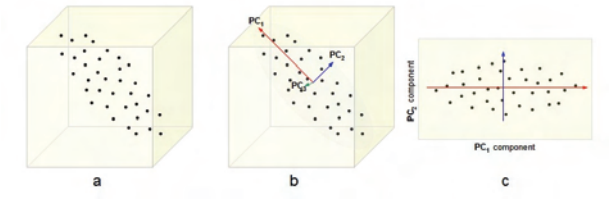


Fig.1.PCA

B. Program Realisation

Model–View–Controller (**MVC**) is a software architecture, currently considered as an architectural pattern used in software engineering. The pattern isolates "domain logic" (the application logic for the user) from input and presentation (GUI), permitting independent development, testing and maintenance of each[2].

The **model** is the domain-specific representation of the data upon which the application operates. Domain logic adds meaning to raw data (for example, calculating whether today is the user's birthday, or the totals, taxes, and shipping charges for shopping cart items). When a model changes its state, it notifies its associated views so they can be refreshed.

Many applications use a persistent storage mechanism such as a database to store data. MVC does not specifically mention the data access layer because it is understood to be underneath or encapsulated by the model. Models are not data access objects; however, in very simple apps that have little domain logic there is no real distinction to be made. Also, the ActiveRecord is an accepted design pattern which merges domain logic and data access code - a model which knows how to persist itself[2].

The **view** renders the model into a form suitable for interaction, typically a user interface element. Multiple views can exist for a single model for different purposes.

The **controller** receives input and initiates a response by making calls on model objects.

An MVC application may be a collection of model/view/controller triplets, each responsible for a different UI element[2].

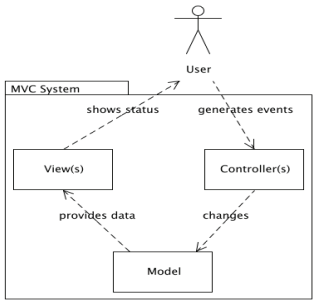


Fig.2. MVC

Model-driven architecture is focused on forward engineering, i.e. producing code from abstract, human-elaborated modelling diagrams. One of the main aims of the MDA is to separate design from architecture. As the concepts and technologies used to realize designs and the concepts and technologies used to realize architectures have changed at their own pace, decoupling them allows system developers to choose from the best and most fitting in both domains. The design addresses the functional (use case) requirements while architecture provides the infrastructure through which non-functional requirements like scalability, reliability and performance are realized. MDA envisages that the platform independent model (PIM), which represents a conceptual design realizing the functional requirements, will survive changes in realization technologies and software architectures.

Of particular importance to model-driven architecture is the notion of model transformation.

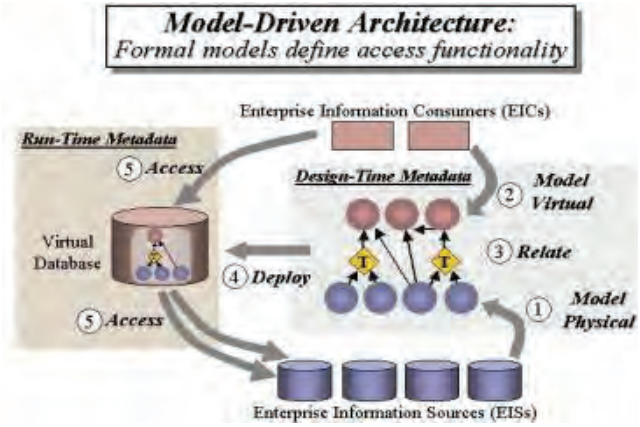


Fig.2.MDA

There are three variants of the system:

1.Desktop application based on PC

2.Client- server approach with connection with enterprise database.

3.Client- server approach, client part for mobile devises.

The cross platform (operating system independences) character of Java granted possibility of variants for realization and using of system.

III. System testing, with diagnostic of main ship engine MAK, from series M32C and M43C

The data for analysis is based on ship engine log-book. A priority are chosen twelve parameters which are marked like most informative (Z, RPM, Pz, Pc, λ, Ps, ε, Tbg°C, Ts, $\Delta P_{T.K.}$, ΔT_{cooler}^{air}, Cu)[1]. Thirteen classes are defined:

- Good working engine – class 1
- Heavy threadbare piston ring – class 2
- Dirty air filter – class 3
- Precipitation over turbine circle- class 4
- Threadbare fuel injector – class 5
- Threadbare nozzle – class 6
- Early fuel injection – class 7
- Late fuel injection – class 8
- Broken piston ring – class 9
- Air between piston and cylinder bush – class 10
- Precipitation over valves – class 11
- Heavy dirty air filter – class 12
- Dirty air cooler – class 13

The data is separated of two parts – for learning of the system and for recognition.

Results from recognition are shown in Table 1 and Fig.3.

TABLE I - CLASS RECOGNITION

	Real Classes													
Recogn. like:	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Class 12	Class 13	accuracy
Class 1	183	0	2	0	0	0	8	0	21	0	0	0	0	85.51%
Class 2	0	163	0	0	0	0	0	0	0	16	0	0	0	91.06%
Class 3	0	0	127	0	0	0	0	0	0	0	0	0	0	100.00%
Class 4	0	0	0	151	0	0	0	6	0	0	0	0	0	96.18%
Class 5	0	0	0	0	106	3	0	0	0	0	0	0	0	97.25%
Class 6	0	0	0	0	15	139	0	0	0	0	0	0	0	90.26%
Class 7	0	0	0	0	0	0	95	0	0	0	0	0	0	100.00%
Class 8	0	0	0	0	0	0	0	103	0	0	0	0	0	100.00%
Class 9	0	0	0	0	0	0	0	0	101	0	0	0	0	100.00%
Class 10	0	15	0	0	0	0	0	0	0	158	0	0	0	91.33%
Class 11	0	0	0	0	0	0	0	0	0	0	115	0	0	100.00%
Class 12	0	0	0	0	0	0	0	0	0	0	0	122	0	100.00%
Class 13	0	0	0	0	0	0	0	0	0	0	0	0	112	100.00%
Recogn.	100.00%	91.57%	98.45%	100.00%	87.60%	97.89%	92.23%	94.50%	82.79%	90.80%	100.00%	100.00%	100.00%	

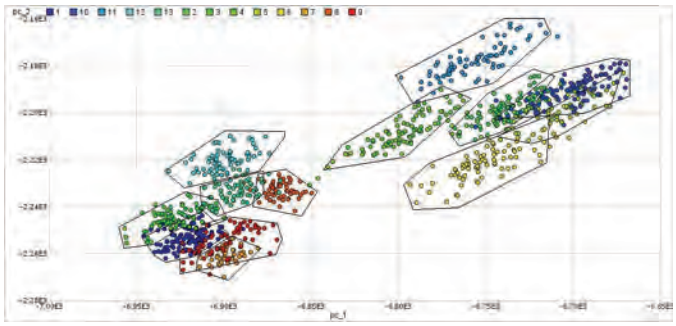


Fig.3. Principal-component analyze of condition classes

IV. Conclusion

The system is fully operational and the results from its work for this hard task are very accurate. This makes it, very powerful tool for diagnostic of complex objects.

Acknowledgement

The work presented on this paper was supported within the project BG 051PO001-3.3.04/13 of the HR Development OP of the European Social Fund 2007-20013.

References

[1] Nenov, Hr. “Algorithms and systems for estimation of condition and managing of complex objects” PhD thesis 2008.

[2] www.sun.com

Innovation in Systems for Analyze and Estimation

Hristo Nenov[1] and Geo Kunev[2]

***Abstract* – In this paper are shown different principals of systems for analysis and estimation. Also is made a preview of the evolution of paradigms in programming and different classificatory and approaches for classification.**

Keywords – Pattern recognition, Clasificatory, DataMining, kNN, KDD, MDA.

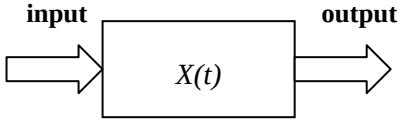
I. INTRODUCTION

The solving of every engineer task is preceded by presence of adequate model. In real human practice this type of model often misses. Main goal of the human (abstract) thinking is to search about regularities, which will help to understand mechanism of some phenomenon. Goal of the scientific searching is to be find orders, describing not only simple objects and processes, but classes of objects and processes. Once found, the model simplified the variations of orders, formulas or structures (images, schemas) and granted possibilities for explaining and/or forecasting.

II. MODELLING OF SYSTEMS

A. System approach

In System approach every object (process) is "System".

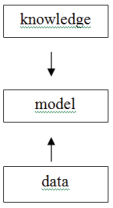


Technical systems like general are managed systems (relatively closed physics systems) and can be described with terms input – output. (Fig.1)

Fig.1.Managed system

Two main methods for system modeling are (Fig.2.):

- knowledge and understanding about system ("white box")
- experimental data for input and output ("black box")



Top	Searching:	Knowledge based	Data based
deduction	synonyms:	top-down	bottom-up
	approach:	deduction	induction
induction	describe:	Internal structure	System behavior
Bottom	method:	analyze	synthesis

Fig.1.Two main approaches for system modeling

B. Complex Systems

When we making preview /modeling of complex real systems, which normally are hierarchically, must note the fact that this kind of systems are parts (subsystems) of other systems and etc. (Fig.2).

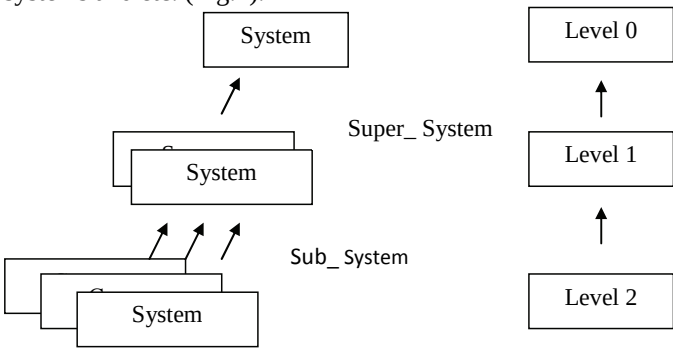


Fig.2. Complex Hierarchical System

This fact on the abstract level is the base of the object approach. It has been realized like class (object) hierarchy.

Searching of knowledge from bottom to top, based on data (induction) is most powerful tool for modeling in case of missing a prior knowledge.

"Black box" approach is inductive and including:

- collecting of (input-output) data from system
- synthesis of models
- application of modeling for describing/prediction

"White box" approach is deductive and including:

- knowing the mechanisms and behavior of the system
- possibility for deductive conclusions, based on knowledge
- base for expert systems

B. Modeling and coding

In computer programming was changed few paradigms.(Fig.3)

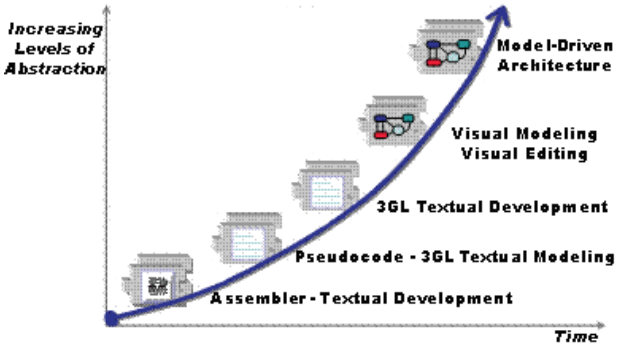


Fig.3.Paradigms and level of abstractions

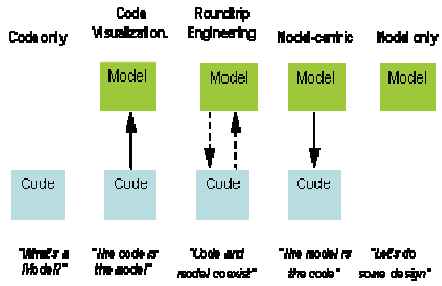


Fig.4. Modeling and Coding

Ratio coding – modeling is improving nonstop, together with evolutions of program paradigms (Fig.4.)[1].

- Current object oriented paradigm is connected with the idea that programming must be oriented not to procedures (describing computer actions) but to declaration languages (describing key abstracts of problem). Nevertheless higher level of abstraction (Fig.3.), its mass using and supporting with different programming environments is focused on the coding.[1]
- Future, model oriented paradigm supposes presence of tools for advanced describing modeling. (XML, UML), and approaches for automotive coding like MDA (Model Driven Architecture), but require considerable change in attitude.

C. KDD and Data Mining

KDD (Knowledge Discovery in Databases) is one relatively new approach for data analysis, which by definition must be *nontrivial extracting of non implicit and a priory unknown data, potentially useful, relatively simple non predefined knowledge from databases*. Knowledge extracting (Data Mining) is one of the steps in KDD technology, but they overlap each other often [1].

Data Mining methods are borrowed from statistic and including: classification, regression, association, clustering. Concept for managed (supervised) learning is borrowed from machine learning (ML), and artificial intelligence (AI).

KDD is used in finances (fraud searching, market prediction, credits), market (user "basket" analyze, target markets, sell predictions),quality control, medicine, astronomy microbiology"text mining", "web mining", "image mining", and etc. [1].

D. Unification of data in data warehouse

One "warehouse" for data (data warehouse) unifies data of different sources and integrates them in significant way. Operational data can be stored in different databases with different formats: relational, transactional, object-oriented, object-relational, active, temporary, text, multimedia, heterogenic and etc. To be unified such kind of data is not very trivial task.

Data warehouse granted data for three levels of data analyze:

- report generation
- OLAP (On Line Analytical Processing),
- KDD

Report generation is one of the functions of working databases (OLTP) and is programmed by SQL clauses, including appropriate functions.

OLAP is advanced techniques for analysis, by which the data is transformed in multidimensional arrays(hyper-cubs).

KDD is the most powerful technology, based on inductive approach, which granted model and knowledge searching in databases.

C. Trained systems and KDD

Approach which is used in KDD, for model searching in data is a case of inductive searching (learning).

There are many classification methods:

- Statistical pattern recognition:
  - Bayes rule
  - Discriminant analyze
  - KNN
  - Probability disperse estimation
  - Tree clasificators
- Neuron networks ("network approach")
- ML - methods
  - Concept training and rules
  - Case – based training
  - Genetic algorithms

E.Classificatory

Concept training.

The searched meaning (concept) c, is a Boolean function

$c: X \rightarrow \{0,1\}$ (1)

where X is multitude of units, which defined the meaning.

Training extract D is consisted from one copy x of X, and his target value $c(x)$ се състои от един екземпляр x от X, заедно с неговата целева стойност $c(x)$. Positive examples for copy are these for which $c(x) = 1$, negative $c(x) = 0$. Current D and examples of c, estimated c or searching hypothesis h in hypothesis space H like:

$(\forall x \notin X)(h(x)=c(x))$ (2)

Linear classificatory

Linear classificatory allowed, that one class can be representing with linear combination of attribute values.

In linear discriminant analysis is allowed that all $f_j(x)$ are multivariate normal density in common covariance matrix and different average vectors.

Relational training models and rules

They are based on logic from first row. The logic expression can work with digit and nondigit data. Models presented by logic languages are readable and understandable.

The following rules are possible:

- Condition rules :

if A then B

- Associative rules:

when A then B

Nonlinear classificatory

This classificatory adjust linear and nonlinear combinations of base functions to combinations of input variables. Examples for this are neuron networks, adaptive splines, nonlinear discriminant analysis and etc.

Genetic algorithms are based on evolutionally model in which survive only most fit members.

Example based methods.

They are based on parameter similarity. Most known method is k^{th} nearest neighbor (kNN).

Validation of classificatory

Based on training and test examples

Definition:

Space of unit X are define like including all metrics vectors of

$x=(x_1 \ldots x_n)$,

where x_i is metric variable(property,attribute).

Frequency and price of errors.

Price for wrong classification of class j is:

$$cost(j)=\Sigma_{i=1,n}\, C(i|j)E(i|j) \qquad (3)$$

Where $E(i|j)$ is the count of classifications, that the class *i is classified like class j.*

Training – test paradigm

Data are separated on two (independent) multitude, L_1 – for training, and L_2.- for test.

F. Software realizations and evolutions

Model–View–Controller (**MVC**) is a software architecture, currently considered as an architectural pattern used in software engineering. The pattern isolates "domain logic" (the application logic for the user) from input and presentation (GUI), permitting independent development, testing and maintenance of each[2].

The **model** is the domain-specific representation of the data upon which the application operates. Domain logic adds meaning to raw data (for example, calculating whether today is the user's birthday, or the totals, taxes, and shipping charges for shopping cart items). When a model changes its state, it notifies its associated views so they can be refreshed.

Many applications use a persistent storage mechanism such as a database to store data. MVC does not specifically mention the data access layer because it is understood to be underneath or encapsulated by the model. Models are not data access objects; however, in very simple apps that have little domain logic there is no real distinction to be made. Also, the ActiveRecord is an accepted design pattern which merges domain logic and data access code - a model which knows how to persist itself[2].

The **view** renders the model into a form suitable for interaction, typically a user interface element. Multiple views can exist for a single model for different purposes.

The **controller** receives input and initiates a response by making calls on model objects.

An MVC application may be a collection of model/view/controller triplets, each responsible for a different UI element[2].

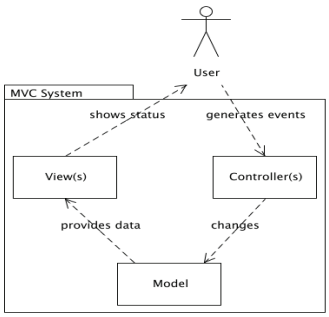


Fig.5

Model-driven architecture is focused on forward engineering, i.e. producing code from abstract, human-elaborated modelling diagrams. One of the main aims of the MDA is to separate design from architecture. As the concepts and technologies used to realize designs and the concepts and technologies used to realize architectures have changed at their own pace, decoupling them allows system developers to choose from the best and most fitting in both domains. The design addresses the functional (use case) requirements while architecture provides the infrastructure through which non-functional requirements like scalability, reliability and performance are realized. MDA envisages that the platform independent model (PIM), which represents a conceptual design realizing the functional requirements, will survive changes in realization technologies and software architectures.

Of particular importance to model-driven architecture is the notion of model transformation.

III. CONCLUSION

The evolution of program paradigms is always on move. In close future maybe will be found more abstractive and more advanced tools and approaches for analysis and estimation.

REFERENCES

[1] Nenov, Hr. “Algorithms and systems for estimation of condition and managing of complex objects” PhD thesis 2008.

[2] www.sun.com

This page intentionally left blank.

Fuzzy Based Approach for Decision Support System

Pavlina St. Vladimirova[1] and Daniela D. Ilieva[2]

***Abstract* – The paper discusses a methodology for decision making over alternative execution of plans represented as control nets, based on a class of Petri nets. The idea is that fuzzy attributes are associated to operations, each with their weight. Fuzzy attributes of operations extend to fuzzy attributes of executions. Optimal executions can be computed as intersections of execution attributes. The methodology is supported of computer tools to graphical plan representation and computations, which are necessary of controlling plan executions in real time.**

***Keywords* – decision making, DSS, fuzzy logic, Petri nets, fuzzy attributes**

I. INTRODUCTION

Conceiving projects and devising plans are characteristic expressions of any goal-directed human activity. In practice, many different plans will be necessary for making a project realizable: scheduling plans, cost plans, plans of resource flow, etc. How to integrate heterogeneous, possibly conflicting, plan attributes for choosing an optimal course of action during the realization of project? In general, plans allow for alternative executions, and always they have different times, costs, failure rates, setup times, etc. Choosing an optimal course of action is a critical question when running plans which allow for alternative moves. In this paper we discuss a multi-attribute methodology for making decision over alternative execution of plans represented as control nets. This methodology is suitable for making decisions at operational level of running a plan.

II. CONTROL NETS

Considered methodology requires plans, where, by constructions, alternative courses of action do not interact, and choice between alternatives is controlled by one decision. Control nets are a class of place/transition Petri nets defined with this issue in mind. In plans expressed by control nets the independence of execution alternatives is guaranteed. Decisions are only made at control places, special places associated with the sets of alternative executions – one control place for each such set.

Control nets are based on four composition modules: T-sequence, S-sequence, synchronization and choice. T-sequence represents sets of operations to be executed one after the other. Synchronization modules represent simultaneous production and subsequent simultaneous consumption of a set of resources. S-sequence represents sets of resources which are available one after the other. Choice modules represent resource sharing by alternative operations, both forward and backward.

Control nets are constrained in structure. The gain of this limitation is the clear-cut definition of execution alternatives. Decision over execution alternatives arise at some places of choice modules, and only there. And every branch out of such a place is guaranteed to represent an executable alternative.

III. METHODOLOGY TO SUPPORT DECISION MAKING

Plans allow for alternative executions, and different executions will mostly have a different degree of desirability, depending on times, costs, failure rates, setup times, etc. The execution supervisor is responsible of making better decision in choice situations. He has to attach fuzzy attributes to operations – such as short time, low costs, acceptable failure rates, low labour rates, etc. From operation attributes he can deduce attributes of the alternative executions – fuzzy too. This done, the supervisor can use different multiatribute decision making techniques: to find optimal alternatives by intersecting attributes or to apply outranking techniques in order to partition the set of alternatives into several preference classes.

A. Fuzzy attributes over Operations

Let $N = (S, T, F, W, M_0)$ is a control net, and $c = (S', T', F', W', M')$ is a primary choice schema of it.

The execution supervisor has to associate a finite set of fuzzy attributes with T_c - the transition set of *c*. A fuzzy attributes over operations is a fuzzy set $A = \{t, \mu(t)\}$ with support T_c. The degree of membership $\mu(t)$ of transition *t* is to be interpreted as the degree, to which operation *t* exhibits attribute A [1], [3].

Short durations, low costs, small failure rates, etc. are examples of properties which may well be expressed as fuzzy attributes. Consider the choice schema in Fig. 1, which has four branches. The property "low cost" could be specified by the fuzzy set $C = \{t, \mu_C(t)\}$ with

$\mu_C(t) = \{(1, 0.9), (2, 0.4), (3, 1.0), (4, 0.2)\}$

and property "short duration" could be specified by the fuzzy set $D = \{t, \mu_D(t)\}$ with

$\mu_D(t) = \{(1, 0.2), (2, 0.8), (3, 0.0), (4, 0.6)\}$.

[1]Pavlina St. Vladimirova is from Technical University – Varna, Faculty of Computer Sciences and Technology, 1 Studentska str., 9010 Varna, Bulgaria, E-mail: pav_varna@yahoo.com

[2]Daniela D. Ilieva is from Technical University – Varna, Faculty of Computer Sciences and Technology, 1 Studentska str., 9010 Varna, Bulgaria, E-mail: ilievadaniela@abv.bg

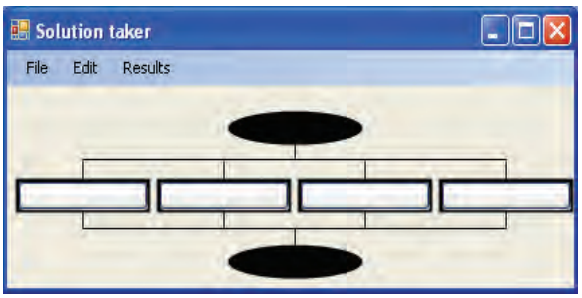


Fig.1. Choice schema

These two attributes mean: the degree to which operation 1 has the attribute “low cost” is 0.9 and “short duration” - 0.2, etc.

B. Attributes over Alternatives

Alternatives of primary choice schema are made up of sequences and of synchronization modules. Hence, the extension of an attribute over operations to an attribute over executions require two operators, one for aggregating the membership degree of transitions belonging to sequences, and one for aggregating the membership degree of transitions belonging to synchronization modules. These two operators must be specified by the supervisor for every operation attribute. The specification of these two operators expresses the supervisor’s beliefs about how the considered attribute spreads from operations to execution alternatives. This requires taking the meaning of the individual attributes into account. For instance, the aggregation of time related attributes must be different from the aggregation of cost related ones. Indeed, it is natural to view the joint cost of two parallel activities as the sum of their costs, but their joint duration will naturally be interpreted as the union of the corresponding time intervals.

Consider a primary choice schema *c*, and a fuzzy attribute A= {t, μ (t)}, defined over its set of transitions. Let C = {c_1, c_2, … , c_n} be the set of alternatives of *c*. Let a fuzzy attribute over the alternatives of *c* to be a fuzzy set A’= {c_i, μ’ (c_i)} with support *C* which membership function is computed on basis of a fuzzy operation attribute *A* by means of two operators assigned by the supervisor: operator *X* for aggregating the membership degree of operations belonging to sequences, and operator *Y* for aggregating the membership degree of operations belonging to synchronization schemata [3]. Operators *X* and *Y* must be specified so that:

for every S or T- sequences *s* with transitions set {t_1, t_2, …, t_n}

X (s) := X (μ (t_1), μ (t_2), … , μ (t_n)) $\in$ [0, 1];

for every synchronization schema *z* with transitions set {t_1, t_2, …, t_n}

Y (z) := Y (μ (t_1), μ (t_2), … , μ (t_n)) $\in$ [0, 1].

The membership degree μ’(c_i) of alternative c_i for the fuzzy attribute over alternatives A’ is obtained in the following way:

1. If c_i is an S-sequence, we set

μ’ (c_i) := X(c_i) and terminate.

2. If not, we run bottom-up through the hierarchy of plans, which we assume indexed by h = 1,2, …, n, and carry out the following computations:

Set μ_n(t) := μ (t) for all transitions t_n, and h := n;

REPEAT

- decrement h by 1;
- for each macro place p of the level h plan which is substituted by an S-sequence *a* in the level h+1 plan, set μ_h(p) := X(a);
- for each macro transition t of the level h plan which is substituted by an T-sequence *b* in the level h+1 plan, set μ_h(t) := X(b);
- for each macro transition t of the level h plan which is substituted by synchronization module *w* in the level h+1 plan, set μ_h(t) := Y(w);
- for all places p and transitions t for which μ_{h+1}(p) or respectively μ_{h+1}(t) were defined, set

μ_h(p) := μ_{h+1}(p) and μ_h(t) := μ_{h+1}(t)

UNTIL in the level h plan c_i contains exactly one transition t’;

3. Set μ’ (c_i) := μ_h(t’) and terminate.

C. Attributes over Executions

Fuzzy attribute over alternatives lead naturally to fuzzy attributes over executions.

Executions of choice schemata are executions of multisets of concurrently enabled alternatives. If c_1, c_2, … , c_n are the actually enabled alternatives of choice schema c, multiset

e := $m_1c_1 \oplus m_2c_2 \oplus \ldots \oplus m_nc_n$

represents the execution of c consisting in executing – in any order and for all i – m_i times alternative c_i.

Let E denote the actual execution set of *c*. The fuzzy attribute over the executions of *c* is the fuzzy set A’’= {e, μ’’(e)} with support C which membership function

μ’’(e) = [$m_1\mu$‘(c_1) + $m_2\mu$‘(c_2) …+ $m_n\mu$’(c_n)] / $\sum_i m_i$,

with $\sum_i m_i$ = M (c’) .

The membership degree of execution e in the fuzzy attribute over executions is the weighted average of the membership degrees of the alternatives constituting *e* in corresponding fuzzy attribute over alternatives. The membership degree of *e* in fuzzy set A’’ is a linear function of the membership degree of c_i in fuzzy set A’.

D. Preference degree of executions. Optimal Executions

Fuzzy attributes over executions can be used to determine the set of optimal execution by the follow way. Consider any choice schema *c*, together with its actual marking M.

Let E = {e_1, e_2, … , e_n} be actual set of executions of *c*, and A = { A_i } a finite set of fuzzy attributes over E, with A_i= { e, μ_i (e)}. Let all the above attributes express desirable features

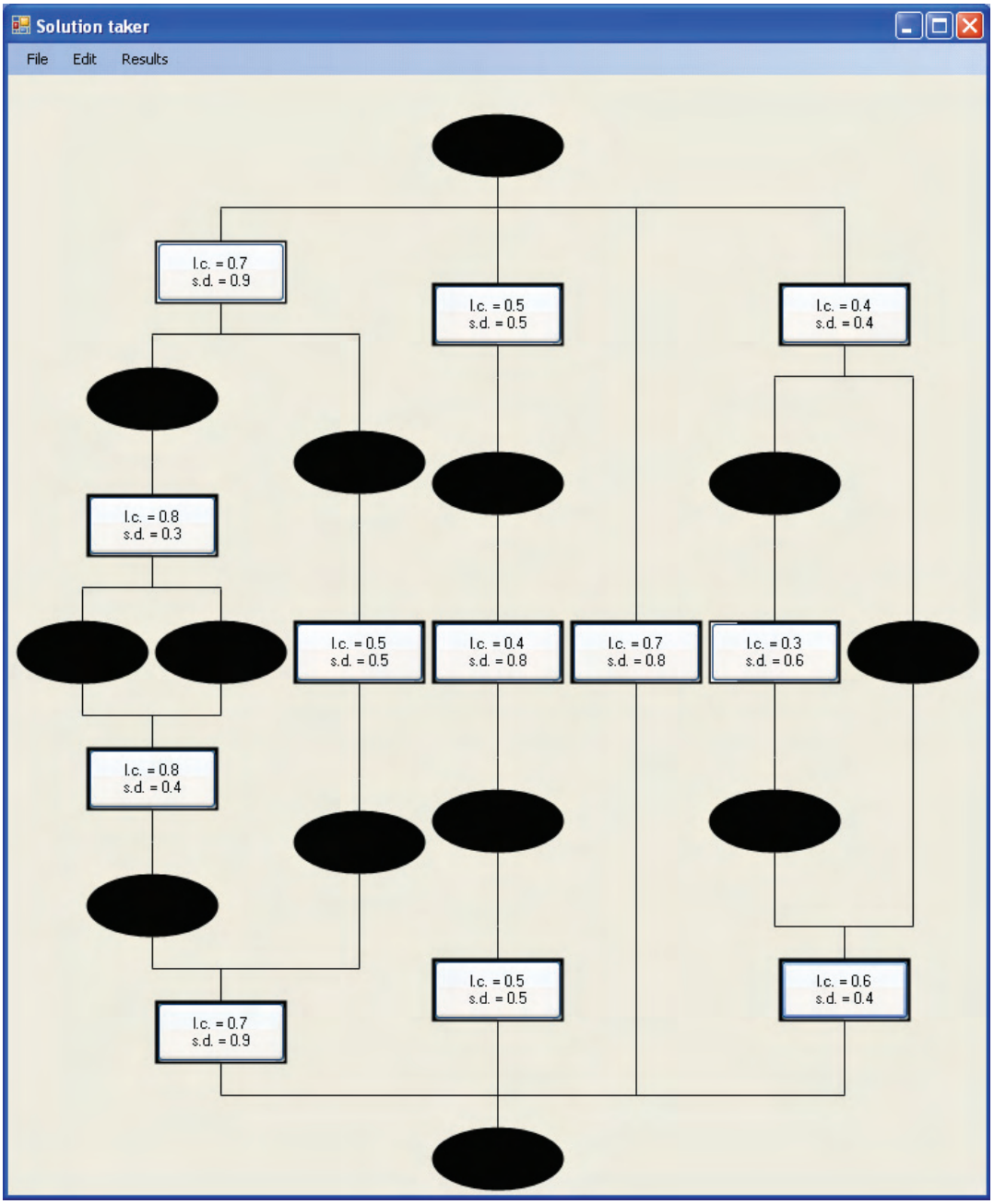


Fig.2. Control net, represented simple project plan with assigned 2 fuzzy attributes for every operation

of executions. The real number μ_i (e) represents the degree to which execution *e* exhibits attribute A*i*.

The intersection Δ of the fuzzy sets A*i* is fuzzy set over E too:

$\Delta := \cap_i A_i = \{e, :\Delta (e)\}$.

The membership degree of execution *e* in fuzzy set Δ is the smallest membership degree of *e* in any fuzzy set A_i:

$\mu_\Delta (e) := \min_i \mu_i (e)$.

μ_Δ (e) represents the degree to which execution e possesses the set of attributes A. Since the attributes A_i express desirable feature of executions, Δ may be interpreted as the fuzzy attribute "preferable". μ_Δ(e) represents then the preference degree of execution *e*.

Optimal executions are executions with maximum preference degree. That is: optimal decisions are execution e' ∈ E such that

$\mu_\Delta (e') := \max_E \mu_\Delta (e)$.

IV. Decision support sysytem "Solution Taker"

The above methodology is embedded in the developed decision support system "Solution Taker". The system allows easy construction of a control net for a project plan, assigning fuzzy attributes to the operations and automated calculation of relevant attributes for the execution alternatives. In Fig.2 is given an example of sample project plan, constructed by this

graphical tools. The figure is the result of top-down expanding of the simple choice module, given in Fig.1. In Fig.3 is shown the way of attaching different attributes to the operations. To each operation are assigned the fuzzy attributes “low costs” and “short time” with their degree of membership.

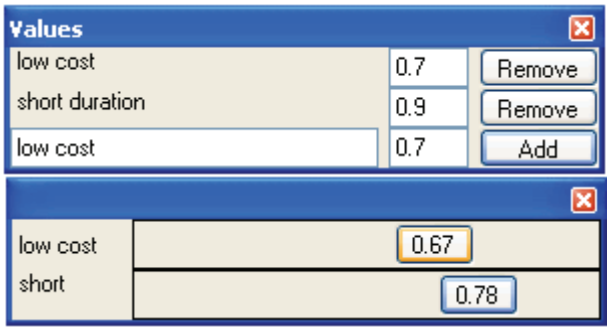
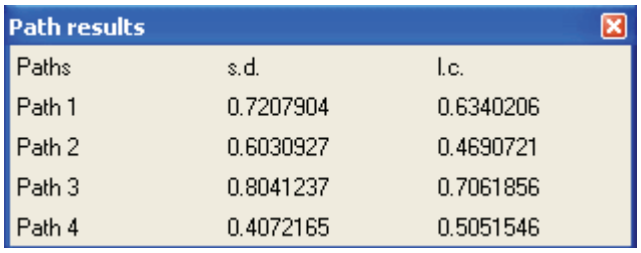


Fig.3. Tools for adding fuzzy attributes to the operations and giving them values

In Fig.4 we can see the values of attributes for every alternative for the choice schema in Fig.2 and that the most effective on the examined criteria is the third alternative.



Paths	s.d.	l.c.
Path 1	0.7207904	0.6340206
Path 2	0.6030927	0.4690721
Path 3	0.8041237	0.7061856
Path 4	0.4072165	0.5051546

Fig.4. Path results for the fuzzy attributes

V. Conclusion

In this paper is proposed a methodology for making operational decisions on the basis of multiple fuzzy attributes and software product is developed to support this methodology. The ability to graphically display and automation of the calculations make the system useful for making and maintaining solutions to manage systems in real time.

References

[1] S. Chen, C. Hwang, Fuzzy Multiple Decision Making. Berlin, Germany: Springer-Verlag, 1992.
[2] H.A. Taha, Operational Research. An introduction. Prentice Hall, 1997.
[3] A. Pagnoni, Project engineering: computer-oriented planning and operational decision making. Berlin, Germany: Springer-Verlag, 1990.
[4] P. Vladimirova, K.Ganchev, G. Kunev, An approach of operational decision making. Second international scientific conference “Computer science”, Greece, 2005.

Information Technology Influence on Bank For Business

Ramaj Xh. Vehbi[1], Hasani I. Vjollca[2] and Shabani H. Enver[3]

***Abstract* – Enterprises in general and particular financial institutions rely on information technology (IT) services to provide the infrastructure required to support their business goals. In order to fulfill these goals, the IT services must be available and should have the capacity to grow smoothly as business requirements expand. In this aspect to verify the influence of information technology it's necessary to make a general analysis of business and technological process on Bank for Business as a research case.**

***Keywords* – Information technology, bank for business, banking software, service, and architecture.**

I. INTRODUCTION

The purpose of this technical analysis on IT infrastructure in (BFB) is simply led to the improvement and development of IT infrastructure in (BFB), and in the development and promotion of banking services in general. A typical enterprise-level organization's data center contains a complex mix of vendor hardware and software as well as technology professionals who provide services throughout the IT life cycle. These components need to be integrated to ensure that they work well together. The components range from networking devices to application servers to storage devices, and each includes the necessary software components. Each of these components has a large number of potentially valid configurations, but only a few of these configurations result in an integrated, functional system. Determining the right configurations and then implementing and maintaining them it's not an easy task but its understood that finding the right technical solution it will influence in particular on information technology department and in general in our institution as in this case Bank for Business (BFB). In General Information Technology in a explicit way influence the business flow and business operations on Banka for Business.

However, it's evident the presence of influence of information technology in all business aspects and especially it would be emphasized here on automation of business process and business development of Bank for Business.

In the cycle of information technology services should have considered various services and equipment, so each service has different orientations for the planning, design and operation in the client needs.

[1]**Ramaj Vehbi** - Faculty of Economy, Magisterial Pristine-Fushe Kosovo,10000, Pristine,Kosovo,E-mail: vehbiramaj@yahoo.com
[2]**Hasani Vjollca**- Faculty of Economy, Magisterial Pristine-Fushe Kosovo, 10000 Pristine, Kosovo, E-mail: vjollcav@gmail.com
[3]**Shabani Enver**-Faculty of Economy, Magisterial Pristine-Fushe Kosovo,10000Pristine,Kosovo,E-mail:enver.shabani@gmail.com

TABLE I

ANALYSIS OF THE OVERALL ARCHITECTURE IS DONE IN THESE AREAS AGE LAYOUT DESCRIPTION.

Architecture	Description
Network and System Infrastructure	Design factors supporting the devices of infrastructure in order to have effective communication with the organization
Security Architecture	Security Strategy including security domains, policies and procedures
IT Management	Factors of Information Technology Department Management
Storage Architecture	Factors affecting the unity and centralization of data storage networking
Core Banking Software	Improvements and updates for the core software infrastructure

II. NETWORK AND SYSTEM INFRASTRUCTURE

A. Network architecture definition

Network architecture combines a set of technological solutions in a complex environment that contains available, secure, flexible, manageable, and achievable services. Network device operate different levels of OSI (Open System Interconnection).The level of the enterprise network architecture is generally composed of the following elements and technologies.

B. Local Area Network segments

Devices that are used for switching from the second level of the OSI model are from well known brand name 3 Com with in models: XM 3 com 3300 switch, 3 Com 4250 switch, 3 Com 4226 switch T – those devices fulfills the current needs and requirements in (BFB).

Switches are connected with vertical cable Fast Ethernet ports while on horizontal cable are connected all other devices as it is required by network standards and regulations.

Those equipments support Wireless LAN technology meaning that has the possibility of traffic optimize and also increasing the level of security for network infrastructure.

TABLE II

SERVICES AND EQUIPMENT ARE INCLUDED IN THIS WORK ARE:

Service	Technologies
Network Devices	Routers, switches, virtual local area network (VLAN), load balancers
Computing Devices	Server classes, hardware requirements
Storage Devices	Direct-attached storage (DAS), network-attached storage (NAS), storage area network (SAN)
Deployment Services	Installing and configuring operating systems WinPE, SysPrep, and RIS.
Network Services	DNS, DHCP, and WINS.
Firewall Services	Perimeter and internal firewalls and Web proxy service (ISA Server).
Directory Service	Active Directory® directory service.
File and Print Services	Distributed File System (DFS), printing configurations, and print devices.
Data Services	Microsoft SQL Server, Windows clusters.
Web Application Services	Microsoft Internet Information Services (IIS)
Infrastructure Management Services	Debug facilities, deployment capabilities, and remote management and tools.
Backup and Recovery	Data and systems backup and recovery infrastructure.
Certificate Services	Public key infrastructure (PKI).
Remote Access Services	Site-to-site virtual private network (VPN), Client VPN, routing and remote access service (RRAS), Internet authentication service (IAS).
Middleware Services	.NET Framework.
Messaging Services	Microsoft Exchange Server 2003

III. WIDE AREA NETWORKS

Bank for Business branches are connected with HQ with 256 kbps speed of connection, sub branches are connected with 128 kbps speed, while HQ has 1 mbps throughput.

Wide area network in BPB is using wireless technology. Frequencies that are used to connect branches are 5 GHZ and sub branches are 2.4 GHz.

Routing devices are from well know brand Cisco Systems, Wide Area Network is managed by internet service provider. Traffic between branches, sub branches with HQ is encrypted.

Regarding the wireless technology for the banking purposes is not one of the best technical solutions if we consider rehabilitees, security and confidentiality factors but for the moment this is the only available solution and as it is right knows it is a good choice of components and technology regarding the elements mention above.

Regarding WAN management it must have some changes specifically Routing devices must be managed by BPB IT department while wireless communication lines can be managed by internet service provider.

IV. SYSTEM DEFINITION

C. Workstations:

(BFB) have various computer equipment "brand-name" and "non-brand name" with various configurations, which affects more negatively in the administrator and maintenance of such equipment. This is a critical point in the infrastructure.

It is imperative that configuring with 64 MB RAM memory must be updated with at least 256 MB. Disk Space does not appear as critical points in the case of infrastructure.

From the perspective of operating systems on workstation computers BPB use Microsoft Windows 2000 and Microsoft Windows XP Professional, those operating systems are fulfilling are need for Core Banking Software requirements.

In this case it’s preferable to have the technical notes and keep books about repairs on equipment managed by IT Department, same time it’s recommended to have one operating system for all workstation in our case best solution is Microsoft Windows XP Professional.

D. Servers

Currently Bank for Business has 2 servers dedicated and 2 workstations adapted for servers use and one Cisco ROUTER3600.

First Server has DNS, DHCP, and Active Directory Roles; second one is database server while other 2 workstations have firewall service and old core system installed.

All those servers use Microsoft Windows Server 2003.

In order to have better service, performance and security it is necessary to made an overall systems review.

This review was made based on the three main principles: confidentiality, integrity and availability including the infrastructure on hardware, network and software platform.

It is recommended to be made the standardization of all the equipments including the personal computers, servers and network. The standardization will help your organization in reduce the costs in maintenance and better services.

Due to the fact that now we have advanced physical gigabit LAN, to have it fully functional, would need for all the computers and servers to purchase network gigabit cards of network and one gigabit network switch.

For security reasons it is mandatory to have licensed antivirus.

For business contingency i would recommend the following solutions:

- For our main current server i recommend to have those rules: Active Directory; DNS/WINS service; DHCP Service
- Antivirus Server,
- Current MSISA Server to have same rule as it has right know,
- New server, i recommend to have those rules: Database Server; File Share Server.

For illustration purposes below it is presented the current situation of IT infrastructure on the bank for business.

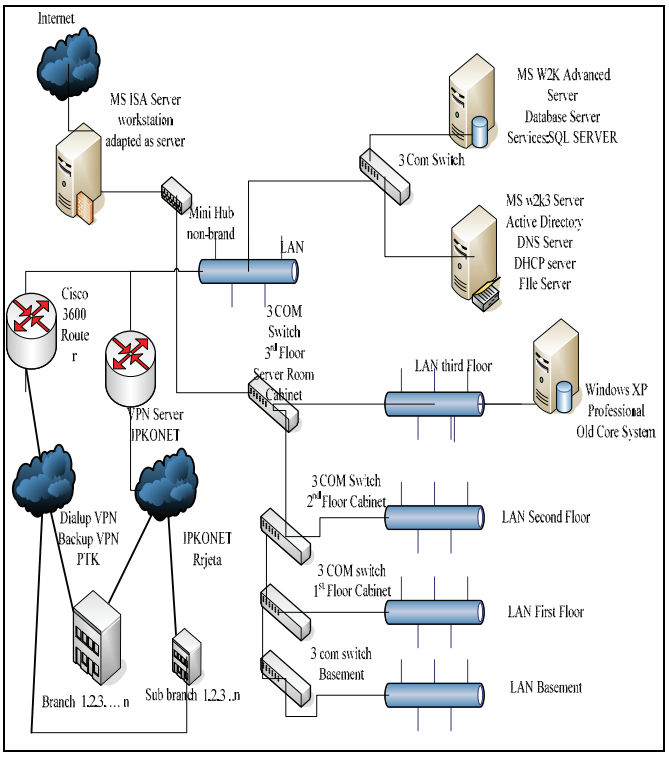


Fig. 1.Current situation of IT infrastructure

V. Security Architecture

The security considerations facing enterprise organizations and their IT (information technology) systems are considerable, with many design issues to be studied and resolved. In order to address security design issues appropriately, however, security architecture must first be looked at from a business need perspective.

People, processes, and technologies are interdependent, and the potential for a security breach increases with a deficiency in any of these elements

To determine how the security is function the most important is to define the architecture security values.

E. Assets Defined

Security architectures are designed to protect the assets of an organization, where assets are defined as resources that provide value to the organization. While there are many types of assets, this blueprint focuses on only two types of IT assets: data and tiers.

VI. Standards and Guidelines

ISO 17799 is an internationally accepted set of controls comprising best practices in information security. This standard was based on and supersedes the British standard BS 7799.

Security except all other security elements those standards mandating use of these services:

- Firewall software
- Internet Protocol Security (IPSec) software
- Anti-virus software
- Cryptography software

The following figure illustrates as a big picture how would be a proper IT (information technology) security program using available technologies for mitigating risks at each of the layers identified most of all it security programs.

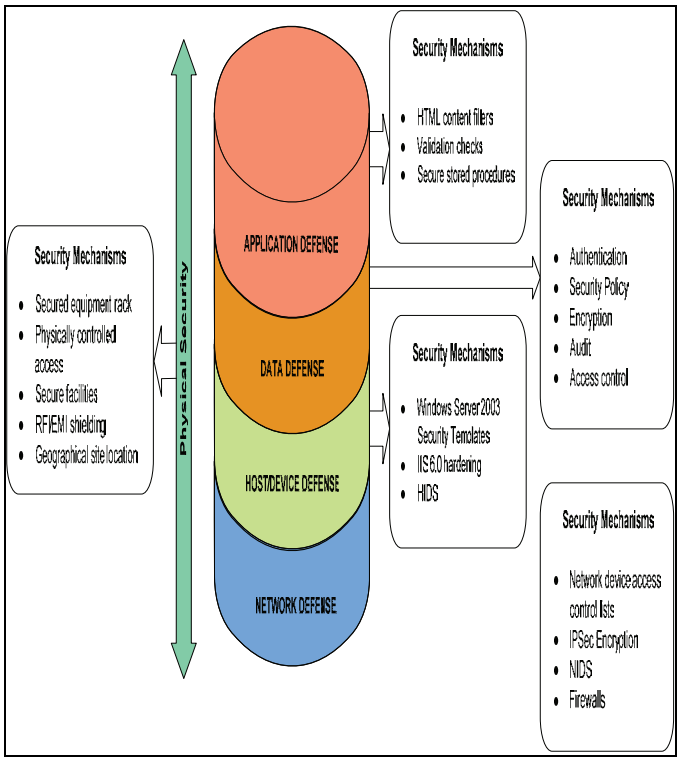


Fig.2.Security program Show Romanization

V. SECURE ACCESS TO LOCAL NETWORK

F. Dialup

A link to existing access from outside the corporate network based on discussions with workers who have made information technology service but there is certainly information technology services has no access to this service so that it is not known what equipment configuration have been essential that due security that IT workers have access to this router. Sometime where should be a policy is existing security. In the same time it dials up connection serves as backup in case of fall of the provider VPN connection. IT Services has no access to the router or the IPKO which represents the critical point in our security.

The following illustrations presents how it is connected the dialup connection:

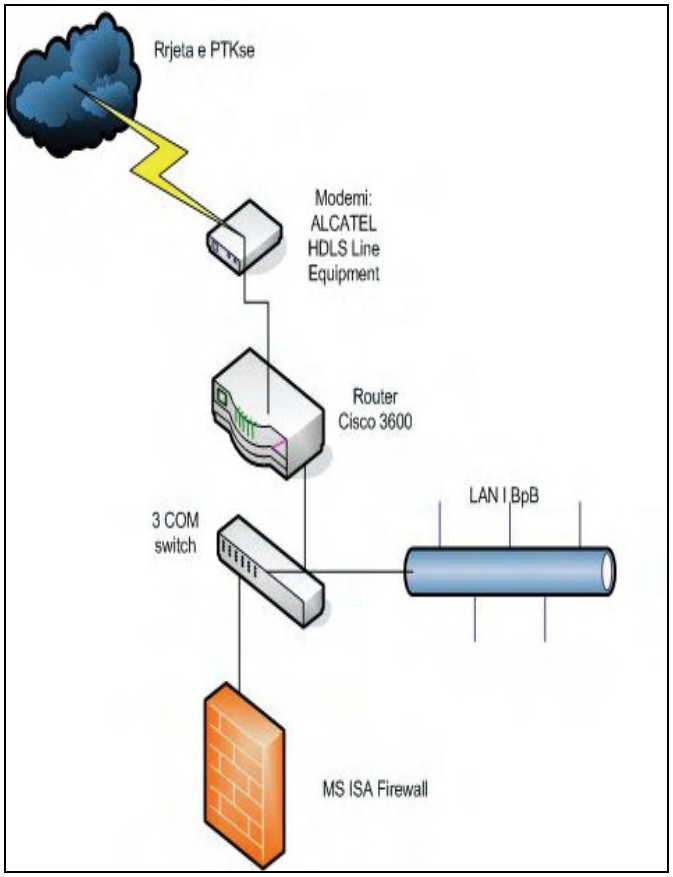


Fig.3.Conected to dialup connection

VII. CONCLUSION

The purpose of this paper is to the role and significance of information technology in the companies such as in this case,Bank for Business. Therefore the purpose of this analysis is that the impact on improving and developing infrastructure in Information Technology for Business Bank with minimum costs that directly affects the development of services and efficiency in the affairs of the bank, the documentation department of information technology, assistance for planning and design of infrastructure.In general, the network is currently operational infrastructural but wanted to make some visible correction which would help services, as well as security architecture is organized on a functional level request but require fundamental changes in architecture because of several critical security releases , a data storage architecture does not exist at all, the banking software infrastructure is well,organized!

In order to clarify the list of critical points that have emerged from this technical analysis we ranked them in several levels ofactions:

-Necessarily denoting actions that require changings in infrastructure.

-Actions for optimum represent actions which help optimazing services.

-Actions that represent continuous action that must be maintained continuously.

REFERENCES

[1] Agarwal, R., "Individual Acceptance of Information Technology," in R. Zmud (ed.), Framing the Domain of IT Management, Cincinnati, OH: Pinnacle Educational Resources, 2000.

[2] Iyer, L. S., et al., "Global E-Commerce: Rationale, Digital Divide, and Strategies to Bridge the Divide," *Journal of Global Information Technology Management*, Vol. 5, No.1, 2002.

[3] Kannen, P. K., et al., "Marketing Information on the I-Way," *Communications of the ACM* 41, no. 3 (1998): 35–40.

[4] Kanter, J., *Managing with Information*, 4th ed. Englewood Cliffs, NJ: Prentice Hall, 1992.

[5] Kleiner, A., "Corporate Culture in Internet Time," *Strategy-Business*, (Booz-Allen & Hamilton Quarterly, *strategy-business.com*), First Quarter, 2000.

[6] Kopp, S. W., and T. A. Suter, "Fan Sites and Hate Sites on the World Wide Web," *Quarterly Journal of Electronic Commerce* 2, no. 4, (2001).*Korean Times*, news item, August 24, 2002.

[7] Liu, S., et al., "Software Agents for Environmental Scanning in Electronic Commerce," *Informations Systems Frontiers*, Vol. 2, No. 1, 2000.

[8] Richardson, R., 2003 CSI/FBI Computer Crime and Security Survey. San Francisco: *Computer Security Institute*, 2003 (*gocsi.com*).

[9] Rogers, M., and T. M. Rajkumar, "Developing E-Commerce Web Sites for the Visually Impaired," *Information Systems Management*, Winter 1999.

[10] Rykere, R., et al., "Online Privacy Policies: An Assessment," *Journal of Computer Information Systems*, summer (2002).

[11] G.B.Davis&M.H.Olson: Management Information Systems, McGrow Hill.

[12] Zelko Panian, „Information Technology", Zagreb, 2006.

[13] Jane Loudon and Kenneth Laudon—Management information Systems-Managiong , the digital firm (9th edition)-2006

[14] Judy Strauss, Adel el-Ansary, Raymond Frost---E-Marketing (third edition), Prentice, Hall, New Jersey, 2003.

[15] Ceric V. Varga. M. "Information technology in business", Zagreb, 2006.

POSTER SESSION PO X

PO X - Computer Graphics

Video Surveillance using Augmented Virtual Environments

Aleksandar Lj. Milosavljević[1], Aleksandar M. Dimitrijević[2] and Dejan D. Rančić[3]

***Abstract* – In this paper we present video surveillance technique based on Augmented Virtual Environments (AVE). AVE represents 3D virtual environment augmented with multiple video streams that are fused with 3D models in a real–time. This approach enables an observer to comprehend real–world video from arbitrary views of the scene. Our solution is based on use of a 3D GIS as the virtual environment. To enable registration of video frames into the 3D GIS we proposed a method for camera view modeling applicable to PTZ video cameras.**

***Keywords* – Augmented Virtual Environments, Video Surveillance, Geographic Information Systems, Virtual Reality.**

I. INTRODUCTION

The ability of GIS to handle and process both location and attribute data distinguishes it from other information systems. It also establishes GIS as a technology that is important for a wide variety of applications [1]. Traditionally, the majority of geographic information systems were limited to the visualization of geospatial data in two dimensions (2D GIS). The fact that we relate to our world in three or more dimensions suggests that some types of data may be more readily visualized and analyzed in 3D [2]. With the development of graphics hardware, virtual reality techniques originally developed for interactive computer games are exerting more and more influence in the field of 3D GIS. Li *et al.* [3] expressed the need for 3D GIS for urban environments in order to understand the 3D landscape with many high buildings.

Research presented in this paper deals with the integration of 3D GIS and video surveillance systems. Real–time video monitoring is playing an increasingly significant role in surveillance systems in various security, law enforcement, and military applications [4]. A typical outdoor urban surveillance system consists of multiple cameras overlooking different areas. However, conventional video monitoring systems have various problems with multi–point surveillance [5]. A typical system of conventional video monitoring connects each video camera directly to a corresponding display screen. Therefore, we have as many screens as video cameras. In these kinds of systems, serious problems can occur when the scale of the monitoring system grows larger than human capacity. Security personnel must mentally map each surveillance monitor image to the corresponding place in the real world, and this complicated skill requires experience and training [5]. To enable multi–camera coordination and tracking, Sankaranarayanan and Davis [6] emphasized the importance of establishing a common reference frame to which each of these cameras can be mapped. They suggested the use of GIS as a common frame of reference because it not only provides a solid ground truth, but more importantly provides semantic information (e.g., locations of roads, buildings, sensitive areas, etc.) for use in applications such as tracking and activity analysis.

Addressing the problem of the human ability (or lack thereof) to successfully fuse and comprehend the information that multi–point video surveillance can provide, Neumann *et al.* [7] proposed a visualization approach based on an Augmented Virtual Environment (AVE). The AVE is a virtual reality model augmented by multiple video streams in real–time to help observers comprehend temporal data and imagery from arbitrary views of a scene.

Our implementation of AVE relies on use of augmented and virtual reality techniques applied to GIS. Augmented reality (AR) aim to combine the real scene viewed by a user and a virtual scene generated by a computer that augments the scene with additional information. Unlike virtual reality (VR), which provides the user with a synthetic environment as a replacement for reality, augmented reality ensures that the user sees the real environment augmented with objects and information from the virtual environment. In order to better understand term "augmented reality", reality–virtuality continuum defined by Milgram and Kishino [8] should be considered (see Fig. 1). The "real world" and a "totally virtual environment" are at the two ends of this continuum while the middle region is called mixed reality (MR). Augmented reality is near to the real environment side, while augmented virtuality (AV) is closer to the virtual environment side. Unlike augmented reality, augmented virtuality adds real images to virtual environment increasing virtual object's reality degree.

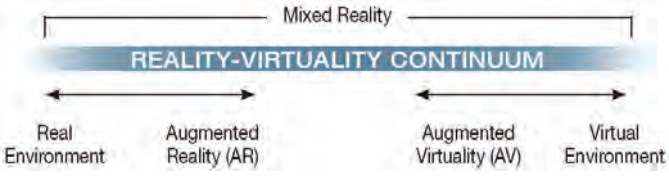


Fig. 1. Reality–Virtuality Continuum [8]

[1]Aleksandar Lj. Milosavljević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E–mail: aleksandar.milosavljevic@elfak.ni.ac.rs

[2]Aleksandar M. Dimitrijević is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E–mail: aleksandar.dimitrijevic@elfak.ni.ac.rs

[3]Dejan D. Rančić is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E–mail: dejan.rancic@elfak.ni.ac.rs

Both augmented reality and augmented virtuality requires 3D registration of real–world images (in our case video) within virtual environment. Our approach relies on use of 3D GIS as such environment. Conceptual diagram that illustrates integration of GIS and video surveillance is show in Fig 2.

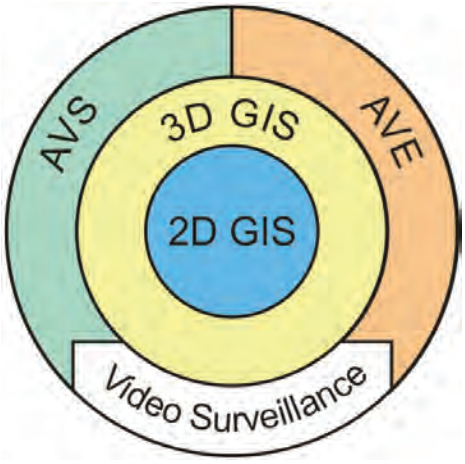


Fig. 2. Conceptual diagram of GIS and Video Surveillance integration

In the proposed conceptual diagram 2D GIS represents a core. It is used for geospatial data management and map visualization. A 3D GIS is built around 2D GIS, i.e. it uses 2D GIS services to enable 3D visualization of geospatial data. Finally, the last shell in this diagram corresponds to systems that integrate 3D GIS with video surveillance. These systems can be divided into two main categories:

- GIS Augmented Video Surveillance (AVS)
- GIS based Augmented Virtual Environment (AVE)

The first category use augmented reality techniques for the integration, while the second is nearer to augmented virtuality. Our previous work [9] included implementation of an AVS system, while in this paper deals with implementation of our AVE prototype *GeoScopeAVE*.

The paper is organized as follows: Section 2 presents a method for registration of camera video into 3D GIS. In section 3 we give an overview of the key features of the implemented prototype. Finally, section 4 presents the conclusions.

II. 3D GIS REGISTRATION OF CAMERA VIDEO

In this section a method for registration of PTZ camera video frames into 3D GIS is presented. PTZ is an abbreviation for Pan–Tilt–Zoom, and in the terminology of video surveillance, it indicates cameras that can rotate in the horizontal (pan) and vertical planes (tilt) and change their level of magnification (zoom). The registration is done in two steps:

- Establishing camera view absolute parameters
- Constructing camera visibility surface

Establishing camera's absolute 3D GIS view parameters require definition of an appropriate view model and transformations from camera's relative coordinate space.

A. Modeling PTZ camera view in 3D GIS

An observer view into the 3D GIS is fully determined by the following seven parameters that can be divided in two groups:

- Position parameters:
  (1) Latitude
  (2) Longitude
  (3) Altitude
- Orientation parameters:
  (4) Azimuth (or yaw)
  (5) Pitch
  (6) Roll
  (7) Field of view (FOV)

When a PTZ camera is in the role of observer, the first group of parameters is fixed and determined by the camera mounting position. Knowing characteristics of the camera lens, the remaining 4 parameters are calculated using the retrieved *pan*, *tilt*, and *zoom* parameters.

The current azimuth, pitch, and roll are determined from the camera *pan* and *tilt* using three calibration parameters:

- Azimuth of the zero–pan camera position ($azimuth_0$),
- Pitch of the north–pan camera position ($pitch_0$), and
- Roll of the north–pan camera position ($roll_0$).

The calibration parameters are determined by the camera mounting. The first parameter ($azimuth_0$) represents the direction of the camera view when the *pan* parameter is set to 0. The other two parameters model small angular deviations in the horizontal plane that arrive from imperfect mounting, and they are represented as pitch and roll when camera is oriented to the north.

After the calibration, all parameters for setting virtual 3D GIS camera orientation are known, and the rotation matrix for aiming the virtual camera is calculated with the following formula:

$$R = R_z(-azimuth_0)\times R_x(pitch_0)\times R_y(-roll_0) \times R_z(-pan)\times R_x(tilt) \quad (1)$$

In the previous formula, R is used to represent rotation matrices. The subscript determines around which axis the rotation is taken, while the parameters in the parentheses are the angles of rotation measured in a counter clockwise direction. Parameters $azimuth_0$, $roll_0$ and $pitch_0$ are previously described calibration parameters, while input parameters *pan* and *tilt* determine the current orientation of the camera in its local coordinate system.

Based on the calculated rotation matrix R, vectors that determine camera orientation in the absolute coordinates of the 3D GIS are calculated with the following formulae:

$$\begin{aligned} \overrightarrow{V_{look}} &= [x_{look} \quad y_{look} \quad z_{look}]^T = R\times[0 \quad 1 \quad 0]^T \\ \overrightarrow{V_{side}} &= [x_{side} \quad y_{side} \quad z_{side}]^T = R\times[1 \quad 0 \quad 0]^T \\ \overrightarrow{V_{up}} &= [x_{up} \quad y_{up} \quad z_{up}]^T = R\times[0 \quad 0 \quad 1]^T \end{aligned} \quad (2)$$

As the subscripts suggest, the first vector V_{look} determines view direction, while the second V_{side} and the third V_{up} determine relative right side and up directions, respectively.

Absolute orientation parameters can be fully determined by two of these three vectors. Formulae that calculate *azimuth*, *pitch*, and *roll* based on V_{look} and V_{side} vectors are the following:

$$azimuth = \arctan\left(\frac{x_{look}}{y_{look}}\right), \quad pitch = \arcsin\left(\frac{z_{look}}{|\overrightarrow{V_{look}}|}\right), \quad roll = -\arcsin\left(\frac{z_{side}}{|\overrightarrow{V_{side}}|}\right) \tag{3}$$

In order to view a 3D scene in the same way that actual camera do, we had to model the virtual camera with the parameters of sensors used in a real camera. To calculate the angular field of view (*FOV*) based on the sensor's dimension (*S*), a lens's focal length (FL_{min}), and the current *zoom* factor, we used the following formula [10]:

$$FOV = 2 \cdot \arctan\left(\frac{S}{2 \cdot FL_{\min} \cdot zoom}\right) \tag{4}$$

B. Constructing camera visibility surface

Display of camera video in the 3D GIS scene is based on OpenGL texture projection technique [11]. Each video frame updates a texture that is projected into the scene from the camera viewpoint (using `GL_EYE_LINEAR` texture coordinate generation). A common issue of this technique is a dual projection – one along the projector's view direction, and another in the opposite direction. Another issue is a texture being projected onto all surfaces along the projector's view direction, while some of them are not truly visible for the camera viewpoint. To avoid these problems we have developed a method for constructing camera visibility surface to which frame texture projection is only applied. The method is applicable to an arbitrary complex scene. The only limitation is computation intensity which can reduce frame rate for scene rendering.

To construct camera visibility surface the following steps should be applied:

1. Setup the scene view to the absolute camera view
2. Setup the drawing viewport to the camera image size
3. Render the scene with light and textures disabled
4. Read the depth of each pixel in the resulting frame
5. Unproject each pixel with depth value using inverse transformation to determine it's absolute (x, y, z) coordinates
6. Using this 3D matrix create triangles from adjacent points.
7. Calculate each triangle normal and test if the angle between this vector and vector that points to the camera is less that (90^o - ε)
8. Create a display list with triangles that passed the test

Steps 1 and 2 are used to setup view to the scene in the exact same way the camera "sees" a real–world. Steps 3 and 4 are needed to acquire depth of the each pixel "seen" by the camera. The scene rendering in the step 3 is optimized for speed and not actually displayed on a screen. Steps 5 and 6 are used to create 3D matrix with absolute coordinates of a surface seen by the camera. To do so we need to unproject all pixels in the frame along with depth information using `gluUnProject` function. Finally, steps 7 and 8 are used to create display list with the camera visibility surface. This display list contains triangles constructed from previous 3D matrix that passed the surface orientaion test. This test is used to eliminate false surface triangles that "connect" different object's truly visible surfaces. The orientation test is illustrated in Fig. 3.

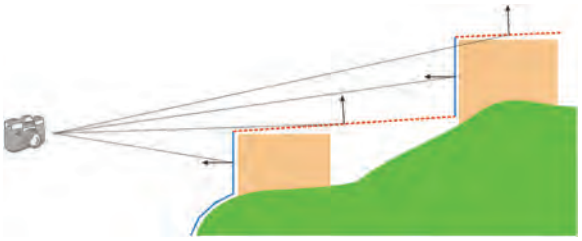


Fig. 3. Elimination of false visible surfaces using the orientation test

The proposed method is compute and data–intensive. In our testing, creation of a visibility surface for 768x576 frame image takes about 0.5–1 second. The resulting display list can contain up to 882050 triangles. Nevertheless, the advantage is that the surface constructed in this way remains valid until the camera orientation or the scene details change.

III. *GeoScopeAVE* Overview

In this section an overview of the developed *GeoScopeAVE* prototype is presented. The application is implemented in MS Visual Studio 2008 as C++ MFC project using OpenGL. For geospatial data access and 2D visualization the application relies on our existing GIS framework [12].

To enable accurate 3D visualization of any place on the Earth, our 3D GIS subsystem use ellipsoid based Earth model. Different terrain levels of detail (LOD) are dynamically created based on the observer's altitude and position. Each LOD consists of several blocks which has assigned texture and digital elevation model (DEM). Beside terrain visualization, the application can automatically create and visualize 3D objects (e.g. buildings) based on a 2D basis and maximum height. An illustration of a scene appearance constructed with the 3D GIS is shown in Fig. 4.

Camera video registration into such 3D GIS scene is based on previously described method. Fig. 5 illustrates registration results viewed from the camera's position, while Fig. 6 represents a view to an augmented virtual environment from an arbitrary position. More comprehensive video demonstration of *GeoScopeAVE* can be found at http://www.youtube.com/watch?v=8S-MhtEj2O0.

From the Fig. 6 it is clearly seen that a system that would use this approach is highly dependent on underlying GIS model. Since we are using simplified buildings representation, which does not include roof models; we have biggest registration errors in those areas. Nevertheless, when the scene is viewed from the camera position (or some near position) these registration errors seams less noticeable.

Without neglecting of identified shortcomings the implemented prototype proved the applicability of the proposed method.

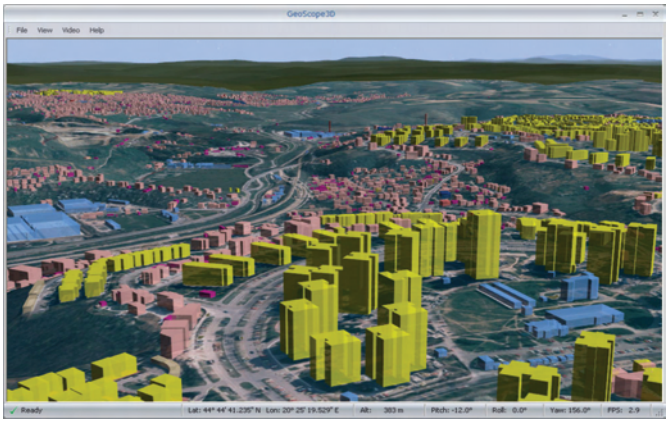


Fig. 4. 3D GIS scene appearance

Fig. 5. Registration of camera video into 3D GIS

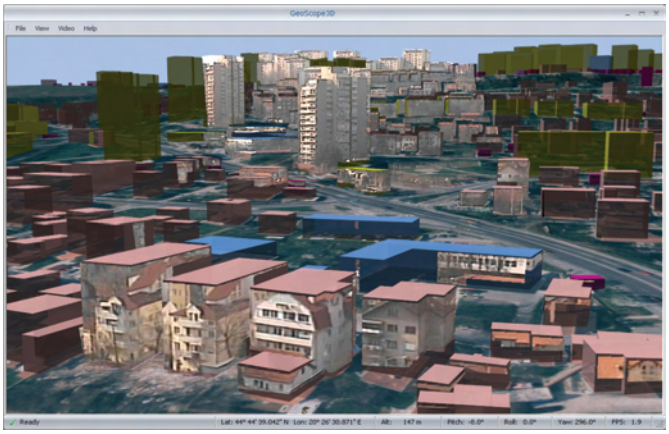


Fig. 6. An arbitrary view to an augmented virtual environment

IV. Conclusions

In this paper, we have presented the method for integration of GIS and video surveillance using augmented virtual environments. To enable fusion of surveillance camera video with 3D GIS virtual environment, we proposed an appropriate PTZ camera view model and registration method. In order to test proposed method we implemented *GeoScopeAVE* prototype which was described in the paper.

The major benefit from the presented approach comes from the fact that camera video is integrated within virtual scene and can be viewed from an arbitrary point and direction. It also has a potential of integration of multiple camera views into single virtual environment. Our future work will go in that direction.

However, there are also some serious drawbacks that should be stated. A system that would use this approach is highly dependent on underlying GIS model. It is also clear that objects (i.e. cars, trees, lamp poles) that are not present in the model are projected onto the buildings or the terrain and may look warped and distorted from other viewpoints. Another issue with AVE approach is performance. Calculation of camera visible surfaces, projection and update of frame textures are both compute and data–intensive.

The presented approach has potential applications beyond video surveillance. For example, it can be used for texture generation, validation, and construction of 3D models.

References

[1] K. Chang, *Introduction to Geographic Information Systems. 3rd ed.*, New York, NY: McGraw–Hill, 2005.

[2] S. Brooks and J.L. Whalley, “Multilayer hybrid visualization to support 3D GIS”, Computers, Environment and Urban Systems, vol. 32, no. 4, 2008, pp. 278–292.

[3] D. Li, Q. Zhu, Q. Liu and P. Xu, “From 2D and 3D GIS for CyberCity”, Geo–Spatial Information Science, vol. 7, no. 1, 2004, pp. 1–5.

[4] I. Oner Sebe, J. Hu, S. You and U. Neumann, “3D Video Surveillance with Augmented Virtual Environments”, First ACM SIGMM international workshop on Video surveillance, 2–8 November 2003, Berkeley, CA, pp. 107–112.

[5] N. Kawasaki and Y. Takai, “Video Monitoring System for Security Surveillance based on Augmented Reality”, Proceedings of the 12th International Conference on Artificial Reality and Telexistence, 4–6 December 2002, Tokyo, Japan, pp. 180–181.

[6] K. Sankaranarayanan and J.W. Davis, “A Fast Linear Registration Framework for Multi–Camera GIS Coordination”, Proceedings of the 5th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'08), 1–3 September 2008, Santa Fe, USA, pp. 245–251.

[7] U. Neumann, S. You, J. Hu, B. Jiang, J.W. Lee, “Augmented Virtual Environments (AVE): Dynamic Fusion of Imagery and 3D Models”, Proceedings of the IEEE Virtual Reality 2003, 22–26 March 2003, Los Angeles, USA, pp. 61–71.

[8] P. Milgram and F. Kishino, “A Taxonomy of Mixed Reality Visual Displays”, IEICE Transactions on Information Systems, vol. E77–D, no. 12, 1994, pp. 1321–1329.

[9] A. Milosavljević, A. Dimitrijević and D. Rančić, “GIS augmented video surveillance”, International Journal of Geographic Information Science, accepted for publication.

[10] J. Titus, “Make Sense of Lens Specs”, Test & Measurement World, 2005. http://www.tmworld.com/article/CA6252635.html

[11] C. Everitt, “Projective Texture Mapping”, 2001. http://developer.nvidia.com/object/Projective_Texture_Mapping.html

[12] A. Milosavljević, S. Đorđević-Kajan, L. Stoimenov, “An Application Framework for Rapid Development of Web based GIS: GinisWeb”, In: J.T. Sample, K. Shaw, S. Tu, M. Abdelguerfim eds. *Geospatial Services and Applications for the Internet*, New York, NY: Springer, 2008, pp. 49–72.

Application of Listed Data Structures in Knitting Industry CAD/CAM Systems' Development

Elena I. Zaharieva-Stoyanova[1]

***Abstract* – This paper treats the problem related to data representation in knitting industry CAD systems. Listed data structures are successfully used to code graphic data related to knitting pattern object. Knitting patterns are represented as sets of listed dots, so it is the reason to use listed data structure to describe knitting pattern object graphically. The goal of the paper is to give an approach of usage of listed structures in CAD systems' data representation. In this particular system the structures linked list is used. The data type does not allow using listed structures in MS VC++ collections. That is the reason to create listed structures related to knitting pattern graphic objects. The algorithms operated with list structures are suggested. These algorithms are applied to particular software module Construction. The application can be embedded in whole CAD/CAM system in knitting industry or it can be used as software application creating patterns of knitting products. The MS VC++ and MFC are used to create a software application. The main reason to create such project is that it can be used as software application from fashion designers to make boutique collections of knitting products. The project can be used as an example for CAD system development.**

***Keywords* – CAD/CAM systems, computer graphics, knitting industry, listed data structures, linked lists.**

I. INTRODUCTION

CAD/CAM/CAI systems are a part of a huge number of software applications launching to the market. Their development is related to software technologies at all. It means that introduction of new programming environments, software techniques, and new Operation Systems (OS) leads to creation of new CAD/CAM software application and to introduction of new versions of existing ones. One of the most important problems solving in CAD/CAM systems' development refer to data representation [1], [3]. Because this paper treats the problem of data representation in CAD systems in knitting industry, the data type will be discussed at first.

Generally, application of computer graphics in knitting industry CAD/CAM systems has two aspects: knitting structures design and design of knitting pattern shapes. The first one refers to design of knitting structures – intarsia, jacquard, lace, arans. The second one is related to knitting machine's capacity to make products by fully-fashion (FF) method. It means that the machine knits the parts (*cuts*) of products, or the whole products. This method allows to avoid cutting the materials, reduces the number of the operations and it leads to reducing the waste products to a minimum.

There is a big difference between data representing knitting patterns objects and data representing knitting structures. While the knitting patterns as graphic objects described by regions or polygons of straight lines and curves – vector graphics, the knitting structures are represented by means of raster graphics – rectangular grid of colored pixels or icons (small bit-map images). The difference between two types graphics objects demands the different approach to the design process and to the data representation, as well. To create a knitting structure a designer needs an editor for raster graphics images. Each item of the raster corresponds to the knitting stitch. A matrix form of an image data representation is suitable [8], [9], [10], [11], [12],[13].

As graphics object the knitting patterns are described as polygons of straight lines and curves. They are determined by set of *control points* (*knots*). Using these knots the computer can generate the whole image [4], [5], [6].

This paper treats the problem related to data representation in knitting industry CAD systems. Listed data structures are successfully used to code graphic data related to knitting pattern object. Knitting patterns are represented as sets of listed dots, so it is the reason to use listed data structure to describe graphically knitting pattern object. The goal of the paper is to give an approach of usage of listed structures in CAD systems' data representation. In this particular system the structures linked list and doubly linked circuit list are proper to use. The data type does not allow using listed structures in MS VC++ collections. That is the reason to create listed structures related to knitting pattern graphic objects. The algorithms operated with list structures are suggested.

II. FEATURES OF KNITTING PRODUCTS

The goal of this paper is to suggest an approach to CAD development using FF knitting method. The core of such system is editing of constructions and the shape of narrow sections of models, there by. The pattern editor, create different knitting structures would be programmed as additional module incorporated to the system [6], [7], [8].

This paper treats the problems related to CAD system development with possibilities to create models of knitting products; their constructions and variety of sizes. The project suggests an approach to data representation. Knitting structures and patterns design module will be a point of further development.

[1]Elena I. Zaharieva-Stoyanova is with the Faculty of Electronic Engineering, Technical University of Gabrovo, 4 H. Dimitar str. , 5300 Gabrovo, Bulgaria, E-mail: zaharieva@tugab.bg.

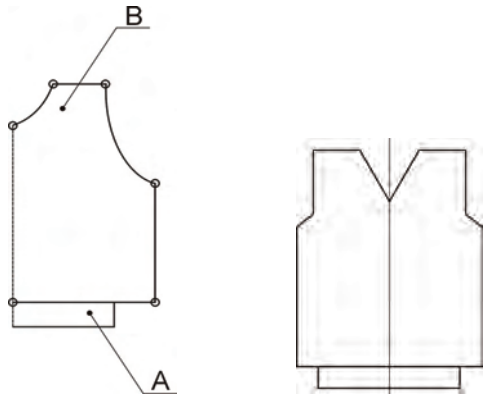


Fig. 1 (a,b). Examples of narrow sections

To choose a method for data representation and pertaining functions, the programmer needs to know the type and the characteristics of designed objects. Figure 1 (a,b) presents examples of knitting products narrow sections. The main difference between them is that figure 1a presents a shape of half symmetric part of the narrow section (cut) while figure 1b shows the whole cut. Section A is a margin – beginning part of the knitted detail. The basic knitted cuts have a margin, so it is a good idea to add it as an extra element. Moreover, the characteristics of this item have more technological aspect. It is obvious that the cuts can be described as a polygon formed by straight lines and curves (fig. 1). It can also be interpreted as a polyline but in general, lines formed knitting cut shape are straight lines and curves, as well [3], [4], [5].

There are two manners to represent a cut of knitting product

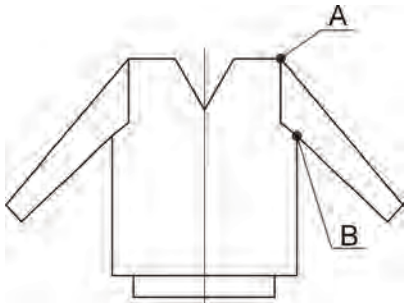


Fig. 2. Examples of a knitted product construction

(detail): as a set of lines or as a set of points. The advantage of the first one is that it is easy to construct a shape - line by line. However, it is necessary to check if the last point of previous line is equal to the first one of the next line. There is a redundancy of information. If the detail is represented as points, the redundancy will be avoided, but it makes some difficulties in line drawing and detail conversion into knitting rows and stitches. Knit products could be regarded as a set of details. Fig.2 shows an example of knitting products construction. A and B are points belonging to two details and they determine how to join the narrow sections. Such points fix a connection between details and determine that the distances between them for both details should be equal.

III. Lists as Dynamic Data Structures

A. Features of lists as data structures

Lists are very flexible dynamic data structures. Unlike arrays items may be added to them or delete from them at will. It means that an item can be added to the list or removed from the list easy nevertheless of its allocation [2],[7].

Another advantage of the lists as data structures is that a programmer need not worry about how many items a program will have to accommodate because these data structures are dynamic. The memory is allocated when a new item is created. Thus the number of items that may be added to a list is limited only by the amount of memory available. If an item must be removed from the list, its memory allocation is released. This allows us to write robust programs which require much less maintenance. In comparison with array lists are more flexible data structures because the programmer can allocate or release a memory for each list item independently but handling for the arrays is much easier than list handling.

In a linked list, each item is allocated space as it is added to the list. A link is kept with each item to the next item in the list. That way it is accepted for item of the list to be called node. Each node of the list has two elements - the item being stored in the list and a pointer to the next node in the list. The last node in the list contains a NULL pointer to indicate that it is the end or *tail* of the list. This is presented on fig. 3.

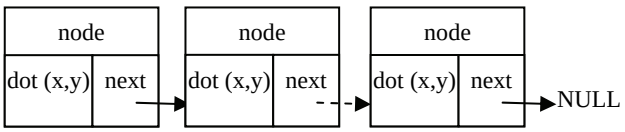


Fig. 3. Data structure **“linked list”**

This paper suggests an approach for particular lists structures application for descriptions of graphics objects. That way go on to describe lists structures instead common term “item”, the particular object part “dot” will be used. In this particular case dots are items of the lists.

There are some variations of the lists structures. Such variants are:

- Circuit Linked Lists
- Doubly Linked Lists
- Doubly Linked Circuit Lists

Circularly or Circuit Linked List is that the tail of the list is always pointing to the head – it means that the pointer of the last node points to the first node, which is the beginning of the list. Doubly linked lists have a pointer to the preceding item as well as one to the next. They permit scanning or searching of the list in both directions. To go backwards in a simple list, it is necessary to go back to the start and scan forwards. Doubly linked circuit list is combinations between doubly linked list

and circuit linked list also have a pointer to the preceding item as well as one to the next.

B. Handle for the lists

The variable (or *handle*) which represents the list is simply a pointer to the node at the *head* (beginning) of the list. Practically, it points to the first node of the list.

The most important operations, applied to the lists structures are: addition to the tail of the list, insert a new item to the list, and deleting an item from the list. The operation *"addition"* is treated as a manipulation to add a new item (node) to the tail of the list, after the last node. The term *"insert"* is used for adding a new item (node) regardless of its new positions. It means to "insert" node, exactly.

Figure 4 illustrates the operation *"insert a new node"* to the list. There are two variants of the operation "Insert before" and "Insert after" according to if the new item is inserted before the selected item or after it.

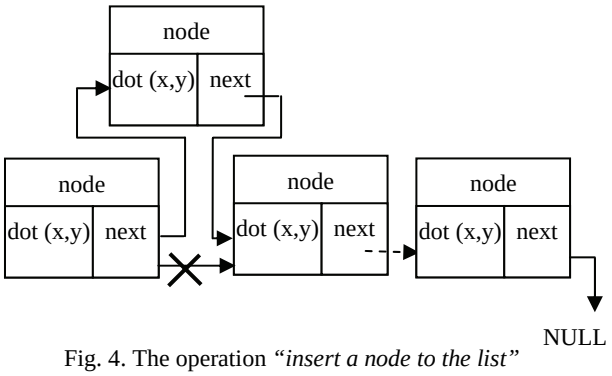


Fig. 4. The operation *"insert a node to the list"*

The simplest strategy (algorithm) for inserting an item to a list is to: allocate space for a new node; copy the item into it; make the new node's **next** pointer point to the next item of the list and make the current node's next point to the newly allocated node.

A delete operation is another important list handler. The simplest strategy (algorithm) to remove a node from the list is to make the previous node's next point to next node (item pointed by the next pointer of the node for deleting) and after that remove allocated memory for the node. This operation is illustrated on fig. 5.

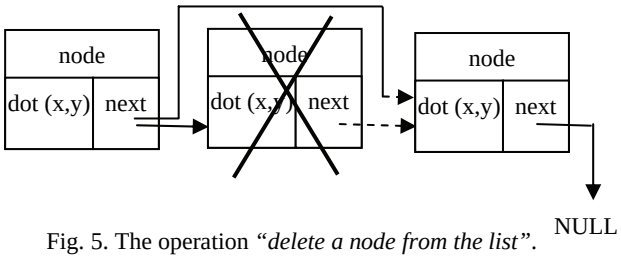


Fig. 5. The operation *"delete a node from the list"*.

IV. Applications of Listed Data Structures in Knitting Industry CAD Systems

There are some possibilities to describe a shape of the detail. The most popular are: arrays and listed structures. The former is the easier way for programming whereas the latter leads to better code efficiency and data representation. The data structure can be a linked list or a circuit doubly linked list. Because a detail described as straight lines and curves have to be converted into knitting rows and stitches, the application of the first kind of list or second one depends on the narrow section type [5]. If the narrow section is a graphic representation of a half symmetric part of knitting cut, it is better to use linked list. On the other hand, if the detail is a whole narrow section, the circuit doubly linked list is better decision.

As it already mentioned, the item of the list is a dot with x, y coordinates. The item can be described with an object of CPoint class. The declarations of structure **node** and class **list** are as follows:

```
struct node
{
        CPoint item;
        node *next;
};

class list
{
private:
        node *begin;
        node *current;

public:
        list();
        ~list();
        void Add(CPoint);
        void Insert(node *, CPoint);
        void Delete(node *);
        void DrawAll();
        void EraseAll();
        bool EmptyList();
        node* Search(CPoint);
        CPoint GetCurrentPoint();

};
```

Member-variable **begin** is a pointer to the node at the head of the list; data member **current** points to current dot position. The last one is used as returned value of **GetCurrentPoint()** method. This value is applied in **OnDraw()** method to draw dots and the contour of the detail. Methods of the class are described on Table 1. Coding of methods **Insert()** and **Delete()** follows the diagrams on fig. 4 and fig. 5, respectively.

Construction is an MFC application use list class. The main window of the application is represented on fig. 6. According to the active mode, the operator can add, or insert a dot to the detail construction or delete (remove) a dot from the construction by left mouse button click. These functions could be started using menu item **Draw**. It consist also menu item **Erase All** called **EraseAll()** method. Because the class

list destructor needs to delete all list's items, it calls **EraseAll()** method, too.

TABLE I
METHODS OF CLASS LIST

method	description
list();	constructor
~list();	destructor
void Add(CPoint);	Add a new dot to the end of the list
*void Intert(node *, CPoint);*	Insert a new dot after pointed dot
*void Delete(node *);*	Delete a dot
void EraseAll();	Erase all dots
bool EmptyList();	Check if the list is empty
node Search(dot);*	Search a dot
CPoint GetCurrentPoint();	Return Current dot

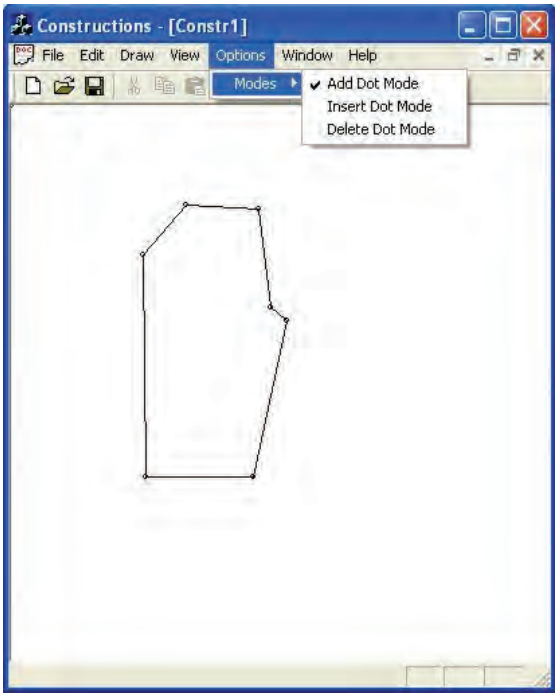


Fig. 6. Main window of the application

V. CONCLUSION

This paper suggests an approach to data representation in CAD system development based on a set of programming modules. Some of these modules can be used in next generation of CAD systems or they can be embedded in another system - CAD software. The main reason to create such project is that it can be used as software application from fashion designers to make boutique collections of knitting products. Listed data structures are successfully used to code graphic data related to knitting pattern object. Knitting patterns are represented as sets of listed dots, so it is the reason to use listed data structure to describe graphically knitting pattern object.

The goal of the paper is to give an approach of usage of listed structures in CAD systems' data representation. In this particular system the structures linked list is used. The data type does not allow using listed structures in MS VC++ collections. That is the reason to create listed structures related to knitting pattern graphic objects. The algorithms operated with list structures are used. These algorithms are applied to particular software module "Construction". The application can be embedded in whole CAD/CAM system in knitting industry or it can be used as software application creating patterns of knitting products. The MS VC++ and MFC are used to create a software application. The main reason to create such project is that it can be used as software application from fashion designers to make boutique collections of knitting products. The project can be used as an example for CAD system development.

REFERENCES

[1] Angelova Y., E. Zaharieva-Stoyanova, Precision of the shape of garment pieces at Fully Fashion Knitting, *Textile & Cloths*, 7/2006.

[2] Baczynski M., Data Structures: Linked List.

[3] Zaharieva-Stoyanova, Computer Aided Design of Knitting Products Using FF Method, *International Scientific Conference UNITECH'2007*, 23-24 November, Gabrovo, Bulgaria.

[4] Zaharieva-Stoyanova E., Methods of Graphic Representation of Curves in CAD systems in Knitting Industry, *XLII International Scientific Conference on Information, Comunications and Energy Systems and Technologies ICEST 2007*, 24-27 June, 2007, Ohrid, Macedonia.

[5] Zaharieva-Stoyanova E., Application of Bezier Curves in Knitting Industry CAD/CAM systems, *XL International Scientific Conference on Information, Communication and Energy Systems and Technologies – ICEST 2005*, June 29 – July 1, Nis, Serbvia and Montenegro, 2005.

[6] Zaharieva-Stoyanova E., Application of Bezier Curves in Knitting Industry CAD/CAM systems, *XL International Scientific Conference on Information, Communication and Energy Systems and Technologies – ICEST 2005*, June 29 – July 1, Nis, Serbvia and Montenegro, 2005.

[7] Zaharieva-Stoyanova E., *Algorithm for Computer Aided Design Curve Shape Form Generation of Knitting Patterns, International Conference on Automation, Quality and Testing, Robots AQTR 2006(THETA 15)*, May 25-28 2006, Cluj-Napoca, Romania.

[8] Data Structures and Algorithms, http://www.cs.auckland.ac.nz/software/AlgAnim/lists.html

[9] *Trend Collection 2004/2005*, Stoll, Department Fashion & Technology

[10] Sirix 110/210, Instructions, Part I, II, Reutlingen, 2001.

[11] DeSignaKnit, Demo, www.knitcraf.com.

[12] Stitch&MotifMaker, www.softwareknitting.com

[13] SweaterWizard2001, Demo, www.softwareknitting.com

[14] WiseTex, Demo, www.mtm.kuleuven.oc.be

A Fast Template Matching Algorithm

Yulka Petkova[1], Nuri Nuri[2] and Milena Karova[3]

Abstract – **Fast template localization in an image has ever been a great challenge to the researchers. The here presented algorithm uses a combination of different approaches for run-time acceleration including computations over extracted sets of representative points, fast elimination of unpromising positions and parallel computations. Experimental results prove the reliability and precision of localization and a significant run-time acceleration comparing with the classic brute-force algorithms for template matching.**

Keywords – **Template Matching, Template Localization, Image Processing, Representative Points, Parallel Computing**

I. INTRODUCTION

One of the most frequently solved tasks in the field of image processing is the task for template matching. It consists of finding approximately or fully coinsidence of the template T^M in the image I^N, i.e. coinsidence with any of sub-images I_i^M, as it is shown in fig. 1.

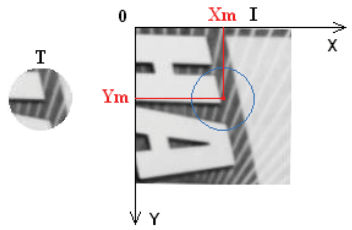


Fig. 1. Illustration of template matching problem

A wide variety of applications need the solution of the pointed problem. Some of them are:

- Fiducial recognition for PCB assembly by pick-and-place robot;
- Registration of alignment marks on printed material to be inspected;
- Robot vision guidance, locating objects on conveyor belts, pallets and trays;
- Template (model) based object recognition, etc.

The problem of template matching is connected with finding of any matching measure and algorithm to evaluate the extent of coincidence of the template with the correspondent part of the image. A great number of matching criteria and searching algorithms are published [4]. One of the basic purposes is the high reliability - the matching criterion has to cope successfully with any type of distortions – translation, rotation, scaling, nonlinear changes in the intensity, etc. The other important goal is the real time processing. The second goal is harder to achieve than the first one, because of the great data sets, which are to be processed.

Some of ways to speed-up the computations for template localization in a greater image are: making computations in the frequency domain [5]; searching around the most probable position [6]; reducing the number of sub-windows for making computations [4, 6]; skipping the unpromising positions [1]; reducing sets of points taking part in comparisons [4, 6]; using pyramids of data for image and template representation [2, 4], etc.

An alternative approach, to speed-up the computations for template localization, is to use effectively the hardware capacity of the modern computers, especially on systems with multicore central processing units (CPUs). This approach is based on scalable parallel computations which will guarantee effective usage of computer hardware.

In this paper we propose to combine the usage of different optimization techniques in order to increase the speed of template localization process in a greater image. We combine:

- usage of extracted sets of representative points;
- fast elimination of unpromising positions;
- usage of parallel computations.

As a result of this combination the run-time of template localization is highly accelerated. The precision and reliability of the localization are preserved.

II. SETS OF REPRESENTATIVE POINTS

One of the ways for speeding-up the algorithms is to use sets of representative points and compute the similarity measures taking intensities of only these points. A great variety of methods for representative point extraction are using in practice. They use different criteria for making selection. For the goals of our study, we have used three types of representative points:

- Points, selected by Sobel edge detector are the first type;
- The second type are selected, using the earlier proposed by us TyPe edge detector [7]. It extracts edge points, determined by the points of local extremes of the second derivative of the intensity function according to the following criteria:

$$If \quad \left| \underset{i}{extr}\left(\frac{d^2 I(x)}{dx^2} \right) \right| > \theta, \quad then \quad i \in E, \qquad (1)$$

where $I(x)$ is the intensity function in a row or in a column of the image, θ is a threshold. E is the set of extracted edge points.

[1]Yulka Petkova is with the Department of Computer Sciences and Technologies at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: jppet@abv.bg.

[2]Nuri Nuri is with the Department of Computer Sciences and Technologies at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: nuri.ismail@gmail.bg.

[3]Milena Karova is with the Department of Computer Sciences and Technologies at Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: milena.karova@gmail.com.

- The third type of points are selected, using another earlier proposed by us method for Equipotential planes [8]. We interpret the image as a three-dimensional object and we propose the choice to be made by equipotential planes, which are parallel to the plane *xOy* and which cut the image (relief) on proper intensity levels. Thus, the extracted points outline the horizontal contours of the local "hollows" and "hills" from the three-dimensional image profile.

We define the following rule (Eq. 2) to extract representative points:

$$\begin{aligned} If \Bigg(& \Big(\big[(PI(i,j) \le Pt_{max}) \cap (PI(i+1,j) > Pt_{max})\big] \cup \\ & \cup \big[(PI(i,j) \ge Pt_{max}) \cap (PI(i+1,j) < Pt_{max})\big] \Big) \cup \\ \cup & \Big(\big[(PI(i,j) \le Pt_{min}) \cap (PI(i+1,j) > Pt_{min})\big] \cup \\ & \cup \big[(PI(i,j) \ge Pt_{min}) \cap (PI(i+1,j) < Pt_{min})\big] \Big) \cup \\ \cup & \Big(\big[(PI(i,j) \le Pt_{max}) \cap (PI(i,j+1) > Pt_{max})\big] \cup \\ & \cup \big[(PI(i,j) \ge Pt_{max}) \cap (PI(i,j+1) < Pt_{max})\big] \Big) \cup \\ \cup & \Big(\big[(PI(i,j) \le Pt_{min}) \cap (PI(i,j+1) > Pt_{min})\big] \cup \\ & \cup \big[(PI(i,j) \ge Pt_{min}) \cap (PI(i,j+1) < Pt_{min})\big] \Big) \Bigg) \ Then \\ & (i,j) \in EP \end{aligned} \quad (2)$$

In this equation $PI(i,j)$ is the intensity at point (i,j); Pt_{min} and Pt_{max} are intensity thresholds - low and high respectively; EP is the extracted set of representative points.

We use the three-dimensional image profile only to extract coordinates of representative points. In the next computations (for template localization) these points take part with their intensities.

III. Similarity Measures

Similarity measures are also very important part of the localization acceleration process because their computational complexity reflects directly on the speed of localization. We focus on two kinds of similarity measures because of their high reliability and low computational complexity:

- Normalized Cross Correlation (NCC);
- M-estimators combined with "sum of robust differences" (SRD).

For the goals of the study, NCC has been computed over sets of representative points:

$$NCC(x,y) = \frac{\sum_{q=1}^{P} I(x+i_q, y+j_q).T(i_q, j_q)}{\sqrt{\sum_{q=1}^{P} \left(I\left(x+i_q, y+j_q\right)\right)^2} \cdot \sqrt{\sum_{q=1}^{P} \left(T\left(i_q, j_q\right)\right)^2}} \quad (5)$$

where $T\left(i_q, j_q\right)$ for $q = 1,2,\ldots,P$ is the intensity at q-th representative point of the template point with coordinates (i_q,j_q), $I\left(x+i_q, y+j_q\right)$ is the intensity at the corresponding point of the image.

A lot of M-estimators have been discussed in literature - Huber's, Tukey's, Geman's and McClure's, Lorentzian's, etc. [1], [3]. We have investigated all of them, but the best results we have received using the first two. They also have been computed over sets of representative points.

In the following equations ρ is the robust error measure, r is the difference between intensities of compared points and σ is a threshold, which can vary from 0 to 255.

- Huber's estimator:

$$\rho_1(r,\sigma) = \begin{cases} \dfrac{r^2}{2} & r \le \sigma \\ \sigma\left(r - \dfrac{\sigma}{2}\right) & r > \sigma \end{cases} \quad (6)$$

- Tukey's estimator:

$$\rho_2(r,\sigma) = \begin{cases} \dfrac{\sigma^2}{6}\left[1 - \left(1 - \left(\dfrac{r}{\sigma}\right)^2\right)^3\right] & r \le \sigma \\ \dfrac{\sigma^2}{6} & r > \sigma \end{cases} \quad (7)$$

The similarity measure is "sum of robust differences - SRD", as it is in [1]:

$$SRD_{\rho,\sigma}(T,I) = \sum_{i=0}^{n}\sum_{j=0}^{n} \rho\left(r,\sigma\right) \quad (8)$$

IV. Search Strategy

For additional acceleration of the localization we use a search strategy based on a regular set of initial positions of the template in the image and consecutive search around the most promising position [9]. This strategy belongs to the so called "coarse-to-fine" searching algorithms and guarantees a significant run-time acceleration comparing with the classic brute-force algorithms.

V. Parallel Computing

In recent years parallel computing has become a dominant paradigm in computer architecture, mainly in form of multicore processors. In order to use the advantages of these processors we have implemented parallel versions of the localization algorithms, using a cross platform multithreading library designed and implemented by us. This library is designed as general purpose library and is used in template localization algorithms at different levels:

- for parallelization of single template localization (parallel localization);
- for parallelization of multiple localizations (parallel experiments).

The strategy used for parallelization of single localization is shown in Fig.2.

The first step is to detect the number of processors (cores) for the current system on which the program is running. If this number is N, then we divide the image in N sectors and create a worker thread for each sector. In the parallel part of the algorithm each worker thread executes computations for template matching for its own sector and finds the best match. After that these results are analyzed and compared in order to find the best candidate for template position in the image.

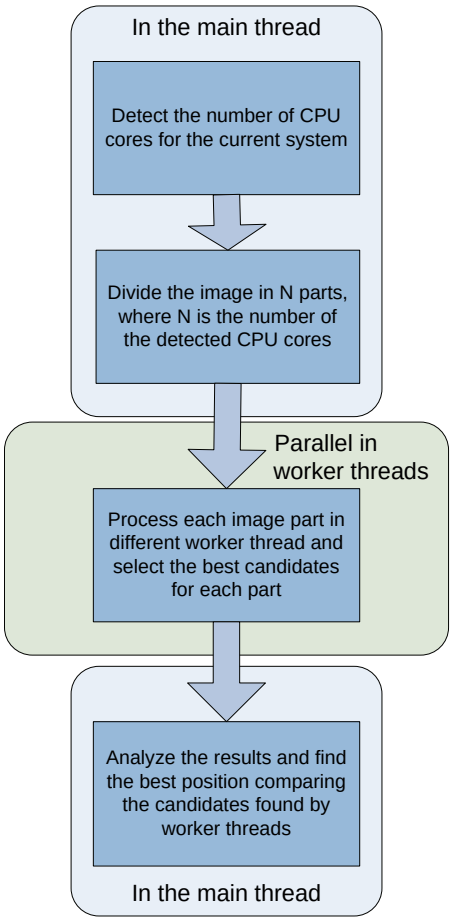


Fig.2. Parallel single template localization algorithm

The strategy used for parallelization of multiple localizations on same image or on different images is very similar to the previous presented parallel single template localization algorithm. The most significant difference is that for multiple localizations we group pending localization tasks in order to distribute them uniformly on worker threads. The number of worker threads is equal to the number of processors (cores) in the system and each worker thread executes one or more localization tasks. Each thread keeps the results for further analysis and registration. The illustration of this strategy is shown in Fig.3.

We successfully combine both of these parallelization strategies with the algorithms presented in previous sections and get significant run-time acceleration.

VI. Experimental results

Series of experiments have been made to estimate proposed approach for fast template localization. These experiments include using different sets of representative points, different template shapes and dimensions, different searching strategies and parallel computing techniques as well. All the series include 100 experiments. A short summary of these experiments is shown in the next tables.

- **Localization reliability in translated and noised images**

The experiments are provided for different sets of representative points, using fast elimination of unpromising positions and parallel computing techniques, presented in previous sections.

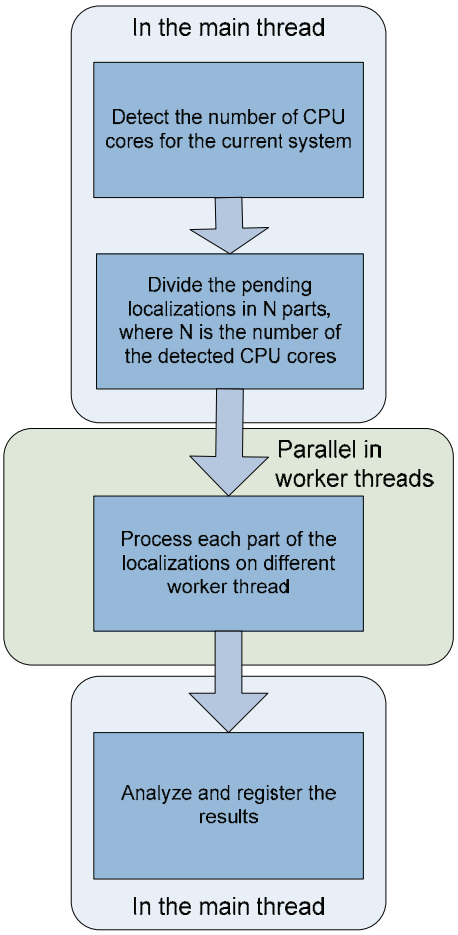


Fig.3. Parallel algorithm for multiple localizations

The results in Table I present the successive template localizations in the image in presence of noise. Table II is about the precision of template matching in presence of noise and present mean square error values for different noise levels. Both of these experiments have been designed and implemented on same conditions.

Experimental results confirm the high reliability of the template localization using the proposed approach of combining the different acceleration techniques with parallel computing.

Table I
Reliability in the presence of noise [%]

Noise [%]	5	10	30	50	70
M-estimators with Equipotential Points	100	100	100	99	99
M-estimators with TyPe edge points	100	100	100	100	99
M-estimators with Sobel edge points	100	100	100	99	98
NCC over all points	100	100	100	100	100

Table II
TEMPLATE MATCHING IN THE PRESENCE OF NOISE - PRECISION

Mean Square Error [*10^{-6}]

Noise [%]	5	10	30	50	70
M-estimators with Equipotential Points	0,02	0,02	0,02	0,03	0,05
M-estimators with TyPe edge points	0,02	0,02	0,02	0,04	0,06
M-estimators with Sobel edge points	0,02	0,02	0,02	0,04	0,06
NCC over all points	0,006	0,01	0,01	0,02	0,02

- **Run-time optimization**

All of the presented experiments in this section have been made on the following PC configuration:

- OS: Microsoft Windows XP Professional SP2;
- CPU: Intel Core 2 Quad Q6600, 2400 MHz (9x267);
- System Memory: 2 x 1 GB DDR2-667 MHz;
- VGA: Intel(R) Q33 Express Chipset Family (128 MB);

The results in Table III present the run-time comparison of classical template localization technique and the combinative parallel acceleration approach presented in this paper.

Table III
RUN-TIME COMPARISON

	Searching algorithm		**brute-force**	**coarse-to-fine**
Total duration [ms]	**Sequential**	**M-Estimators**	10078	844
		Correlation	11060	1078
	Parallel	**M-Estimators**	2880	275
		Correlation	3210	305

The processor of the test system is multicore (4 cores) and the usage of parallel algorithm results in great performance improvement (at about 350%) for both classical brute-force and accelerated coarse-to-fine search strategies.

The next images show the CPU usage for single-threaded and for multi-threaded parallel algorithm on the test system.

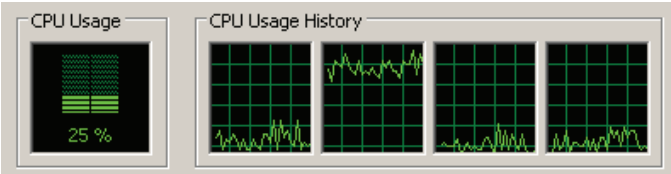


Fig.4. CPU usage for sequential single-threaded algorithm

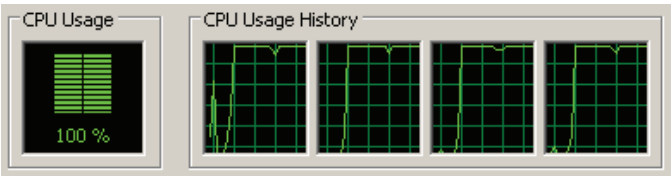


Fig.5. CPU usage for parallel multi-threaded algorithm

It is seen in the images that the parallel computing techniques guarantee effective usage of computer hardware.

VII. CONCLUSION

To conclude we can point out the advantages of our approach:

- It uses a combination of different strategies in order to decrease the computational complexity of template matching.
- It uses parallel computing techniques for effective usage of computer hardware.
- Decreasing the computational complexity and using parallel versions of the algorithms does not lead to significant reliability and precision decrease.

We can generalize that the received results are good enough, and we can use the proposed method to solve tasks for template matching in applications for which the received precision is satisfactory.

ACKNOLEDGEMENTS

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

REFERENCES

[1] J-H. Chen, C-S. Chen, Y-S. Chen, “Fast Algorithm for Robust Template Matching with M-estimators”, IEEE Transactions on Signal Processing, Vol. 51, Issue 1, Jan. 2003, pp. 230 - 243

[2] R. Dahyot, P. Charbonnier, F. Heitz, “Robust Visual Recognition of Colour Images“, CVPR-2000 International conference on Computer Vision and Pattern Recognition - pp. 685-690 - Hilton Head Island, USA

[3] K.V. Arya, P. Gupta, P.K. Kalra, P. Mitra, “Image registration using robust M-estimators”, Pattern Recognition Letters 28, pp. 1957–1968, 2007

[4] B. Zitova, J. Flusser, “Image Registration Methods: A Survey”, Image and Vision Computing, 21, pp. 977 – 1000, 2003

[5] S. Yoshimura, T. Kanade, “Fast Template Matching Based on the Normalized Correlation by Using Multiresolution Eigenimages”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No 12, pp. 2086-2093, 1993

[6] Lee Chang-Hsing, Ling-Hwei Chen, “A Fast Motion Estimation Algorithm Based on the Block Sum Pyramid”, IEEE Transactions on Image Processing, Vol. 6, No11, pp. 1587-1591, 1997

[7] D. S. Tyanev, Y. P. Petkova, “About the Possibilities of a New Edge Definition”, 28th ISSE 2005, 19th – 22nd May 2005, Wiener Neustadt, Austria, IEEE Catalog Number 05EX1142C, ISBN: 0-7803-9325-2, pp. 273 - 378

[8] D. Tyanev, Y. P. Petkova, “A New Method for Important Points Extraction”, International Conference on Computer Systems and Technologies CompSysTech’2004, 17th – 18th June 2004, Rousse, Bulgaria, ISBN 954-9641-384, pp. IIIA.11-1 – IIIA11-8

[9] Y. Petkova, N. Nuri, “A Strategy for Fast Template Localization in an Image”, Fourth International Conference “Computer Science” - 2008, Kavala, Greece, Sep. 2008, pp. 675 - 680.

Cost Effective 3D Architectural Visualisation

Igor I. Nedelkovski[1] and Vladimir S. Gjorgjieski[2]

***Abstract* – Visualization is any technique for creating images, diagrams, or animations to communicate a message. Visualization through visual imagery has been an effective way to communicate both abstract and concrete ideas since the dawn of man. Architectural Visualization involves visualization software technology for the viewing and manipulation of 3D models, technical drawing and other related documentation of architectural buildings. Architectural visualization software typically provides high levels of photorealism so that a building can be viewed before it is actually manufactured.**

***Keywords* – Visualization, 3D modeling, Analysis, 3D laser scanning, Photogrametry, Hybrid method.**

I. INTRODUCTION

The primary goal of research in the field of 3D graphics is exact representation of the 3D real world. One of the possibilities for precise 3D representation of the real world is to create an interactive 3D model where the user could move into a simulated environment. This feature of 3D graphics software is used in many areas, from architecture, automotive, and ending with marketing and 3D catalogs of products.

For architectural visualization, is important to say that computer graphics is widely used in architectural visualization. First, for better presentation of new projects, then the restoration of old buildings and ruins, in the area of spatial planning, interior elements, the very arrangement of the interior or exterior of the building and ultimately of course for leaving a better impression on investor in the facility or the public.

But despite all these benefits allows our 3D architectural visualization, it must be noted (highlighted) that this is usually an expensive operation, requiring construction of complex models with many details, high quality rendering, long time, and expensive hardware.

Also classical architectural visualization is based on movement along a path, without the possibility of free virtual tours and interaction with the user.

[1] Igor Nedelkovski is with the Faculty of Technical Sciences,St.Kliment Ohridski University - Bitola, address: ul. Ivo Ribar Lola,7000 Bitola, Macedonia, e-mail: igor.nedelkovski@uklo.edu.mk.

[2] Vladimir Gjorgjieski is with the Faculty of Technical Sciences,St.Kliment Ohridski University - Bitola, address: ul. Ivo Ribar Lola,7000 Bitola, Macedonia, e-mail: vladimir.gjorgjieski@gmail.com

Therefore aim of this paper is analysis of existing methods for 3D architectural visualization, and demonstration of hybrid methods and tools that will satisfy the quality of the finished product (3D interactive visualization of architectural structure), which will be cheap, simple and will not require much time.

II. METHODS FOR 3D ARCHITECTURAL VISUALISATION

In the world of architectural visualization, there are three standard methods, they are:

1. 3D laser scanning
2. Photogrametry, and
3. 3D modeling.

Each of these methods has its field of application, has its advantages and disadvantages. In this section we explain the methods with examples of models made by each of the methods. Also in this section we analyze advanteges (benefits) and disadvantages of these methods.

A. Method of 3D Laser scanning

3D scanner is a device that analyzes the real object or the natural environment in order to gather data on its shape and possibly for his outward (exterior) appearance. The collected information are used for the construction of digital 3D models that have broad application. These devices are used mostly in the entertainment industry in the production of films and video games. Another similar application of this technology include industrial design, prothetics, and of course most in architecture for testing, documentation, and reproduction of objects of cultural heritage.

The method for obtaining the 3D model using 3D scanner comprises three stages, as follows:

1. Scanning (often multiple)
2. Aligment, and
3. Reconstruction

In most situations, the singular scan does not produce a complete model of the scanned object. Often multiple scans are needed, often hundreds, of different directions to obtain full information about the scanned object. This information should be brought into a common coordinate system and then be drawn in order to obtain a complete model. This process is usually called alignment.

A classic example of this is digitalization of the temple of Angkor in Cambodia and it is of exceptional importance. An example of this visualization is given in the Figure 1.

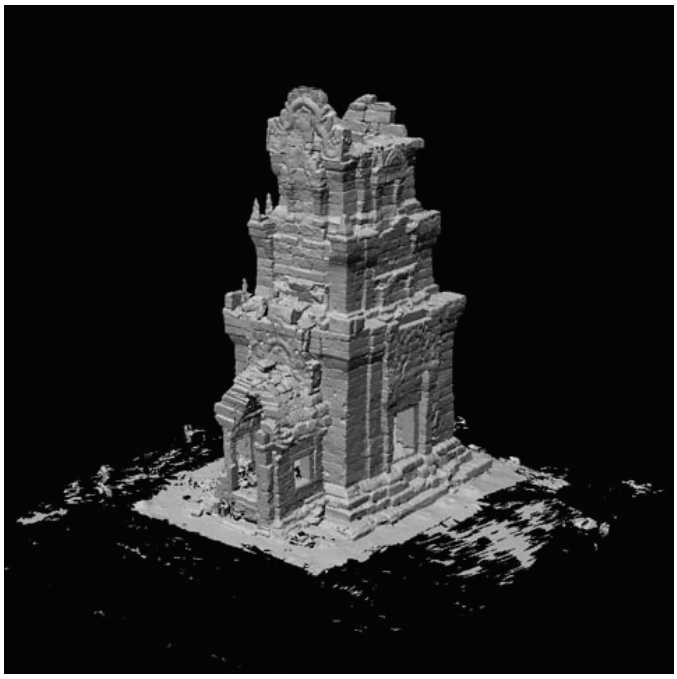


Fig. 1. Digitalization of the temple of Angkor in Cambodia with method of 3D laser scanning.

The largest advantage of this method is precision with which the final model is obtained. So its use is limited in projects of digitalization of cultural heritage and the model then is used for further examinations. But despite the high quality models that are obtained using this method should be noted that the method is very expensive and also the equipment. Average 3D scanner costs about U.S. $ 30 000, plus the method is expensive, requires engagement of many people who also need to pay. So if the use of that model is not for scientific research purposes, use of this method is unnecessary.

B. Method of Photogrametry

Photogrametry is the first technology in which geometric properties of objects are determined from the photos. The principle of photogrametry can simply be explained as follows, the distance between two points which lie on a plane parallel to the plane of the photograph can be determined by measuring their distance if scaling of the photograph is known.

More sophisticated technique is stereogrametry. With this technique it is possible to obtain 3D coordinates of object points. They are determined by measurements made on two or more photos taken from different positions. The common points are identified on each image. The line of sight (or ray) can be constructed from the location of the camera to the point of the object. The intersection of three beams determines the exact location of the point in 3D coordinates. For this purpose are used very sophisticated algorithms that can determine the 3D coordinates of the point from just one photograph.

Photogrametry process can be divided into three basic stages as follows:

1. Taking pictures of the facility,
2. Metrology (measurement object), and
3. Reconstruction of the object (obtaining the model).

Photographing the facility is the first stage of this method. There are two ways of photographing Aerial photogrametry, and Close-range photogrametry. These methods have their use depending on field of use.

Measurement is an integral part of photogrametry. Namely this is the phase in which are obtained real dimensions of the object from the images. So the use of more photos from different sides are recomended. Then the rays are projected and we determine the common points of the images, from wich we can reconstruct the required model.

This method is often used in architectural visualization, especially if it is a digitalization of existing facilities. Compared to the 3D scanning, models are not as detailed but would satisfy the needs of a digital tour guide or a 3D map for a city. It should also be noted that Close-range photogrametry is often used in hybrid technologies. classic example of photogrametry is visualization of North-German castles, figure 2.

Fig. 2. Digitalization of the North-German castles with method of Photogrametry.

Unlike the method of 3D scanning, Photogrametry is used often due to economic reasons. In terms of quality, models are simple and could not be used for scientific purposes as was the case with 3D scanning. Therefore, this method is widely used in hybrid technologies. There is a technology based on photogrametry and is called Image - based modeling (modeling based on the picture) which is often used in hybrid methods. As disadvantage of this technology can be mentioned expensive computer hardware and software, and also aerial photogrametry is very expencive As for photography equipment should say that the professional photographic equipment does not come in cheap products. Another important problem that arises in this method is digitalization of art and natural forms (such as decorations on walls with images of animals, people, plants, irregular shapes, etc.). However photogrametry is often used for architectural visualization and gives satisfaying results.

C. *Method of 3D modeling*

In 3D computer graphics, 3D modeling is the process of creating mathematical (string) representation of any three-dimensional object, using special software. Popular softwares in the world of 3D modeling are 3D `s MAX, Maya, Blender and so on. The product is called a 3D model. It can be shown as two-dimensional image through a process called rendering, or be used for computer simulation of physical phenomena. The model can also be physically created with device called 3D printer.

Although programs for 3D geometric modeling may seem very complex, mainly any modeling is based on six tasks (steps):

1. Creation of simple elements of the model,
2. Their assembly into complex structures,
3. Order of objects in 3D scene,
4. Selecting materialsAdjust the lights,
5. Selecting a views and methods for rendering.

In the architectural visualization 3D modeling is widely used. We should note that these models usually do not represent real objects with their real dimensions, but are widely used in many fields.

3D modeling is often used for the design of new buildings and their presentation to the public, which was not the case with the previous techniques. In figure 3 is given example of architectural 3D visualisation made with this method.

Fig. 3. Example of architectural visualization of a new facility developed in 3D `s MAX.

The method of 3D modeling is the most popular method in the world of architectural visualization, because of it`s economic feasibility. Models have high quality. They are very widely used in industry with games, also in the film industry. Disadvantage of this method is long time of modeling, of course depending on complexity of the model. It`s not excluded the application of 3D modeling in hybrid methods.

III. HYBRID METHOD FOR 3D ARCHITECTURAL VISUALISATION

Hybrid method involves using several methods and software that are simple to learn, require low or average hardware, and of course are cheap or free.

For this method we use combination of previous methods (3d modeling, and close range photogrametry). It is a mixture of three cost free softwares. Google Sketch Up for polygonal modeling, Gimp for image processing and SUBDO, software used for interactivity, lights and shadows of the model that allows real-time rendering.

This hybrid method consists of the following steps:

1. Taking pictures of the model
2. Measurement
3. Selection of photos
4. Processing of the images,
5. Modeling from photos,
6. Texture mapping,
7. Render in real time and interactivity.

Below in figures 4, 5, and 6 are given steps of this method.

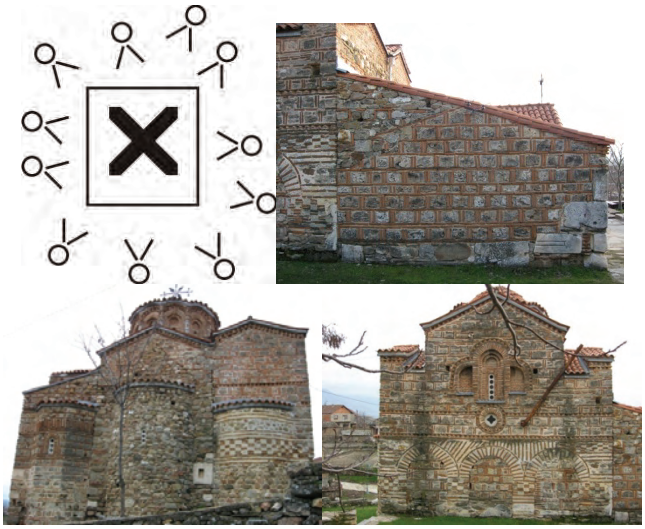


Fig. 4. Taking pictures of the model (close range photogrametry).

Fig. 5. Processed photos.

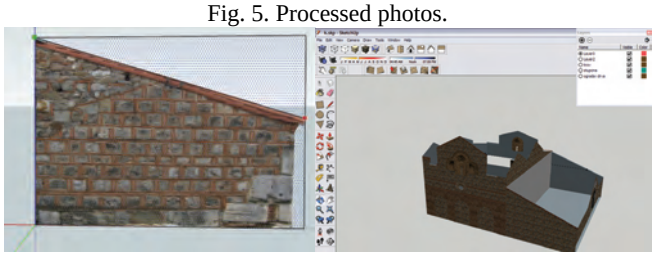


Fig. 6. Modeling.

After modeling follows the phase of setting lights, and interactivity in SUBDO. In the end of the process we choose the format and export the interactive architectural 3D visualization.

The hybrid method is cheap, needs short time for developing. Disadvantages of this method is deficit of natural effects. Quality of the interactive 3D architectural visualization developed with this hybrid method can`t be compared with traditional methods like 3D modeling or 3D scanning, but it`s on the same level with other very expensive hybrid technologies. This method can be used in spatial planning, also for presentation of new architectural designs, documenting of architectural heritage, presenting the tourist attractions on the Internet and multimedia presentations, navigation systems and so on.

IV. Conclusion

3D architectural visualization is process for creating models of existing or new architectural structures with different methods. There are few conventional methods for 3D architectural visualization depending for what purpose is visualization you can use different methods and software.

In this paper we analyze existing methods with their advantages and disadvantages. Also we demonstrated low cost hybrid method for creating satisfying 3D architectural visualization that can be used in many fields. Further research will continue toward improving the real time rendering, exploring new combinations of software, and application of this method in navigation systems.

References

[1] Anne Morgan Spalter, The Computer in the Visual Arts.Adision – Weesley, 1999.
[2] James D.Foley, Andries van Dam, Steven K.Feiner, John F. Huges, Richard L. Phillips, Introduction to Computer Graphics. Adison – Weesley, 1993.
[3] Malcom Kesson, An Introduction to 3D Computer Graphics. Malcom Kesson, 1995.
[4] Igor Nedelkovski, Computer Graphics, 3D Modeling and Animation. Faculty of Technical Sciences - Bitola 2008.
[6] Blender 3D: Architecture, Buildings, and Scenery (Create photorealistic 3D architectural visualizations of buildings, interiors, and environmental scenery), Allan Brito, 2008.
[7] http://www.architectural-media.com
[8] http://subdo.com
[9] http://sketchup.google.com
[10]http://3dwalkthrough.blogspot.com
[11]http://www.vanishingpointstudio.com/news
[12]http://www.light-up.co.uk
[13] http://www.cybercity3d.com/
[14] http://www.sketchup4architect.com
[15] http://sketchuptips.blogspot.com

Creating Panoramas and Eliminating Ghosting

Svetlana P.Vrskova[1], Igor I. Nedelkovski[2]

Abstract – **A panorama is any wide-angle view or representation of a physical space, whether in painting, drawing, photography, film/video, or a three-dimensional model. Panoramic photography soon came to displace painting as the most common method for creating wide views. As panoramic photography becomes increasingly popular, there is a greater need for high-quality software to automatically create panoramic images. Existing algorithms either produce a rough "stitch" that cannot deal with common artefacts, or require user input. Our first contribution is a method for dealing with objects that move between different views of a dynamic scene. If such moving objects are left in, they will appear blurry and "ghosted".**

This paper gives analysis of methods for creating panoramas and resolve problems during image stitching using blending algorithm.

Keywords – **Panoramas, image stitching, image blending, "ghosting".**

INTRODUCTION

A panorama is a picture that is made by combining a series of photos into one large picture. By combining a series of photos, you can provide a complete view of an area or location that cannot fit in a single shot. Panoramas are usually made by taking a series of photos across a horizontal (or, at times, vertical) line. The word "panorama" indicates an unbroken view, since a panorama can be either a series of photographs or panning video footage.

One of the problems in automatic image stitching is that of de-ghosting. When the images are taken, there is no guarantee that objects in the image stay stationary from one image to the next. This becomes a problem in the areas of overlap between images. When the images are stitched together, we take a composite of the overlapping images in order to create a smooth transition between neighbors. However, if regions of the scene are not stationary, the overlap image will be slightly different in each image contributing to the overlap. Thus, those regions of the composited image will contain combinations of pixel values from entirely different parts of the scene. For example, if a person moves his head in an area of overlap, the region containing his head in the stitched image will be a combination of head and background from both images. It will give the head a ghosted look, not to mention that this ghosted head will appear in two locations.

II. METHODS OF CREATION PANORAMAS

There are many methods of creation panoramas. You can produce any type of panoramas with any type of camera, digital or analog. Depending on the lens you usually stitch 8-16 images for a 360º panorama. Wide angle lenses with coverage of 60 degrees or more are usually used. The images are taken with 20%-50% overlapping. Important for all panoramas: [1]Use manual mode with same exposure for all images. [2]If possible do not use automatic white balance. [3]Use a tripod with a special panohead which rotates the camera around the nodalpoint.

Fisheye lenses are popular for spherical panoramas as you can cover a full view with only 3-8 images.To be able to make a full spherical with 3 images you need a full circle of min 180º. This can be done by using a 35mm film camera or a fullframe digital and an 8 mm fisheye.

Fullframe Fisheye lenses usually 15 - 16 mm has been used for many years for panoramas with 35 mm cameras. The quality is much better than with the 8 mm fisheyes but they need 6-8 images around + top and bottom images.

One shot 360º panoramic solutions are parabolic mirrors which can be used as an ad on to many digital cameras. The resolution is low and the quality is also affected by the quality of the mirror.

Digital panoramic scanning cameras can make very high quality panoramas but the price is also high.

One popular method of creation is to stitch photographs taken from the same point in space but of varying pitch and yaw in order to create a panoramic image which, when rendered using appropriate software (usually through web browser plugins), allows the end user to pan and zoom at will.

The benefit of this method is that it does not require any specialized equipment to capture the images, although using such can greatly speed up the process and render a higher quality result. This photographic technique results in a limited depth of field, meaning that the space may appear warped.

Image stitching or photo stitching is the process of combining multiple photographic images with overlapping fields of view to produce a segmented panorama or high-resolution image. For image segments that have been taken from the same point in space, stitched images can be arranged using one of these graphical projections:

Rectilinear projection, where the stitched image is viewed on a 2D plane.

[1]Svetlana Vrshkova is with the Faculty of Technical Sciences, St.Kliment Ohridski University - Bitola, address: ul. Ivo Ribar Lola bb, 7000 Bitola, Macedonia, e-mail: svetlanavrskova@yahoo.com.
[2]Igor Nedelkovski is with the Faculty of Technical Sciences, St.Kliment Ohridski University - Bitola, address: ul. Ivo Ribar Lola, 7000 Bitola, Macedonia, e-mail: igor.nedelkovski@uklo.edu.mk.

Cylindrical projection, where the stitched image shows a 360° horizontal field of view and a limited vertical field of view. Panoramas in this projection are meant to be viewed as though the image is wrapped into a cylinder and viewed from within. When viewed on a 2D plane, horizontal lines appear curved while vertical lines remain straight.

Spherical projection, where the stitched image shows a 360° horizontal by 180° vertical field of view. Panoramas in this projection are meant to be viewed as though the image is wrapped into a sphere and viewed from within. When viewed on a 2D plane, horizontal lines appear curved as in a cylindrical projection, while vertical lines curve as they get closer to the poles of the sphere.

The workflow of this method of creation panoramas with stitching photographs taken from the same point in space but of varying pitch and yaw includes:

[1] Browsing folder that contains the photos taken from the same point in space

[2] Detecting panoramas

[3] Editing panoramas (cropping)

[4] Save and render panoramas

III. TIPS FOR TAKING PHOTOS FOR YOUR PANORAMA

When you are taking the photos that you plan to use in your panoramas, here are a couple of quick tips that can help you to make better panoramas more quickly.

Take a few practice shots. Before you take the sequence of pictures, make sure that the photo quality is good. When taking your practice shots, make sure you take pictures in different parts of the scene because the lighting can change from one part of the scene to the next. However, make sure that you have high-quality photos to start with and use in your panorama.

Mark the point where the sequence of photos begins and ends. Before you take the first photo that you plan to use in your panorama and after you take the last photo that you plan to use, "mark" the place where the sequence of photos begins and ends. For example, you could mark the place where the sequence starts by taking a photo of something unrelated, such as the ground, a blank sheet of paper, or another unrelated object. Do the same at the end of the sequence. This is helpful to do especially if you camera does not have a mode for taking digital photos for a panoramic images. If you do not want to take up storage space on your storage card with these unused shots, another way to mark the beginning and end of the sequence is to simply write on paper the name and number of the first and last photo in the sequence of photos that you plan to use in your panorama.

Use a tripod when you take the photos to help keep the camera level across the sequence of photos you take. Also, using a tripod lets you smoothly pan horizontally across the scene from one part to the next. Tripods come in many different sizes, so choose one that will work well for your needs.

Make sure there is overlap between one photo and the next. When taking the photos, make sure that each photo overlaps the other by at least one-third or half. This prevents gaps from appearing in your final panorama. In the following image, the amount that the two photos overlap is shown in the area that is shaded white.

Minimize the number of fast moving objects in your photos. If possible, minimize the amount of fast moving objects you have in each photo that you take. Fast moving objects will move from one part of the scene to the next. Therefore, you might only capture half of the object in the final panoramic, or the moving object might appear two times in your final panorama. This might occur if the object moves from one part of the scene to the next, and you take different photos that have the same moving object in different parts of the landscape.

Keep the focal length consistent throughout your photos. As you take one photo and then another, be careful not to zoom in or zoom out.

IV. IMAGE BLENDING

One of the problems in automatic image stitching is that of de-ghosting. When the images are taken, there is no guarantee that objects in the image stay stationary from one image to the next. One way to resolve the problem of “ghosting” is using blending algorithm.

A blender is an algorithm aiming at blending two images (or more) together. The goal of the blender is to produce a resulting image where no transition can be seen between the original source images.

The following blenders are available:

None: no blending is performed; the "top picture" is used. The stitching seams are obviously visible (geometry and/or brightness).

Fig.1 None

Linear: this mode is very fast and can be a good compromise between quality and speed if you are not too demanding on quality.

Fig.2 Linear

Multiband: this slower mode will produce much better results than the "Linear" mode. The transitional areas between images will hardly visible. However this mode does not take into account objects that could have moved from one shot to another and some “ghosting” can occur.

Fig.3 Multiband

Smartblend: this blender combines a Multiband blender and picture analysis engine in order to identify common objects present in the source images (a moving object, a strong line..). Based on the results of this analysis, Smartblend tries to keep the features common between images (edges, lines, curves,) and automatically discard objects that have moved between shots. Even though much slower than the other algorithms, this blender produces a much higher rendering quality.
Be aware: this blender is not HDR compatible yet. Smartblend, very often, is the best blender: it does a better job at harmonizing the colors, even when no color correction is applied. It prevents "ghosting", and produces a sharper image than Multiband.

Fig.4 Smartblend

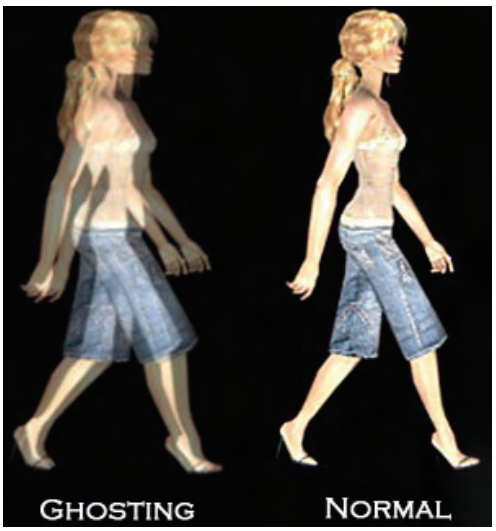


Fig.5 Example of "ghosting"

You have to know that smartblend is using a pyramidal blender like in a multi-band approach, plus a seams detection algorithm that will find where to cut pictures.

Smartblend parametars

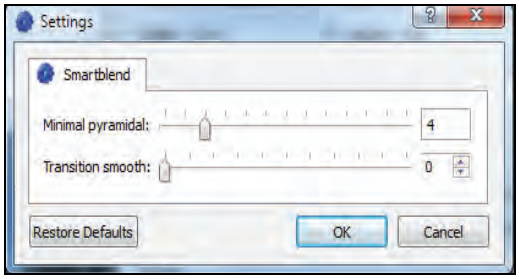


Fig.6 Smartblend paramaters

Minimal pyramidal

This parameter indicates the number of stages the blender should use. It's not coded as a pyramid level but as a stop condition. When reaching this value (in pixels), it stops the building of the pyramid for the blending. So, if you have 4096, the blender will not do any level of the pyramid as the end condition is reach at first stage. The default value is 4 pixels. That's really short but allow to have large color equalization over large areas.

Fig.7 Minimal pyramidal 4 - The blend is done over large zone making a good equalization of color

Fig.8 Minimal pyramidal 16 - You can see that just changing the value a little, it can have a large influence on the result.

Fig.9 Minimal pyramidal 64

Fig.10 Minimal pyramidal 256

Fig.11 Minimal pyramid 1024

Fig.12 Minimal pyramid 4096

Transition smooth

This value tells how sharp the transition has to be done between both pictures. It's not exactly like the first parameter as it doesn't work on pyramid number of levels but directly on images and how they are blended together at each pyramid stage. This value in fact allows a kind a linear blend on the transition zone.

This study shows the influence of Transition smooth only. I put 4096 as Minimal pyramidal value to prevent its influence.

Fig.13 Transition smooth 0 pixels, Direct blend, no smooth transition

Fig.14 Transition smooth 8 pixels, A small smooth transition

Fig.15 Transition smooth 16 pixels, A fairly large transition

Fig.16 Transition smooth 32 pixels, A really large transition zone over 32 pixels

Another way to resolve the problem of "ghosting" is to edit pictures in appropriate program for editing raster images (Adobe Photoshop) and using Brush tool with seamless colour to overlay the objects which are not mutual in every picture. After this there will not be any problem with image stitching.

IV. Conclusion

A panorama is a picture that is made by combining a series of photos into one large picture. By combining a series of photos, you can provide a complete view of an area or location that cannot fit in a single shot. But, while creating panoramas in appropriate software the problem of "ghosting" must be resolved. One way to resolve this problem is to use Smartblend algorithm for blending pictures, and another is to edit pictures before they've been stitching (in appropriate program for editing raster images - Adobe Photoshop). Panoramas can be used in many areas, but the most popular is using them for creating virtual tours. Virtual tours are popular today because of their wide area of implementation.

References

[1] Malcom Kesson, An Introduction to 3D Computer Graphics. Malcom Kesson, 1995.

[2] Anne Morgan Spalter, The Computer in the Visual Arts.Adision – Weesley, 1999.

[3] James D.Foley, Andries van Dam, Steven K.Feiner, John F. Huges, Richard L. Phillips, Introduction to Computer Graphics.Adison – Weesley, 1993.

[4] Igor Nedelkovski, Computer Graphics, 3D Modeling andAnimation. Faculty of Technical Sciences - Bitola 2008.

[5] http://www.freepatentsonline.com

[6] http://citeseerx.ksu.edu.sa

[7] http://www.computer.org

[8] http://www.autopano.net

[9] http://www.easypano.com

[10] http://www. canoma.com

[11] http://www.wikipedia.org/

[12] http://www.0-360.com

[13] http://www.panoramas.dk/

[14] http://www.panoguide.com

[15] http://www.thegnomonworkshop.com/

[16] http://www.cambridgeincolour.com

[17] http://www.commission5.isprs.org

[18] http://www.goodtutorials.org/

[19] http://www.photomodeler.com/

Optimization of Laser Marking Process with the Help of Stimulation Models

Lyubomir Lazov[1], Nikolaj Angelov[2]

Abstract **– In the article we consider the result of several numeral experiments carried out with a program product working under MATLAB for marking of tool steel and electronics elements. We search laser impact of laser CuBr and Nd: YAG laser. We discussed three different simulations with direct attitude toward process of laser marking by modification of pattern surface. The numeral experiments we carried out until the submelting point for examined patterns as we obtained the optimum ranges about basic laser parameters and those for marking technological process.**

Keywords **– Graphic images, Laser, Temperature field**

I. INTRODUCTION

Laser marking takes more and more lasting position in the recent output of electronics and machine-building products since it enables the firms to response to up-to-date requirements for control of quality. The laying of serial numbers, matrix-codes, barcodes, technical parameters, tables and other operative information is the main factor for correct optimization and monitoring of production processes [1]. We have to denote that laser technology for marking equally satisfies requirements for marking of super fragile and miniature elements in electronics as well as of super solid tools and articles of machine-building. Impetuous entering of new technology in the industrial output is due first of all to its special features such as contactless process realization; possibility for variance of marking zone range managing the parameters of lasing, selectivity of impact and simultaneous opportunity to achieve heating, melting or vaporization of material from the processing zone [2].

II. PRESENTATION

The purpose of the report is to present the opportunity for obtaining prognostic working intervals of the basic properties of lasing and technological regimes for laser marking of electronic and machine-building elements.

In our exploration we applied the program product TEMPERATURFELD 3D [3] for simulation of different models and at last we obtained as a final result three dimensional temperature fields in the laser impact zone. In some particular case of laser marking serial number experiments were accomplished to obtain temperature fields at laser marking of pattern, as some of follows parameters are changing [4, 5]: λ - – length of laser wave; q_S - laser power density; τ – length and ν - frequency of repetition of pulses, d - diameter of working spot, v - velocity of marking.

In the simulation optics and thermo-physics material properties values are reading (R – reflection coefficient , δ - depth of penetration of laser beam, k - thermal conductivity, c - specific heat capacity, ρ – density of the material, as well as the way in which they vary when temperature in the zone of activity increases.

The numerical experiments are done with samples of: Si-silicon, Ge-germanium, SiO_2, silicone dioxide, tool steel У10, У11, У12 and У13 and fast-cutting tool steels P9 and P6M5 [6]. The sources of heat created as a result of the interaction of lasing and the samples for the first three materials are considered as volumetric and for the steels – as surface [7]. Laser sources with length of laser wave λ = 1064 nm and λ = 511nm are used for the calculations, their basic parameters are specified in table 1.

The following simulations are considered in the research:

- Marking through a surface modification for samples of steel with CuBr laser;
- Marking through a surface modification under impact with Nd: YAG lasers with different duration of impulses for steel samples.
- Marking through a melt of silicone samples with CuBr laser

Table 1. Basic parameters of laser courses

Laser Parameters	IS 1064	JenLaser MOPA N45	JenLaser MOPA M45	CuBr
Length of wave λ, nm	1064	1064	1064	511
Power P, W	3	20	20	20
Frequency ν, kHz	10	100	100	19
Length of pulse τ, ns	10	100	1000	30
Diameter of working spot d, μm	28	30-80	30-100	30-80

Results and analysis from the research

Subject of research in the first simulation were the temperature profiles in samples of carbon steel under CuBr lasing with density of lasing power q_s ($6{,}00.10^9$ W/m^2 – $1{,}40.10^{10}$ W/m^2) providing maximum temperature on the surface below the melting point (1820 K). This was due to the

[1]Lyubomir Lazov is with the Technical University of Gabrovo, Hadzhi Dimitar 4, 5300 Gabrovo, Bulgaria, E-mail: lazov@tugab.bg

[2]Nikolaj Angelov is with the Technical Sciences of Gabrovo, Hadzhi Dimitar 4, 5300 Gabrovo, Bulgaria, E-mail: bo232001@yahoo.com

fact that marking through modification of the product's surface through oxidation or structural changes in a very small surface layer is a method often used in practice.

The following conclusions can be drawn from the progress of temperature profiles during the numerical experiments made and shown in fig. 1 and fig. 2:

- Temperature quickly decreases when moving away from the zone of impact in radial direction for distances 200 μm from the centre of the working spot reaching about 3 times values close to environmental temperature;
- When changing the density of power in the range examined, the temperature on the sample's surface varies within limits 850 K – 1780 K.

From the diagram of the temperature field (fig. 1 the power density $q_s = 6.82.10^9$ W/m^2) it is seen , that the surface temperature of the sample is below that of structural changes. In our other real experimental studies with the same power density a pale , low contrast slightly visible line appears (fig.3) . The obtained marking is due to the oxidizing processes on the steel surface. The temperature field on fig.2 indicates the maximum temperatures on the sample surface, exceeding 1003 K, from where structural changes occur (q_s= $9.98.10^9$ W/m^2).

After the comparison of the results from the conducted numeral experiments were made, the intervals for the power density q_s , for which the temperature is in between the allowable for appearing of structural changes in the surface layer were determined as follows: q_s ($8{,}30.10^9 – 1{,}15.10^{10}$ W/m^2)

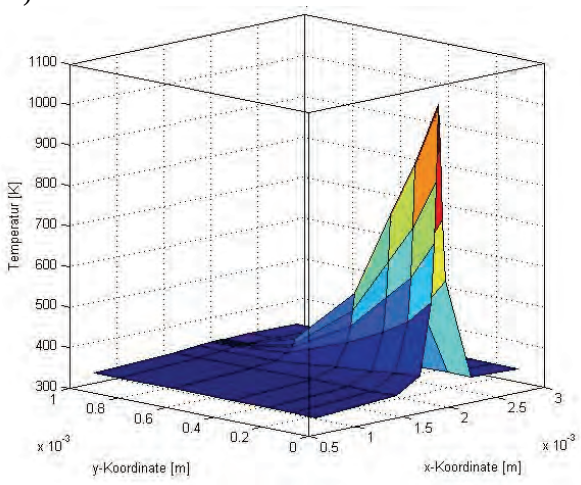


Fig. 1 Graphic of temperature field of instrumental steel У11 for power density $q_S = 6{,}82.10^9$ W/m^2 with CuBr laser

The critical power density for reaching the melt point in the processing area was determined as q_{Skp}= $1.47.10^{10}$ W/m^2 (Fig 4).

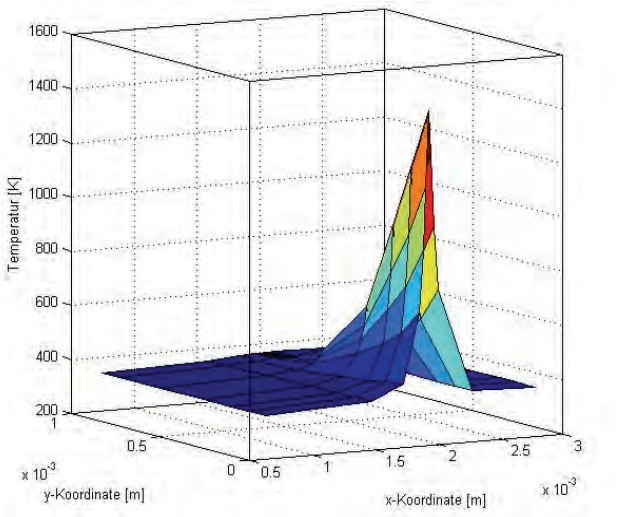


Fig. 2 Graphic of temperature field of instrumental steel У11 for power density $q_S = 9{,}98.10^9$ W/m^2 with CuBr laser

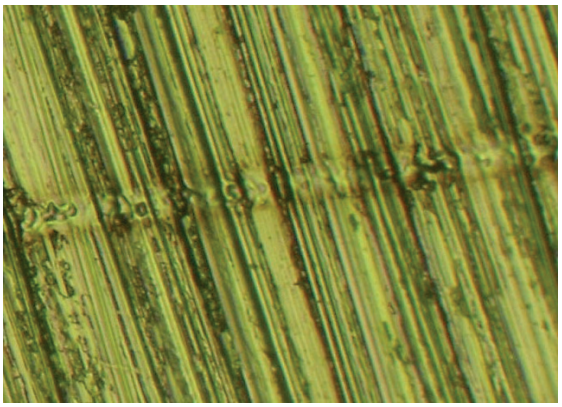
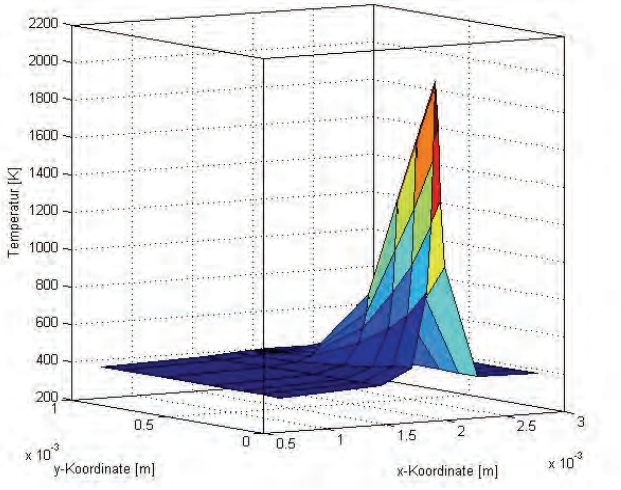


Fig. 3 Laser marking of instrumental steel У11 for power density q_S = $6{,}82.10^9$ W/m^2 with CuBr laser



Figr. 4 Graphic of temperature field upon reaching the melting point of instrumental steel У11 (q_{Skp} = $1{,}47.10^{10}$ W/m^2), CuBr laser

The investigation of the influence of the pulse length τ on the process of laser-marking was conducted under conditions of other simulation, by using Nd:YAG laser systems in three

different operating modes, respectively (τ = 10 ns; 100 ns; 1000 ns) , see table 1.
On fig. 5, 6 and 7 are presented the temperature fields in the process of laser marking through structural changes of instrument steel У12 , with a velocity v = 100 mm/s with three Nd:YAG lasers.

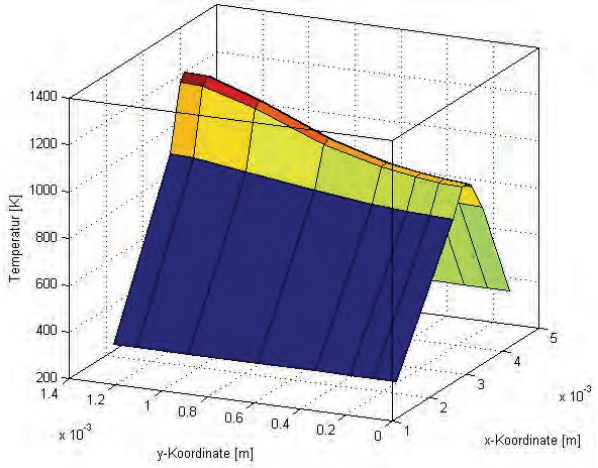


Fig. 5 The temperature field in the process of laser marking through structural changes of instrument steel У12 , with Nd:YAG laser, pulse length τ = 10 ns

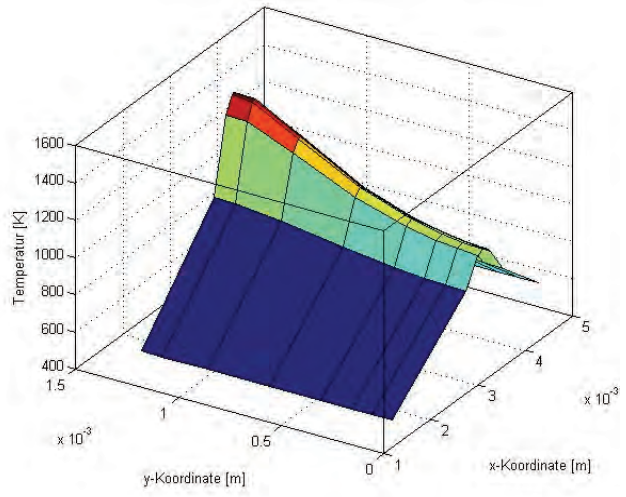


Fig. 6 The temperature field in the process of laser marking through structural changes of instrument steel У12 , with Nd:YAG laser, pulse length τ = 100 ns

From the analysis of the three-dimensional graphical images was determined the following:

- The surface temperature of the samples decreases by 100 K (from 1250 K at 10 ns to 1150 K at 1000 ns) when the pulse length is increased by two orders, nevertheless, that for the three processes is absorbed one and the same quantity of energy in each of the investigated processes;

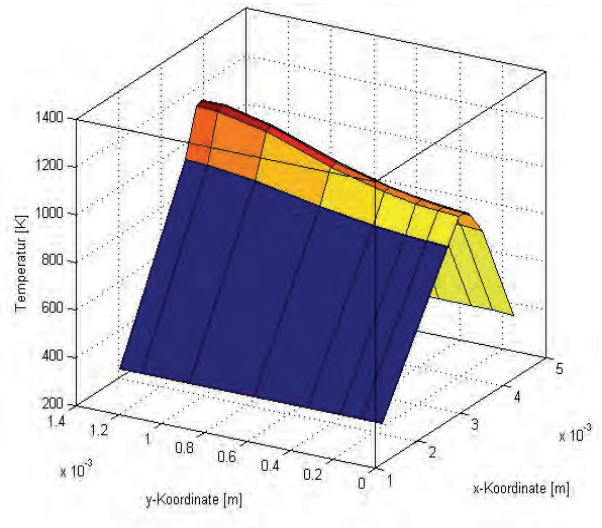


Fig. 7 The temperature field in the process of laser marking through structural changes of instrument steel У12 , with Nd:YAG laser, pulse length τ = 1000 ns

- The heat-impacted area increases nonlinearly with the pulse length increase.

The obtained result has its explanation in the physical model of interaction of the laser radiation with the substance [8] namely: The length of the short pulses is ratable with the time for absorbance of a part from the incident photons from the free electrons and the redistribution of the energy of the electronic gas to the crystal grating . Rapid heating of the impact area appears and accordingly – rapid cooling appear.
The heat transfer through heat conductivity is epsilon squared and practically, the whole absorbed energy remains in this area and its significant temperature increase appears. In the case of bigger pulse length, the heating of the sample surface is slower, and accordingly, slower cooling, compared with the shorter pulses.
The energy transfer through heat conductivity in this case cannot be neglected, the heat impacted area is expanded as part of the absorbed energy is accumulated in it.
This results in obtaining of lower temperatures in the impact area , compared to the case with using of shorter pulses.

In the third simulation the goal is to study the heating of the samples, made of silicon in the process of marking through melting in the impact area, i.e. above the melting temperature. The studies were conducted by using of a Nd:YAG laser at the following initial conditions:
Constant marking velocity v = 50 mm/s ; pulse length τ = 100 ns , spot diameter d = 80 μm.
The changes of the optical and thermal and physical properties with the temperature were also taken in mind in the time of impact, latent heat of melting the material too.
The heat source in the material is considered as volumetric.

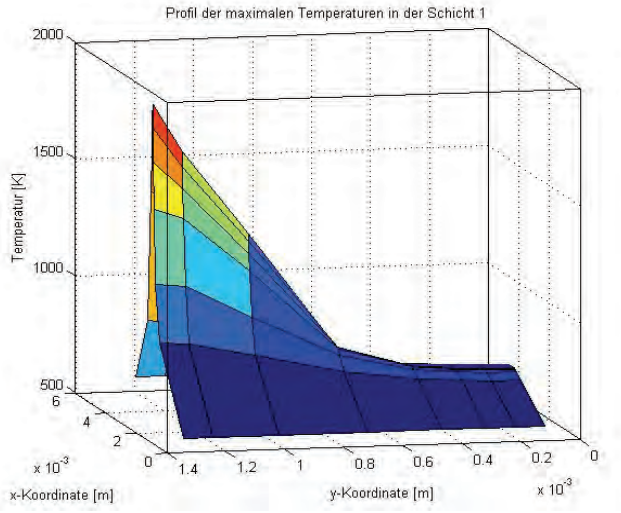


Fig. 8 The temperature field in the process of laser marking of Si with Nd:YAG laser, power density q_S = 2,90. 10^9 W/m^2.

In the process of investigation the power density is changed in the interval (2.6 .10^9 - 4.3 .10^9 W/m^2).
The temperature field for one of these numeric experiments is presented on Fig.3. The power density of the laser radiation for this case is q_s= 2.9.10^9 W/m^2
From the analysis of the results from the numeric experiments in this simulation the following conclusions can be made:

- The heat impacted area is comparable with the diameter of the working spot. The temperature gradient in this area is ≈ 10 K / μm;
- The interval for the power density of the laser radiation , where is produced marking through melting on samples, made of silicon should be maintained in the interval 2.7.10^9 W/m^2 - 3.6.10^9 W/m^2

III. Conclusion

The use of numeric methods and simulations help for the proper determination of the border zones and operation modes for different laser technological methods for processing of materials. They contribute also for the clarification of complicated issues associated with the thermo-chemical reactions, phase transitions, outbreak of substance in a liquid and evaporated form the zone of the laser impact.
These preliminary experiments result in sparing of funds and time, which is of great importance for the companies, intending to introduce different laser methods in their industrial production process.

References

[1] Системы лазерной маркировки U–серии компании LaserMark, научно-технический журнал Фотоника, 2008, бр.2
[2] Angelov N. Classification of laser marking according to requirement to it and changes in the sample, International science conference AMTECH,07, Gabrovo, Bulgaria, 23-24 November 2007
[3] Белев И. Среда за пресмятане на лазерно индуцирани температурни полета, Дипломна работа, Технически университет, Габрово, 2009
[4] Lazov Lyubomir, Nikolaj Angelov Main Factors Determining the Quality of Metal and Alloy Marking, International science conference AMTECH,07, Gabrovo, Bulgaria, 23-24 November 2007
[5] Angelov N., L. Lazov Interconnections some Parameters Affecting Quality of Laser Marking of Articles of Tool Steel, International science conference UNITECH,09, Gabrovo, Bulgaria, 20-21 November 2009
[6] www.splav.kharkov.com
[7] Динев Н. Лазерите в модерните технологии, изд. Алфа, София, 1993
[8] Angelov N. Determining Temperature Fields under Laser Treatment of Solid Bodies with a Model of Semi-Bounded Body Having an Internal Head Source, International science conference UNITECH'04, Gabrovo, Bulgaria, 18-19 November 2004

3D Graphic Images of Laser Inducion Temperature Fields

Lyubomir Lazov[1] and Nikolaj Angelov[2]

***Abstract* – A possible to obtain three-dimensional graphic images of laser-induced temperature fields at different moments of time for samples of tool steel is shown. Optimal intervals for certain parameters when using Nd: YAG-laser are determined.**

***Keywords* – graphic images, laser, temperature field**

I. INTRODUCTION

Theoretical model investigations in the field of laser radiation interaction with the substance are of key importance to the analysis of experimental results obtained during various technology processes [1]. Similarly, this appears to be a principal factor which determines the trends in conducting a number of novel experimental investigations. In summary their role in the specific practical applications of lasers in treating materials can be rendered in the following way:

- Determining of destruction threshold in treating certain type of material by laser source;
- Optimizing the course of various laser technology processes such as marking, etching, cutting , welding, boring of holes, hardening etc.;

II. PRESENTATION

With laser impact on tool steel a portion of the radiation is absorbed whereas the remainder part is reflected by the surface and another third part passes through it thus satisfying the condition

$$E = E_A + E_R,$$

Where $E$is the energy of falling radiation, E_A - is the energy of absorbed radiation, E_R, - is the energy of reflected radiation.

In fig. 1 is shown a general diagram of energy distribution of the process of laser impact upon alloys and metals(including tool steel).

Apart from the energy gained through absorbed radiation it is possible to obtain in the impact zone some energy which is produced by way of chemical reactions E_{ch} while its temperature change. The energy balance equation is then

$$E_A + E_{ch} = E_{proc} + E_c + E_r + E_{con} \qquad (1)$$

where E_{proc} is the energy of the process; E_c – is the energy which has entered the thermally impacted zone around the work space; E_r – is the energy of radiation; E_{con} – is the energy loss due to convection.

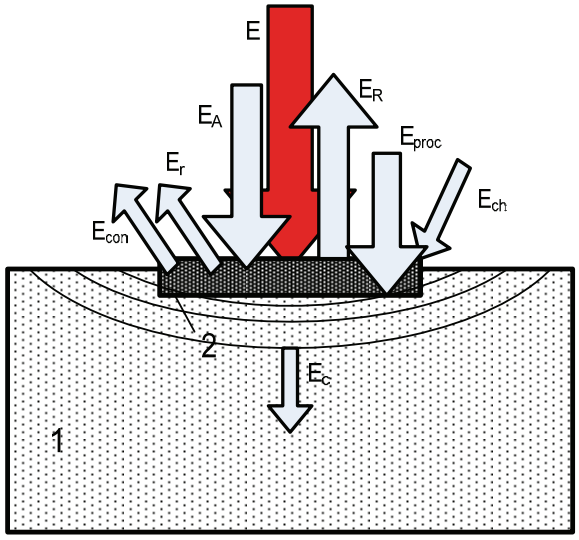


Fig.1 a general diagram of energy distribution

of the process of laser impact upon alloys and metals

Physical model

In the theoretical model under observation the calculations for the laser induced temperature fields are made by taking into account these restrictions:

- Ignoring thermal losses resulting from convection and radiation;
- Ignoring non-uniformities and non-homogeneities in the impact zone.

Likewise, the thermal impact of laser radiation in the treatment zone in the model is assigned by the heat conduction equation of the kind

$$c\rho\frac{\partial T}{\partial t} = div(k\ grad\ T) + q_S(1-R)\alpha\ , \qquad (2)$$

where T is temperature; t - time; c - specific thermal capacity; ρ - density; k - coefficient of thermal conductivity; q_s - laser radiation power density; R - coefficient of reflection; α - coefficient of absorption.

Since the absorption of the falling laser radiation takes place in a thin layer close to the surface [] in the theoretical model it is possible to regard thermal impact as surface heat source. In laser marking of metal surfaces with pulse durations being in the range τ = (10 – 1000) ns this requirement is in effect because the thickness of the sample

[1]Lyubomir Lazov is with the Technical University of Gabrovo, Hadzhi Dimitar 4, 5300 Gabrovo, Bulgaria, E-mail: lazov@tugab.bg

[2]Nikolaj Angelov is with the Technical Sciences of Gabrovo, Hadzhi Dimitar 4, 5300 Gabrovo, Bulgaria, E-mail: bo232001@yahoo.com

under marking is much greater than the depth of penetration δ of laser radiation ($\delta = 1 / \alpha$).

By means of the theoretical model it is possible to estimate the required power density at which the surface material is being melted. The latter is of great importance for the method of laser marking effected by way of structural changes of second order or by oxidization.

Laser marking by means of taking off a pre-marked layer on the sample enjoys a considerable amount of practical application [2].The theoretical model allows to perform calculations in various solutions such as:

- Laser radiation is allowed through the layer and absorption takes place at the layer/pad boundary;
- Laser radiation is absorbed by the layer,then evaporates it thus uncovering the base material.

An important aspect in laser marking is the calculation of laser induced temperature fields in treatments utilizing phase changes in the material.The proposed theoretical model allows the solution of this task as well. Heat conduction equations in this case are to be solved for both solid phase and the molten metal.

In terms of practical application it is interesting to make observation of the case in which the material underneath the surface in the absorption zone starts boiling and bursts out due to the generated high pressure; part of that material is deposited aside of the marking zone in the form of splashes. During work with pulses of τ = 10 ns sublimation (direct evaporation from the impact zone) is observed. This process is of great importanece in the case of precision etching of metal surfaces.

Theoretical model estimations could be subjected to the following principal tasks which are of substantial importance to the specific processes and methods of laser marking and etching:

- Clarifying the effect of the wave length λ of laser radiation for the concrete material. This facilitates the selection of appropriate laser source (fiber laser λ = 1064 nm, Nd:YAG laser with λ = 1064 nm, laser of CuBrwith λ = 511; 578 nm, eximer lasers (in the ultra-violet range) etc.;
- Clarifying the role of the duration of pulses τ in spatial distribution of temperature field and the physical process of radiation interaction with the metal τ = (10 – 1000) ns;
- Determining the effect of diameter *d* at the work spot upon temperature distribution *d* = (10 – 200) μm;
- Clarifying the role of pulses frequency upon the temperature field;
- Clarifying the role of the laser beam relative speed of displacement ,with regard to the treated surface, upon the profile of the temperature field;

Basic elements of related software

Temperaturfeld3D software is the working medium for calculating temperature fields with a wide set of input parameters and possibilities for analysis of estimated results [3].

Two models of estimation are used : dynamic and static. In the dynamic model the laser beam moves within xy-co-ordinate system and affects the material. Temperature fields are determined both at the surface of material and in depth. In this case there are 3 dimensions *x, y* and *z*, which determines the complexity and time for calculations. In static model laser beam remains static and the material has radial symmetry. In Thus dimensions are reduced to 2 – *r* and *z* and calculation is accelerated considerably.

After selecting a model the main window drops. From there the user has access to the three principal sections of software- input/output ; model parameters and calculation.

To calculate one of the models it is necessary to enter a lot of parameters:

Program/Software parameters

- File name
- Initial time step/interval
- Minimum time step/interval
- Maximum time step/interval
- End up of calculations
- Ambient temperature
- Radiative capacity
- Calculations accuracy
- Initial time for storing results
- Coefficient of storing
- Liquid phase transition
- Method for model calculation

There are two methods available: Gear and Adams with a variety of sub-versions.

Geometry parameters

Here are defined geometry dimensions of treated sample as well as the number of composite materials it is made of.

- Geometry model

Three geometry models are available: rectangular prism, layer, semi-space.

- Computerized generation of elements
- Geometry dimensions

Laser parameters

Laser radiation and movement are assigned here.

- Type of laser
- Time distribution of power density
- Gaussian radius of beam
- Time for following of pulses
- Initial co-ordinates and laser beam speed

Materials parameters

The relationship between the number of various materials indicated by the geometry parameters and the parameters of materials themselves is made inhere

- List of available materials
- Booting data base
- Defining a new material

In a series of windows wherein the four groups of parameters are defined: reflection coefficient; heat conduction coefficient; specific heat capacity; absorption capacity.

- Change of parameters of certain material

Calculation

The stages of calculation are described in the items that follow:

- Preparation

Prior to starting calculations all necessary parameters are checked for availability.

- Calculation and results

Following the writing of data in the respective file, the calculation kernel is activated to perform calculation of temperature fields.

- Calculations analysis

Upon opening analysis window all data from the file with calculated results are booted automatically:

-approximation of results;
- animation of the entire process;
- temperature profile of material at certain moment;
- maximum temperature profile
- temperature time dependence;
- material in-depth change of temperature ;

Examples of numeric experiments

By means of this product(software) numeric calculations are done aiming at finding out appropriate operation mode for marking samples made of tool steel type У11 with Nd:YAG laser by way of the surface structural changes method. The aim is to determine the distribution of laser radiation absorbed energy inside the sample according to the developed physical model [4] and to compare the results with the experimental investigations. Marking done by this method is carried out below the surface destruction threshold for steel type У11 that is ,the temperature in the treatment zone in numeric experiments should not exceed melting temperature (1820 K). In numeric experiments the zone under laser impact is divided into three layers with thickness of Δh_1 = 12,5 μm, Δh_2 = 30,0 μm, и Δh_3 = 300 μm. The time for laser impact is kept constant i.e the speed of movement, duration and frequency of pulses in numerical experiments remain unchanged.

In fig. 2 and fig 3 and fig. 4 are shown temperature profiles for different layers of the sample for time moment t = 3 ms following the start up of the process. The movement of the laser beam is along the Oy axis and the process starts from a point with co-ordinates (2mm,0). The parameters of the laser radiation and the technology process are:

Wave length	λ = 1064 nm;
Power density	$q_s = 5{,}70.10^9$ W/m^2;
Duration of pulses	τ = 100 ns;
frequency	ν = 10 kHz
velocity of marking	v = 20m/s..

The graph in fig 2 concerns the first layer of the tested sample; fig 3 concerns the second layer and fig 4 concerns the third layer. Temperature drop in depth is in accordance with the physical model which describes laser radiation absorption by the free electrons in metals and alloys. The depth of penetration of the of laser radiation is ~ 10^{-7} m, which suggests the consideration of surface heat source.Heat is distributed across material by way of heat conduction (1). For the short time interval of laser impact ($\sim ms$), depending on the rate of marking and cooling of the impact zone, (~ 10^{-6} K/s) heat cannot reach a substantial depth . From the graph in fig. 4 it is evident that in depth the maximum temperature is equal to 450 K.

Numerical experiments lead to the conclusion that the depth of marking done by the method of structural changes

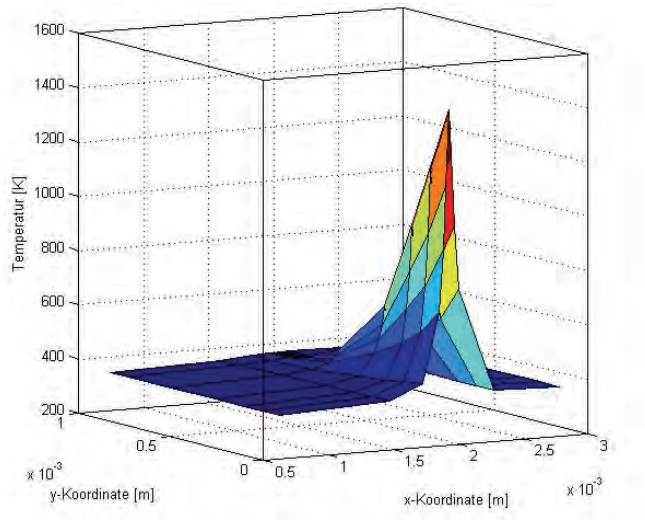


Fig.2 Temperature field of the upper layer with thickness of Δh_1 = 12,5 μm for sample made of steel type У11 impacted by Nd:YAG laser

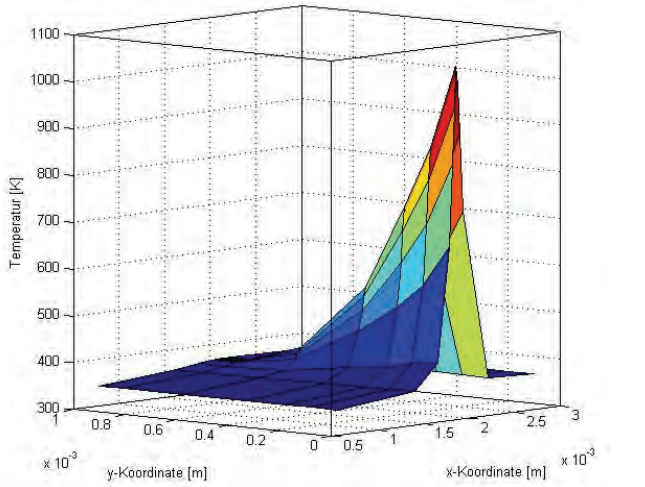


Fig.3 Temperature field of the upper layer with thickness of Δh_2 = 30,0 μm for sample made of steel type У11 and impacted by Nd:YAG laser

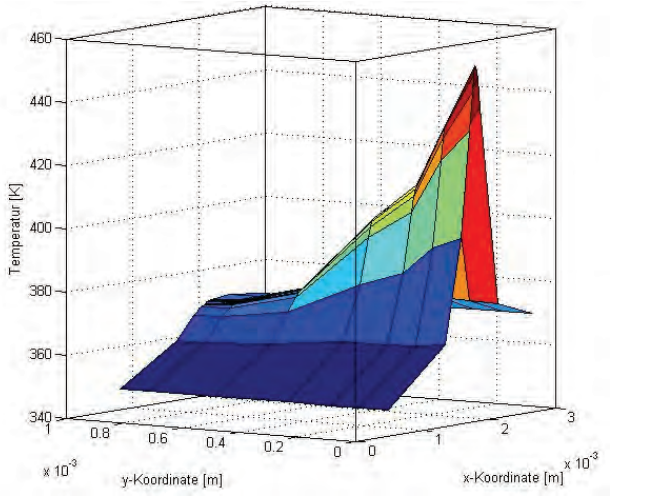


Fig.4 Temperature field of the upper layer with thickness of Δh_3 = 300 μm for sample made of steel У11 impacted by Nd:YAG laser

in the surface layer of the sample is approximately equal to 40 μm. It is a known fact that structural changes in steel take place within the temperature interval K.Results obtained from numerical experiments correlate well with those obtained by real experiments [5].

III. CONCLUSION

Numeric experiments for determining temperature fields by means of this product are appropriate not only for optimizing the process of laser marking but also for other laser technology processes such as cutting, welding, boring, hardening etc. Compared with the real experiment results they contribute to the rise in quality and effectiveness of laser treatment processes. The software features options for laser treatment numeric experimentation concerning various materials from mechanical engineering, electronics instrument engineering etc.

REFERENCES

[1] Reinhart Poprawe Lasertechnik für die Fertigung, Springer 2004 ISBN 3-540-21406-2

[2] Динев Н. Лазерите в модерните технологии, изд. Алфа, София, 1993

[3] Белев И. Среда за пресмятане на лазерно индуцирани температурни полета, Дипломна работа, Технически университет, Габрово, 2009

[4] Angelov N. Determining Temperature Fields under Laser Treatment of Solid Bodies with a Model of Semi-Bounded Body Having an Internal Head Source, International scientific conference UNITECH'04, Gabrovo, Bulgaria, 18-19 November 2004

[5] Lazov L., N. Angelov, A. Atanasov Optimization of the Parameters of the Laser Irradiation during the Marking on Instrumental Steel, International scientific conference UNITECH'04, Gabrovo, Bulgaria, 24-25 November 2006

A Low Cost Method for 3-D Documentation of Cultural Heritage Objects

Jove P. Pargovski[1]

***Abstract* – In this paper a low cost method for 3-D documentation of cultural heritage objects is presented. Cultural heritage is a testimony of past human activity. As such, cultural heritage objects exhibit great variety in their nature, size and complexity. From small artefacts and museum items to cultural landscapes, from historic buildings and ancient monuments to city centres and archaeological sites.**

Cultural heritage around the globe suffers from wars, natural disasters and human negligence. The importance of cultural heritage documentation is well recognized and there is an increasing pressure to document our heritage.

***Keywords* – Cultural Heritage, 3D Modeling, Photogrammetry, Documentation.**

I. INTRODUCTION

The need for digital documentation, preservation and conservation is given by the constant threat and danger that are affecting the Cultural Heritage. Currently three main approaches can be distinguished, for the optical recording, documentation and visualization of heritage sites, monuments and finds:

1. Image - based methods (e.g. photogrammetry): these methods are widely used for 3D reconstruction of architectural objects [1],[2], for precise modeling of terrain and cities [3] or monuments and statues [4] and lately also for precise and detailed modeling of complex objects using consumer-grade digital cameras [5]. Image-based methods use projective geometry or a perspective camera model. They are highly portable and the sensors are often low priced.
2. Range - based methods (e.g. laser scanning): these techniques are based on active sensors that directly capture the geometric 3D information of an object using artificial laser light (e.g. ShapeGrabberTM) or projecting a pattern (e.g. BreukmannTM) without requiring a mathematical model to derive 3D information from 2D observations. Applying different measurement principles, they capture millions of points in relatively short time but require great editing efforts to correctly model the recorded 3D data. They are quite expensive and often unpractical in some field campaigns.
3. Combination of image - and range -based methods: in many applications, a single 3D modeling method that satisfies all project requirements is not yet available. Photogrammetry and active sensors have been often combined in particular for the recording of large architectural objects or complex archaeological sites, where no technique by itself can efficiently and quickly provide a complete and detailed model [6 - 9].

Comparisons between image-based and range-based modeling are reported in numerous papers [10], [11] and the experts agree that at the moment, for all types of objects and sites, there is no single modeling technique able to satisfy all requirements of high geometric accuracy, portability, full automation, photo-realism and low costs as well as flexibility and efficiency.

Anyway, for some projects, like those described in this paper, currently available image based technologies can offer a lot of improvements in the current practices, and the generated documentation.

Currently on the market there are a lot of low cost software packages, and digital cameras with constantly increasing performances. Every conservation work, field work, excavation, according to the regulations, must include documentation and photography. Since the photographs are already taken, there is no reason not to do that in photogrammetric manner, which would also enable their use in the image based 3D modeling of the object.

There are three case studies presented in this paper. The main goal of these examples is to accentuate how easy 3D documentation can be implemented in the current work practices. The first two case studies are two churches, which complete 3D models were made in less than 4 hours. From these models, all the measurements can be derived and they can also be used in a presentation proposes. The third case study is a measurement check project, in which the precision of the approach presented in this paper is checked.

As it can be seen from the conclusion, the approach presented in this paper, can efficiently be implemented in the current work practices. The recent developments in the new technologies enabled this, so there is no reason not to use the full potential of image based 3D modeling.

II. CASE STUDY 1: CHURCH ST. ATANASIJ IN BITOLA (NEAR VERO SUPERMARKET)

The church “St. Atanasij” is a new church build in 2008/2009. As it can be seen the exterior is almost finished, and this year it is planned to start with the interior.

[1]Jove Pargovski is with the Ministry of Culture of R. Macedonia, Cultural Heritage Protection Office, Bitola branch.
Email: j.pargovski@uzkn.gov.mk

Fig.1. Church St. Atanasij

The 3D image based model of this church, was made using low cost digital camera (previously calibrated) – Canon PowerShot A430 (4MP) and PhotoModeler application. Reference measurement points were taken during the field work, which later were used for adjusting the scale and dimensions of the object and also for precision checking.

The whole project of 3D modeling was finished for less than a 4 hours. During the field work, 10 photographs were taken, covering the object from all sides. The location of the near market, enabled the generation of "high" angle photos, for one side of the church.

High ground photos from the south side could not be taken, so as an alternative, photos from a greater distance were taken. This on other side reduces the precision of the generated 3D model, due to small number of pixels available in the digital camera – 4MP. In order to increase the precision of the generated 3D model, it is advisable to use a digital camera with higher resolution, and also printed target marks which should be placed on the object before it should be photographed.

Fig.2. Image based model of the church St. Atanasij

III. Case study 2: Church St. Spas, location Cebren, Mariovo

The church St. Spas, near the monastery complex "Cebren" is a small church (6.3 x 4.3 m) build from simple river stone.
The interior is divided in two with a wooden altar. All four walls from the church interior are decorated with painted frescos. According to the available scientific literature, the frescos are dated back from the XVI century. Their preserved area is ~ 60 m^2. Aldo the exterior of the church is restaurated several times through the centuries, the interior is kept untouched, which gives the real value to this church.

Not so far from the church in the future is planned a big hydro plant – "Cebren" to be build. This means that the location of the church would be under water, and the church, along with the nearby St. Dimitrij church, and the archeological site "Chebren" has to be dislocated.

Dislocation of this objects, means that the generated documentation must be elaborate and precise, so they can be brought back in their original form, in the new location.

The photographs used for generation of 3D model of the church St. Spas, were taken by an amateur photographer, during his field work. The camera model used was Sony DSC R1 (10 MP) and the photographer was told, to keep the parameters of the camera constant and to take photographs from all sides of the object. The other day in the office, the camera was calibrated with the same parameters.

The goal of this "assignment" was to prove how easily and efficiently the Photogrammetry and 3D modeling can be used, and which are the benefits of their use in the field work:

The church St. Spas is located in the Chebren locality - 4 hours driving with terrain vehicle, on a bumpy road, on which you can drive only when the soil is dry meaning only in summer. This means that when you arrive on the location, you must generate as much information's as you can, because you don't know when you will have another chance. The generation of the photos used for the 3D model was in a time interval which was less than 10 minutes, but yet, from these photos, all the dimensions of the object can be derived, back in the office.

In the same manner, the interior of the church (which is most valuable) can be 3D modeled, but for this kind of project, a special preparation work is required and appropriate lighting.

Fig.3. Church St. Spas

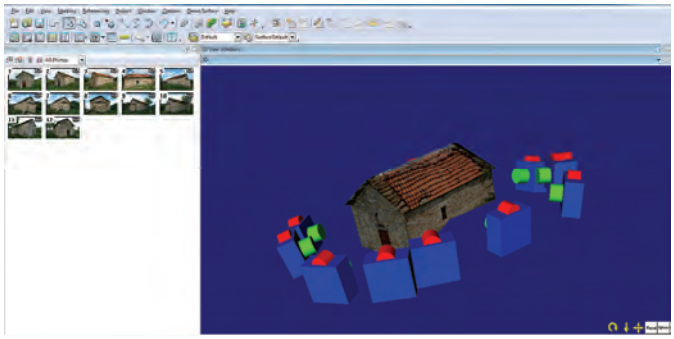


Fig.4. Generated 3D model of the Church St. Spas, along with the camera stations

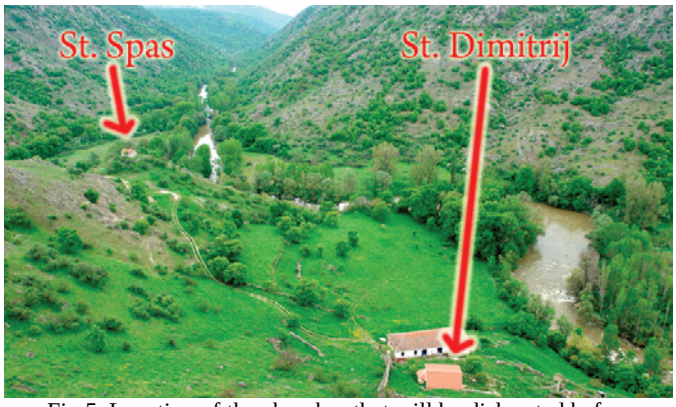


Fig.5. Location of the churches that will be dislocated before building the "Chebren"hydro plant

The nearby church St. Dimitrij (Fig. 5) is also planned for dislocation, this church is bigger, it is reconstructed 4 years ago, and its altar part is dated back from the XVI century.

The previous two case studies have shown how easy is to generate a 3D model of an object, and by doing that, to generate an elaborate and precise documentation.

The question that should be answered is how precise is the generated documentation.

In the previous two projects, real measurements of the objects have been taken, but simply for orientation and scale proposes, and some minor accuracy tests. Also for these two objects, an analog generated documentation already exists, and from which by comparison it is determined an error less than 2 cm in all measurements. Yet, the problem with this documentation is that we don't know how accurate the analog measurements were, and from which points the measurements were made.

The achieved accuracy in a 3D modeling project is dependent of a number of factors, concerning the used camera and its handling, the object and its size and not to forget the clear definition (natural vs. targeted) of the measured points.

For accuracy check in this paper a third case study is presented, and in this case the goal was not to model in 3D the object, but simply to use the approach as a measurement tool and to check its precision.

IV. Case Study 3: Entrance of the Bitola Museum

The building of the museum in Bitola is build in the middle of XIX century, to serve as a military school. In this school the father of modern Turkey – Mustafa Kemal Ataturk had studied, which makes this building one of the most valuable in Bitola. It is big object, and to generate a 3D model would be a serious project, in which a lot of preparations should be made. Yet, its entrance gives a perfect opportunity to check the precision of the approach elaborated in this paper. The balcony above the entrance makes it possible to place and measure the distance of a printed target marks and by doing so a big coverage of the object is ensured and the precision can be checked. 20 target marks were placed on the object, and the distances between them were precisely measured. The placement of the targeted marks also enabled more precise referencing, and greater precision of the measurements.

The goal in this project was not to model the entrance, but simply to check the precision of the elaborated approach, and its possible use in the cultural heritage documentation.

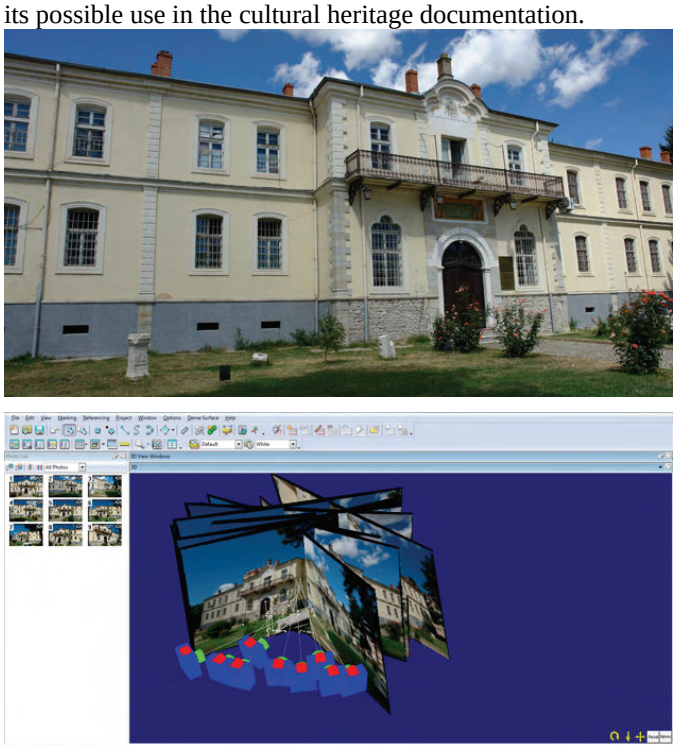


Fig.6. Entrance of the museum in Bitola

Table 1 shows 10 previously measured (real distances RD) between the targeted marks. MD (3D Measured Distances) - the distances measured in the Photo modeler application. The camera used in this project was previously calibrated Sony DSC R1 (10 MP).

As it could be seen from table 1, the Absolute Error (AE) in the largest distance (3,3 m) is near 1 cm, which means Relative Error (REL) ~ 0.3 %.

The generated errors in the other measurement are even smaller and in the analog measurements they are often neglected. Also it should be noted that the used camera had a resolution of 10 MP, which today is a standard for entry level models. Semi professional and professional digital cameras used in the field works, have a resolutions even double than

this, which means even greater precision of the generated 3D models.

Table 1. Accuracy check of the 3D generated model

RD (cm)	MD (cm)	AE	REL	REL (%)
54.2	54.3	0.1	0.00185	0.1845
66.5	66.2	0.3	0.004511	0.451128
73.4	73.7	0.3	0.00409	0.40872
78.3	78.1	0.2	0.002554	0.255428
86.4	86.9	0.5	0.00579	0.5787
95.8	95.6	0.2	0.002088	0.208768
98.5	98.3	0.2	0.00203	0.203046
122.2	122.5	0.3	0.00245	0.2455
324.1	322.9	1.2	0.003703	0.370256
335.3	334.2	1.1	0.003281	0.328064

This example clearly shows that the generated image based 3D models of the cultural heritage objects, offer a level of precision greater than the analog measurements.

V. CONCLUSION

The previous case studies have shown that the image based 3D modeling offers a great potential in the processes of documentation and presentation of cultural heritage.

We can see that it is easy to educate the current staff, to take field photos of the heritage objects, in a photogrammetric way. As it is said in the beginning of the text, since the photographs must be taken during the field works, there is no reason not to do this in a photogrammetric manner.

The experiments made in the area of image based 3D modeling, had shown that if the digital camera is properly calibrated, and if 3D modeling procedures are followed, than the generated models, have far greater accuracy, than the drawings generated trough the current practices.

Another positive side that should be mentioned is the speed for generation of 3D model. If the cultural heritage object is properly photographed, than the appropriate 3D model can quickly and efficiently be reconstructed back in the office.

The 3D modeling approach has also the advantage of reconstruction an object from old photographs. There are a numerous examples of long gone (ruined) objects, from which old photographs exist. In some of these examples, the only requirements are two photographs from different angles, and some kind of reference point (for example the base of the ruined object).

We can freely say that the current analogue practices are old, imprecise and can't match the new technologies. With the current development of the technology, we have mobile phones equipped with digital cameras with more than 5 MP, whose photos can also be used for generation of 3D models. Also with the images from mobile phones, we have the possibility of geo referencing, which means that every generated image, has also meta information about the location on which it is taken.

Trough the heritage sites around the world there are numerous examples of successful implementations of image based 3D modeling.

Image based 3D models can also be used in educational and presentation proposes. There are endless possibilities offered by this technology, and according to the experts it will become standard practice in the documentation activities.

REFERENCES

[1] Remondino F., El-Hakim S., 2006: "Image-based 3D modelling: a review". Photogrammetric Record, 21(115), pp. 269-291

[2] Debevec P.E., Taylor C.J. and J. Malik, 1996: "Modelling and rendering architecture from photographs: a hybrid geometry and image-based approach". ACM Proceedings of SIGGRAPH '96, pp. 11-20

[3] Grün A., 2000: "Semi-automated approaches to site recording and modeling". IAPRS&SIS, 33(5/1): 309-318.

[4] Grün A., Remondino F. and Zhang L., 2004: "Photogrammetric Reconstruction of the Great Buddha of Bamiyan, Afghanistan". Photogrammetric Record, 19(107): 177-199

[5] Remondino F., Zhang L., 2006: "Surface reconstruction algorithms for detailed close-range object modeling". IAPRS&SIS, Vol. 36(3), pp. 117-121, Bonn, Germany

[6] El-Hakim S.F., Beraldin J.A., Picard M. and Godin G., 2004: "Detailed 3D reconstruction of large-scale heritage sites with integrated techniques". IEEE Computer Graphics and Application, 24(3): 21-29

[7] Lambers K. Eisenbeiss H., Sauerbier M., Kupferschmidt D., Gaisecker T., Sotoodeh S., Hanusch T., 2007: "Combining photogrammetry and laser scanning for the recording and modelling of the Late Intermediate Period site of Pinchango Alto, Palpa, Peru". Journal of Archaeological Science 34

[8] Voltolini F., Remondino F., Pontin M., Gonzo L., 2006: "Experiences and considerations in image-based-modeling of complex architectures". IAPRS&SIS, Vol. 36(5), pp. 309-314

[9] Voltolini F., El-Hakim S., Remondino F., Girardi S., Rizzi A., Pontin M., and Gonzo L., 2007: "Digital Documentation of complex architectures by integration of multiple techniques – The case study of Valer Castle". Videometrics IX, Proc. SPIE-IS&T Electronic Imaging, Vol. 6491

[10] Böhler W. and Marbs A., 2004: "3D scanning and photogrammetry for heritage recording: a comparison". Proceedings of 12th Int. Conf. on Geoinformatics, 7-9 June 2004, Gävle, Sweden, 291-298

[11] Remondino F., Guarnieri A. and Vettore, 2005: "3D modelling of close-range objects: photogrammetry or laser scanning?" Videometrics VIII, Proc. SPIE-IS&T Electronic Imaging, Vol. 5665, pp. 216-225

POSTER SESSION PO XI

PO XI - Electronic Components, Systems and Technologies

Realization of Direct Digital Synthesis Generators Based on FPGA and PSoC Integrated Circuits

Eltimir Ch. Stoimenov[1] , Georgi S. Mihov[2] and Ivailo M. Pandiev[3]

***Abstract* – This paper presents an implementation of a Direct Digital Synthesis (DDS) generator incorporating FPGA and "System on Chip" (PSoC) technologies. The DDS is a digitally-controlled method of generating multiple frequencies from a reference frequency source. An evaluation board equipped with a FPGA Spartan-3A and a PSoC (from Cypress Inc.) integrated circuits were used for development. PSoC is a programmable integrated circuit obtaining certain analog and digital recourses with general purpose. In the current research the FPGA is synthesize sinusoidal, or other, signal in digital form. The synthesis itself is achieved by a Look-Up Table (LUT) containing 256 samples. An UART interface is used to transfer the information from the FPGA to a PSoC where an 8 bit digital-to-analog conversion (DAC) took place. Moreover, a two pole low pass filter (LPF) and a programmable gain amplifier (PGA) is incorporated in the PSoC. In addition the parameters of the created DDS generators can be controlled by a personal computer through a USB interface.**

***Keywords* – Mixed-signal systems, Function generators, Direct Digital Synthesis, FPGA, PSoC.**

I. INTRODUCTION

The function generators have been found useful in many applications, such as analogue and mixed-signal processing, telecommunications, and measurement systems. Additionally, the principals of oscillation can be extended to construct other types of generators, such as voltage-controlled oscillators (VCOs), quartz crystal resonator sensors, voltage-to-frequency converters, etc. [1-4]. The traditional hardware realization of function generators uses closed-loop analog circuits with active and passive elements or digitally-controlled method of generating multiple frequencies from a reference frequency source called Direct Digital Synthesis (DDS) [1]. The main advantage of the DDS generators is that the typical amplitude stabilization is not higher than 0,01% and the typical THD is 0,1% [5]. Generally, the DDS generators are a type of frequency synthesizer used for creating arbitrary waveforms. In them stable clock oscillator drives a Programmable-Read-Only-Memory (PROM) which stores one or more integral number of cycles of a sine waveform (or other arbitrary wave-form, for that matter). As the address counter steps through each memory location, the corresponding digital amplitude of the signal at each location drives a Digital-to-Analog Converter (DAC) which in turn generates the analog output signal. Nowadays, one of the most effective techniques for realization of DDS generators is by using Application Specific Integrated Circuits (ASICs) with fixed electrical parameters (such as the DDS function generators from Analog Devices) or by developing electronic circuit based on FPGA (Field Programmable Gate Array) and PSoC (Programmable-System-On-a-Chip). The FPGA ICs are becoming more popular and are used in many mixed-signal applications [6-8]. To the authors' knowledge, DDS function generators based on FPGA and PSoC ICs have not yet been reported in the literature. It is, therefore, the purpose of this paper to present a programmable DDS function generator employing FPGA and PSoC ICs.

II. DIRECT DIGITAL SYNTHESIS (DDS) : GENERAL OVERVIEW

A frequency synthesizer, including DDS, generates multiple frequencies from one or more frequency references**.**

The block diagram of a classical DDS system is shown in Fig.1.

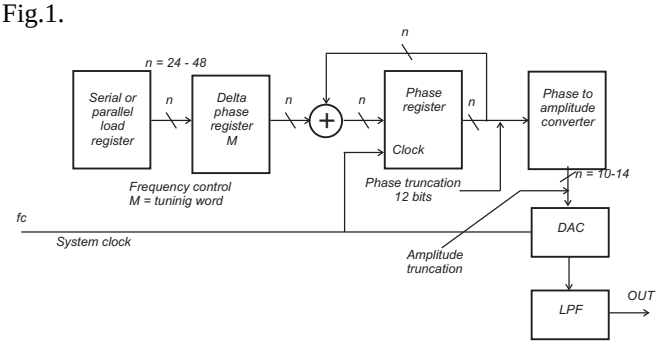


Fig. 1. Block diagram of DDS system.

The heart of the system is the *phase accumulator* whose content is updated once each clock cycle. Each time the phase accumulator is updated, the digital number, M stored in the *delta phase register* is added to the number in the phase accumulator register.

The truncated output of the phase accumulator serves as the address to a sine (or cosine) look up table. Each address in the look up table corresponds to a phase point on the sine wave from 0° to 360°. The LUT contains the corresponding digital amplitude information for one complete cycle of a sinewave. The LUT therefore maps the phase information from the phase accumulator into a digital amplitude word, which in turns drive the DAC. Because of the symmetry of the sine wave, only data for 90° could be use for reconstruction.

[1] Eltimir Ch. Stoimenov is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: e_stoimenov@tu-sofia.bg

[2] Georgi S. Mihov is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: gsm@tu-sofia.bg

[3] Ivailo M. Pandiev is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ipandiev@tu-sofia.bg

For an n-bit phase accumulator (n generally ranges from 44 to 32 in most DDS systems), there are 2^n possible phase points. The digital word in the delta phase register, M, represents the amount the phase accumulator is incremented each clock cycle. If f_c is the clock frequency, then the frequency of the output sinewave is equal to

$$f_0 = \frac{M \times f_c}{2^n}. \quad (1)$$

This equation is known as the DDS "*tuning equation*". The frequency resolution of the system is equal to $f_c / 2^n$. For n = 32, the resolution is greater than one part in four billion [1].

III. DDS Function Generator Development Based on FPGA and PSoC

The current paragraph discusses the particular hardware and software realization of DDS function generator. Also the objectives and the approach are presented.

A. Objectives

Design a function generator with the following parameters:

- Output signal forms : *sine, triangular, square;*
- Frequency band : *0 – 2 MHz;*
- Frequency tuning method : *digitally - minimum 8bits resolution;*
- Amplitude tuning method : *digitally;*
- Output signal amplitude : *±10V*;
- THD: 0.1%
- PC connectivity: *via standard USB ports*.

B. Approach

In order the above objectives to be achieved a DDS function generator is designed with the block schematic shown in Fig. 2.

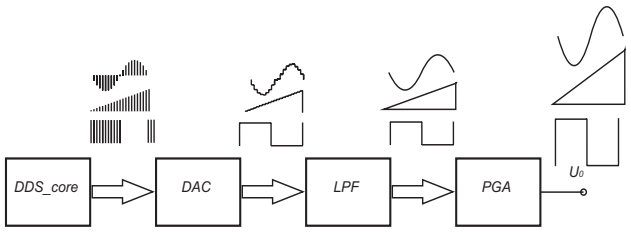


Fig. 2. DDS development block diagram.

The DDS core, shown on Fig. 2, generates a digital signal with a sine, triangle or square shape and passes it to a DAC. After the conversion the analog signal is filtered by an anti-imaging second-order Low-Pass Filter (LPF) and amplified by a Programmable Gain Amplifier (PGA).

In our development the DDS generating core is implemented on a FPGA, which allows best versatility and makes the design open for constant improvement. VHDL is used for programming FPGA IC. On the other hand the DAC function is accomplished by a PSoC from Cypress Inc. Using the PSoC technology also allows implementation of the LPF and PGA on the same integrated circuit. The above functions of DAC, LPF and PGA are realized by the graphical editor of PSoC Designer IDE from Cypress Inc.

Xilinx® Spartan®-3A Evaluation Kit from Avnet is used as a development platform for realization the circuits. It poses the following key features related to the current design [9]:

-Xilinx XC3S400A-4FTG256C Spartan-3A FPGA;
-Cypress 8C24894 PSoC;
-USB connectivity.

The block diagram of the Evaluation Kit is shown in Fig. 3. The blocks of interest are patterned.

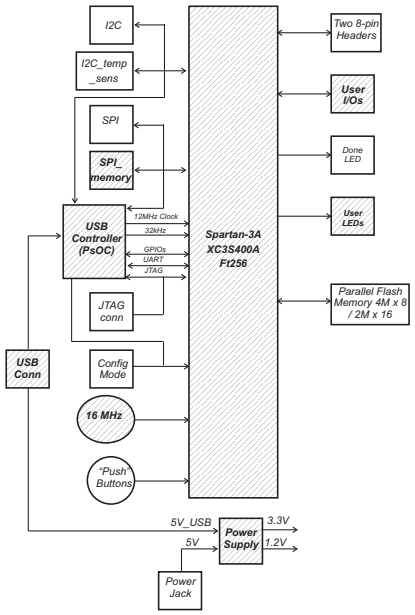


Fig. 3. Spartan 3A Evaluation board.

The block diagram of our DDS function generator development is presented on Fig. 4.

As it shown the FPGA contains the DDS core which generates different waveform signals, according to our objectives. The generated signal is transmitted through UART interface to the PSoC for further processing.

The PSoC itself contains all the necessary blocks for digital to analog conversion, filtering and amplifying the output signal. Moreover the embedded USB interface is used for communication with a PC. The control unit block decodes the commands sand by a PC. Than it adjust the gain of the PGA or transmit a command to the FPGA for frequency tuning or signal form. The programming of the PSoC is done on C programming language. The algorithm flowchart of the control unit is shown in Fig. 5.

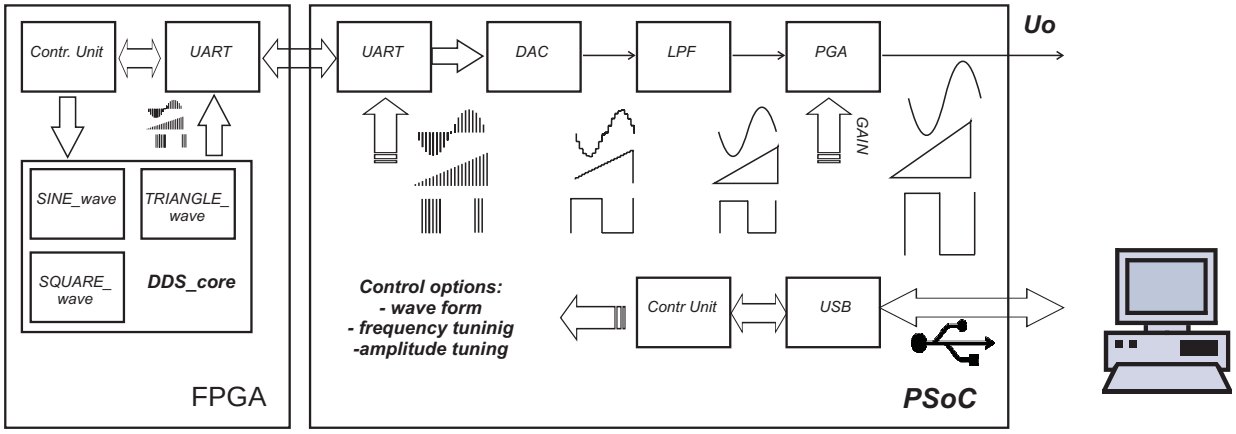


Fig. 4. Block diagram of DDS function generator.

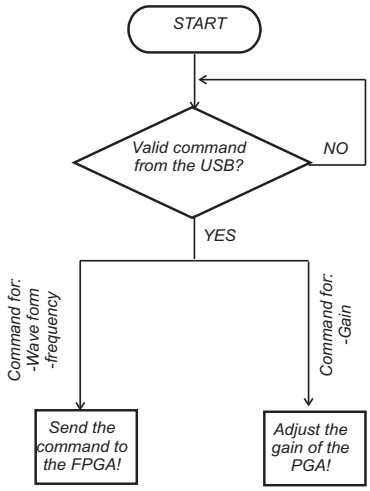


Fig. 5. Control unit flowchart.

The DDS core block from Fig. 4 incorporates three sub blocks. Each of these blocks produces one of the three output shapes – sine, triangle or square. The sine wave sub block has the block scheme represented in Fig. 6.

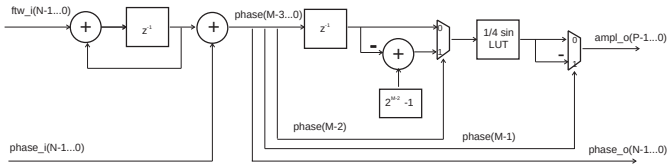


Fig. 6. Sine wave sub block.

As it can be seen from the figure only one quarter of the sine-wave is stored in the LUT, the rest is computed by simple operations (negating, subtraction), resulting in a reduced memory requirement. The resolution of the frequency tuning word (FTW), the phase and the amplitude can be defined separately [10].

A screen shot taken from PSoC Designer IDE (Cypress inc.) is presented in Fig. 7. As is shown in the "Analog Configuration Blocks" section the 8-bits DAC is connected with a second-order anti-imaging LPF. The filtered out signal is passed to a PGA for an output signal amplitude adjustment.

On the other hand the digital section contains an UART interface for communication with the FPGA. Also a 16-bits Timer for FPGA reset impulse production is implemented. The reset impulse duration is about 1*s* .

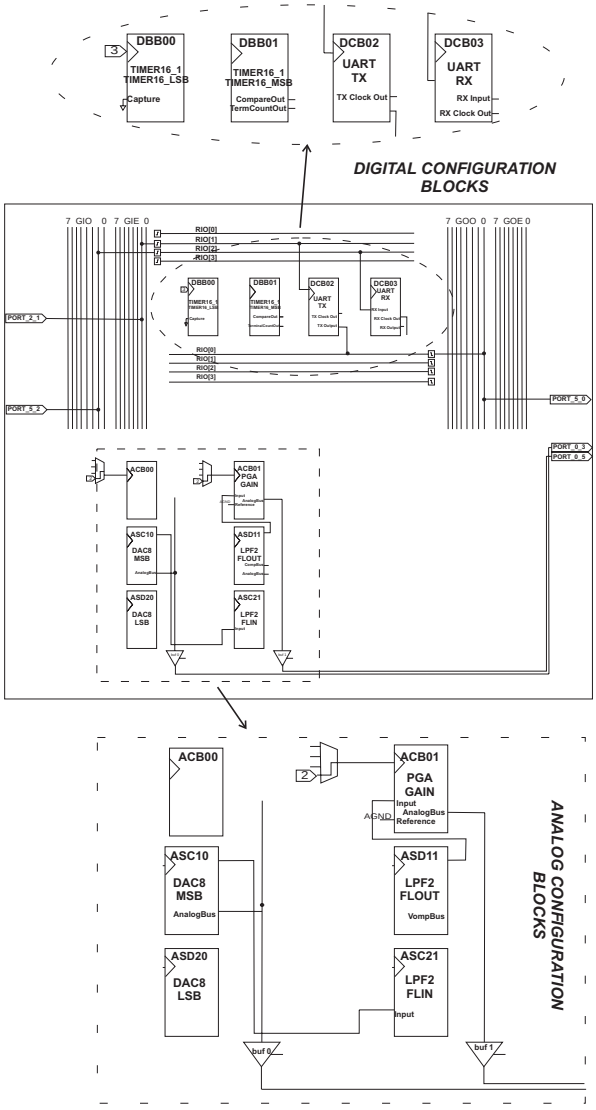


Fig. 7. PSoC Designer IDE – DDS project screen shot.

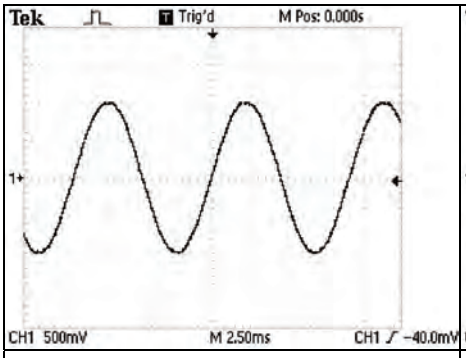


Fig. 8. Sinusoidal output signal.

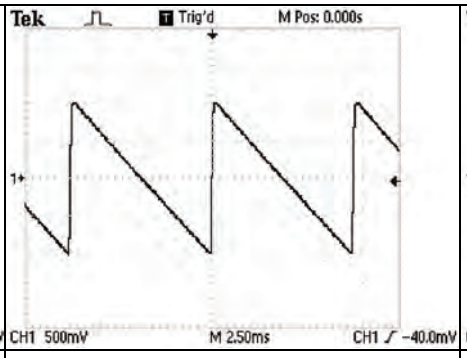


Fig. 9. Triangle output signal.

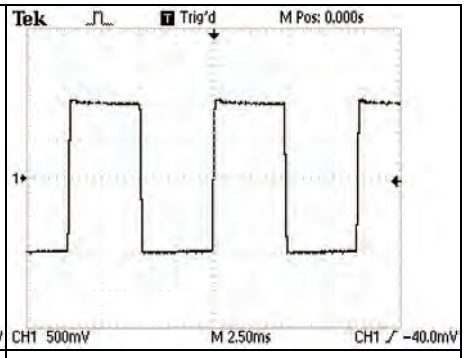


Fig. 10. Square output signal.

IV. Experimental Results and Discussions

This chapter presents some oscillograms from the experimental studies. The following technical equipment was used: Xilinx Spartan – 3A Evaluation Kit; Tektronix 1002B Oscilloscope (oscilloscope setup: AC mode and 10X probe attenuation); Personal computer.

Fig. 8 through Fig. 10 shows the sinusoidal, triangle and square output signals respectively. The signals amplitude and frequency are $1V$ and $114Hz$, respectively.

After analyzing the results a comparative table between the desired and achieved results could be make (Table 1).

TABLE I
COMPARISON BETWEEN DESIRED VALUES AND ACHIEVED EXPERIMENTAL RESULTS

Parameter	Desired	Achieved
Output signal shapes	*sine, triangular, square*	*DONE*
Frequency band	*0 – 2 MHz*	*0 – 200 Hz*
Frequency tuning method	*digitally / minimum 8bits resolution*	*DONE*
Amplitude tuning method	*digitally*	*DONE – 16 levels*
Output signal amplitude	±10V	0 - 3.3V
THD:	0.1%	*Future study*
PC connectivity	*via standard USB ports*	*DONE*

The table reveals the main problems of the development:

Problem 1 - Frequency bandwidth:

The achieved frequency bandwidth is in orders lower than the desired. Our study and analysis shows that the root cause is the low data transfer rate between the FPGA and PSoC. The UART, connecting the two ICs, appears to be constrained to about 37,5 kbps. Higher transfer rates lead to data errors.

Solution: Using another connecting protocol and finding better communication lines between the FPGA and PSoC may improve the results.

Problem 2 - Output signal amplitude:

The achieved output signal is unipolar and with small amplitude. The explanation is clear – the output of the PSoC is directly connected to the output of the system.

Solution: Buffering and amplifying the signal from the PSoC will resolve the problem.

IV. Conclusion

In conclusion we can restate that DDS is very modern and pliable technology. It allows flexible control of all the parameters of interest. Moreover the technology is not resource demanding and is convenient for implementation in microprocessor and mixed signal systems. Also using a FPGA and PSoC ICs greatly reduces the design workload and makes the system even more flexible.

By our personal opinion the objectives stated in chapter III are fulfilled sufficiently. Of course, the main problems, described in chapter IV, should be addressed in future developments.

Acknowledgement

This paper is a part of a project under contract № 102pd209-3/2010, which is sponsored by the research program of the Technical University of Sofia, Bulgaria.

References

[1] W. Kester, *Analog-Digital Conversion*. Analog Devices, MA, 2004.
[2] V. Tietze, Ch. Schenk. *Halbleiter-Schaltungstechnik.* 12. Auflage. New York. Springer-Verlag, 2002 (in German).
[3] M. Ferrari, V. Ferrari, K. Kanazawa, Dual-harmonic oscillator for quartz crystal resonator sensors, Sensors and Actuators, vol. 145 (6), pp. 131-138, 2008.
[4] W. Sansen, *Analog design essentials*. New York, Springer 2006.
[5] M. Seifart, *Analoge Schaltungen*. 6 Auflage. Verlag Technik Berlin, 2003 (in German).
[6] A. Amara, F. Amiel, T. Ea, FPGA vs. ASIC for low power applications, Microelectronics Journal, vol. 37, pp. 669–677, 2006.
[7] J. Kalomiros, J. Lygouras, Design and evaluation of a hardware/software FPGA-based system for fast image processing, Microprocessors and Microsystems, vol. 32, pp. 95–106, 2008.
[8] Y.-H. Seo, H.-J. Choi, J.-S. Yoo, D.-W. Kim, An architecture of a high-speed digital hologram generator based on FPGA, Journal of Systems Architecture, vol. 56, pp. 27–37, 2010.
[9] http://www.em.avnet.com/spartan3a-evl
[10] http://opencores.org/project,dds_synthesizer

Modeling and Simulation of Hall Elements

A.T.Aleksandrov[1], P.D. Petrova[2], N. D. Draganov[3]

***Abstract*: In this paper, some Hall element models in view of their applicability for simulation testing in different operation modes are presented. The theoretical-experimental and PSpice structural models of a Hall element have been developed. The conversion characteristics of the discrete semiconductor Hall elements have been simulated. Simulated characteristics have been compared with experimental ones and the results obtained have been discussed**

***Keywords* – Hall elements, model, PSpice simulation.**

I. INTRODUCTION

The wide practical application of Hall elements operating as magnetic field sensors in areas such as electronics, automatics, medicine, automobile industry, etc., requires studying and gaining knowledge about their characteristics and parameters (Hall voltage, control current, magnetic field variation range, and temperature dependences). This is an objective prerequisite for developing suitable models of the elements in view of their applicability for simulation testing of their operation in different operation modes.

The main conversion characteristic of Hall sensors is the dependence of Hall voltage (U_H) on magnetic flux density (B) and the current flowing through the element (I_S), i.e. $U_H=f(I_S,B)$. Object of research interest is the simulation modeling of this characteristic depending on D.C. operation mode and the applied magnetic field.

It is well known [3, 4, 5, 6, 7] that the voltage of discrete semiconductor Hall elements can be described in terms of the dependence

$$U_H = -G\frac{r_H I_S B}{qnd}, \qquad (1)$$

where: **G** - coefficient showing the effect of the geometric dimensions of the element and its recording electrodes, $\mathbf{r_H}$ - Hall coefficient; **d** - depth of the semiconductor plate; **n** - charge carrier density; **q** - electron charge.

[1]Anatolii T. Aleksandrov is with the Technical University, H. Dimitar, str. 4, 5300 Gabrovo, Bulgaria, E-mail: alex@tugab.bg

[2]Pesha D. Petrova is with the Technical University, H. Dimitar, str. 4, 5300 Gabrovo, Bulgaria, E-mail: daneva@tugab.bg

[3]Nikola D. Draganov is with the Technical University, H. Dimitar, str. 4, 5300 Gabrovo, Bulgaria, E-mail: nikola_draganov@mail.bg

II. PRESENTATION

The presence of design and engineering parameters in the mathematical model, which are normally unknown, limits its application in engineering practice. In this respect, the theoretical-experimental models which are synthesized according to characteristics measured at the elements outputs are more correct. To obtain such a model, the planned experiment approach has been applied [1, 2], the control current I_S and flux density B being chosen as factors affecting Hall voltage.

Complete factor experiments of the type $N=2^n$ have been conducted for 4 types of Hall elements (type VHE101, type L1, type D1 and type D2), three observations being carried out in each point [1, 2, 4]. The variation levels of the factors have been determined on the basis of conducted preliminary single-factor experiments for obtaining the conversion characteristics of the Hall elements and they guarantee nominal operation mode. The following values have been assigned for the lower and the upper variation level of the two chosen factors: for I_S – lower level 1mA and upper level 10mA, and for B – 0.1T and 1.5T, respectively.

After the experimental results for the four types of Hall elements have been processed using the methodology in [1, 2], the following general theoretical-experimental model has been obtained

$$U_H = AI_S^a B^b, \qquad (2)$$

Where: **A**, **a** and **b** are constants whose values depend on the type of the specific element and its D.C. operation mode (Table 1).

TABLE I

Hall element	**A**	**a**	**b**
VHE101	399,7903	0,7519	0,8560 - 0,0111.ln(Is)
L1	462,5437	0,7175	0,4539 - 0,0133.ln(Is)
D1	482,1914	0,7057	0,4767 - 0,0139.ln(Is)
D2	233,7463	0,7002	0,1985 - 0,0441.ln(Is)

The theoretical-experimental model has been modified in view of its compatibility with the CAD systems of the electronic circuits. For this purpose a simulation modeling approach has been applied on the basis of structural units whose operation is described using suitable analytic dependences. For simulation of the Hall elements conversion

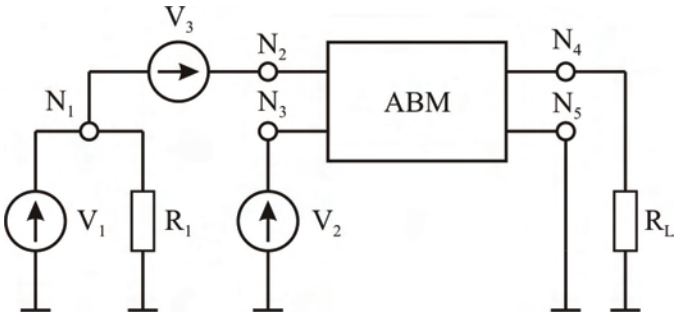


Fig. 1. PSpice structural model of a Hall element

characteristics, an analogue functional block /AFB/ has been used as a structural unit, and for its modeling – a non-linear dependent voltage source is used, which is controlled by the current flowing through the element I_S and by voltage proportional to magnetic flux density B. The generalized PSpice structural model of a discrete semiconductor Hall element is presented in Fig. 1.

The non-linear transfer function of AFB in PSpice format is modeled in the following manner:

Ename N+ N- VALUE = {functional dependence}, Where **Ename** is an arbitrary name of the non-linear dependent voltage source, **N+** and **N-** – node numbers (in sequence +, -), between which the source is connected, **VALUE** – a mandatory key word, after which an analytical dependence is assigned that defines the output voltage of the dependent source as a function of the control quantities, i.e. the non-linear dependence $U_H = f(I_S, B)$. This voltage is actually the Hall voltage.

With view to forming library files with PSpice models of various types of Hall elements, the latter are presented in the form of sub-circuits according to the following general version:

```
.SUBCKT HALLTYPE N1 N3 N4 N5
Ehall N4 N5 VALUE = {A*Ia(V3)*Vb(N3)}
V3 N1 N2 0
.ENDS HALLTYPE
```

The specific type of Hall element is assigned by means of **HALLTYPE**, and the operation of all modeled elements is based on the non-linear dependence described by equation (2). The source V3 has a zero voltage, i.e. it is fictitious. Its existence in the model is justified by the PSpice requirement that the control currents for the dependent sources be defined as currents flowing through voltage sources.

The control current I_S in the structural model in Fig.1 is simulated by means of the D.C. voltage source V_1 and resistor R_1, while the magnetic flux density is simulated by means of voltage source V_2.

To simulate the conversion characteristics of the discrete semiconductor Hall elements, PSpice modules of the following type have been developed:

```
Hall Characteristics
.TEMP 25
R1 N1 0 VALR1
V1 N1 0 VALV1
.STEP V1 LIST <V1_VALUES>
V2 N2 0 DC 0
.DC V2 V2START V2STOP V2STEP
RL N4 0 VALRL
.LIB HALL.LIB
X1 N1 N3 N4 0 HALLTYPE
.PROBE
.END
```

Simulated conversion characteristics for the 4 types of Hall elements have been compared with conversion characteristics obtained experimentally in different operation modes using the possibilities provided by the PSpice integrated medium – Matlab, in order to assess the adequacy of the functional PSpice model. For this purpose the results simulated using the model, which are automatically set up in file filename.out, are edited and saved as a data file filename.dat. This file is set up in Matlab, and its data are read using either the command textread or load. A file containing the experimental results is also set up in Matlab. The two types of results, simulation and experimental ones, corresponding to the same operation mode for a specific element, are processed and visualized. This technology for processing and visualization of the characteristics, allowing the conduction of a precise assessment regarding the correctness of the simulation model, is presented by means of the generalized flow chart in Fig. 2.

Fig. 3 shows simulated and experimental characteristics $U_H=f(B)$ for an element of the type VHE101 in a common coordinate system for 4 different control currents.

III. CONCLUSION

The results show that

- The experimental conversion characteristics and the ones simulated using the functional PSpice model, vary in the same manner.
- Both the Hall voltage and the characteristics steepness rise with the rise in the control current when B=const. Fig. 2 shows that when the flux density B = 0.05T and the control current changes from I_S=0.5mA to I_S= 10mA, i.e. when it increases 20-fold, the ratio between the output voltage obtained through simulation and the output voltage obtained by experiment is almost the same. When control current I_S=0.5mA, the ratio is 1.468, while for I_S = 10mA, it is 1.212, respectively. Under the same conditions, but for B = 0.1T, these ratios are 1.286 and 1.091.
- In case of small and large values of the control current, below 1mA and above 6mA, respectively, the experimental results and the ones simulated using the functional model are very close, and even coincide completely in case of small currents. This statement is proved not only by the graphical dependences in Fig. 3, but also by the one in Fig. 4, which illustrates the variation in the mean-square error obtained through the processing of the measured conversion characteristics and the simulated ones. As can be seen in Fig. 4, when the current flowing through the Hall element

increases up to a certain value, the error increases, after which it decreases. The error is the least (0.03261) for current I_S=0.5mA, and it is the greatest (0.11912) for I_S=5.25mA.

- Considering the fact that Hall elements practically operate in a relatively narrow area of flux density variation, it follows that the accuracy of both the theoretical-experimental model and the PSpice model is really much greater.
- The results are an indisputable proof for the adequacy of the functional PSpice model and the possibility of its practical application for simulation study of Hall elements characteristics and parameters.
- The synthesized model can be used to determine the optimal operation modes of discrete semiconductor Hall elements relative to both control current and magnetic flux density, which is an important foundation for the design of precise sensors.

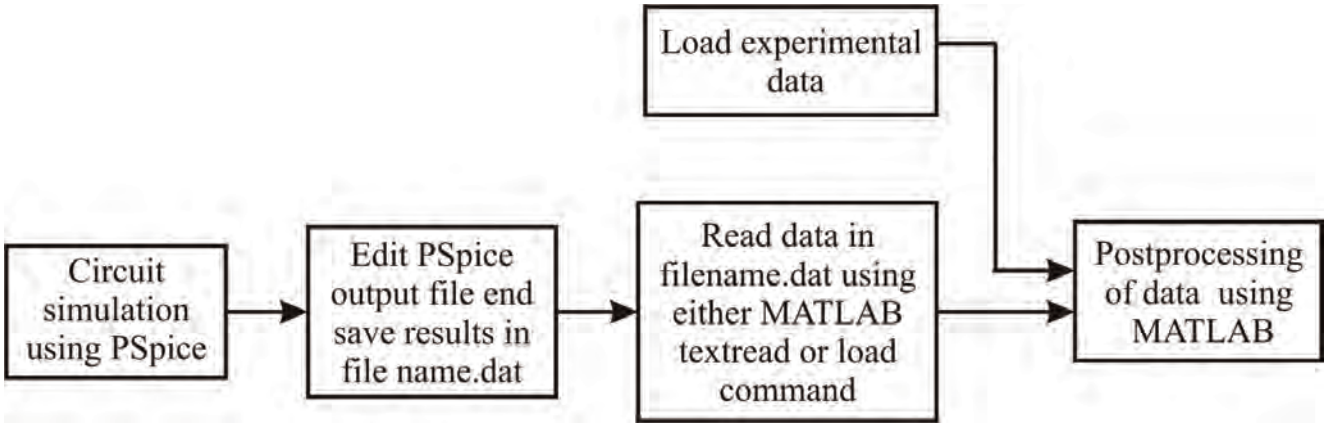


Fig. 2. Flowchart of circuit simulation using PSpice and postprocessing by Matlab

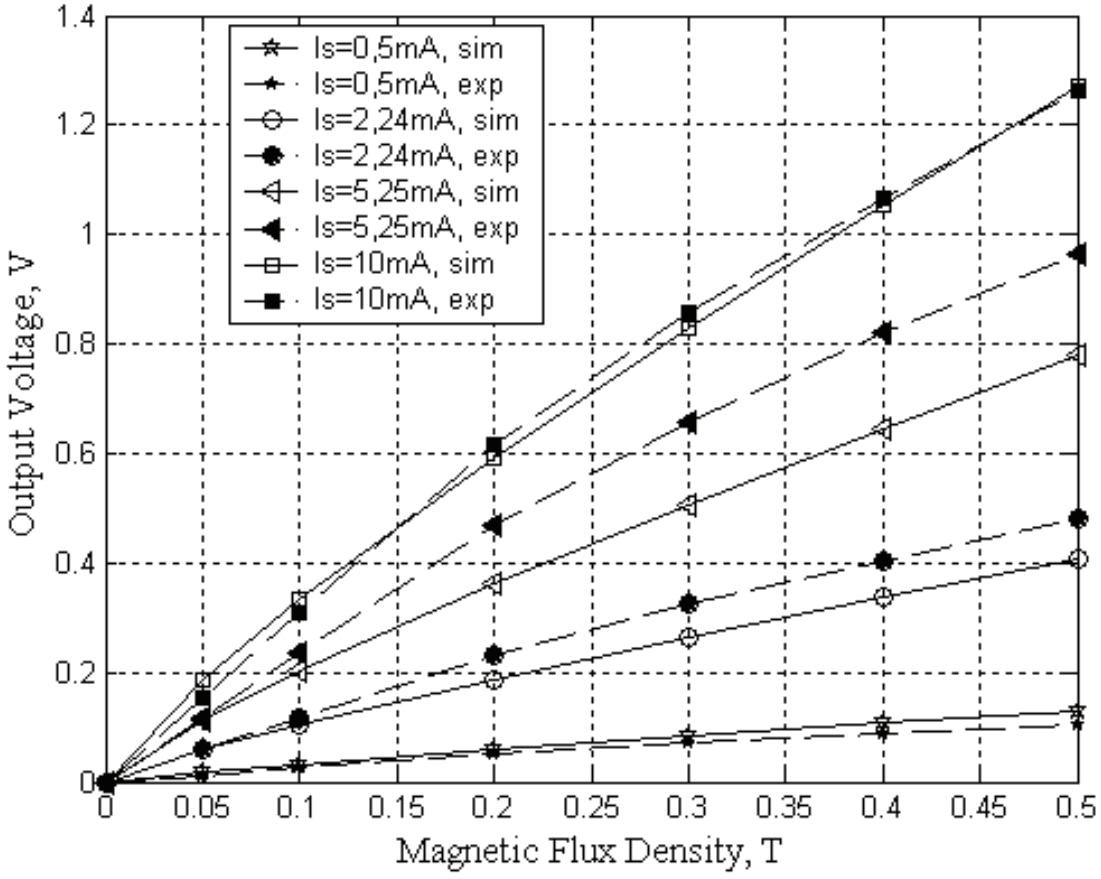


Fig. 3. Simulated and experimental conversion characteristics for element VHE101

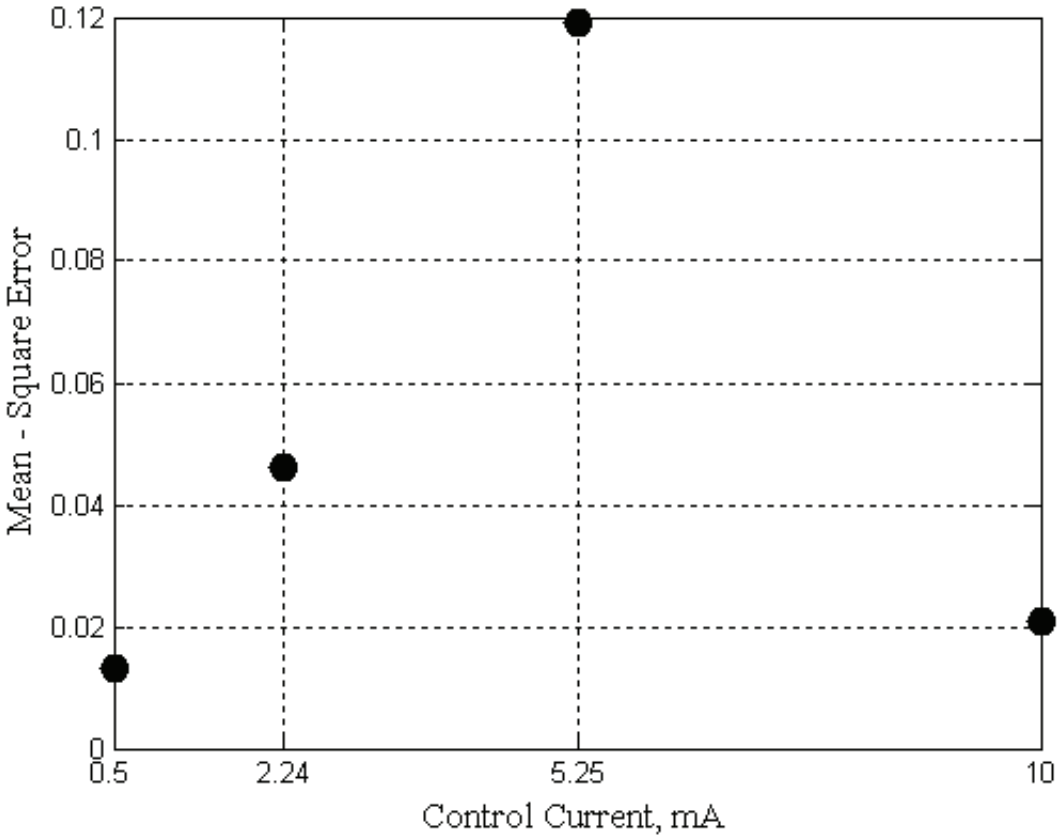


Fig. 4. Mean-square error depending on the control current

REFERENCES

[1] Aleksandrova, I. S. Basis of engineering investigation. University publishing houses "V. Aprilov", Gabrovo, 2003.

[2] Aleksandrov, A.T., N. Draganov. Theoretical-experimental modeling of Hall elements. International Scientific Conference „UNITECH`06", Vol. 1, Gabrovo, 2006, 197-201.

[3] Gurtov, V.A. Solid-state electronics. Moscow, Technosfera, 2008.

[4] Draganov, N., A.T.Aleksandrov. Research over Hall elements. International Scientific Conference „UNITECH 06", Vol.1, Gabrovo, 2006, 193 - 196.

[5] Draganov, N., A.T.Aleksandrov. Research over Hall elements. International Scientific Conference „UNITECH 06", Vol.1, Gabrovo, 2006, 193 - 196.

[6] Draganov, N., A. Aleksandrov. Research of joint work of Hall elements. Book of papers – UNITECH`07. Vol. 1, Gabrovo, 2007, 193-196.

[7] Roumenin, C. S. Solid state magnetic sensors. Amsterdam, Elsevier, 1994.

[8] Takov, T. Semiconductor sensors. Sofia, Techniques, 1986.

Optoelectronic Devices Characterization Using Low Cost Data Acquisition System

Georgi T. Nikolov[1] and Todor S. Djamiykov[2]

***Abstract* – Development of a simple and cost-effective system dedicated for optoelectronic device characterization is reported in present paper. The system consists of low-cost data acquisition module and graphical programming environment. Designed with education in mind, the presented system is a comprehensive tool for teaching the base characteristics of optoelectronic devices.**

***Keywords* – Data Acquisition System, I-V characteristics, LabVIEW, Optoelectronic, Virtual instrumentation**

I. INTRODUCTION

Optoelectronics has become an important part of human lives. Examples include light-emitting diodes in various appliances, photodetectors in elevator doors and digital cameras, and laser diodes that transmit phone calls through glass fibers. Such optoelectronic devices take advantage of sophisticated interactions between electrons and light [4]. For these reasons recently growing number of courses and educational laboratories are developed in this area. The need for enhanced laboratory teaching aids is obvious for better understanding of basic optoelectronic concepts to students. Although there are currently a number of powerful educational products in the market, they are often very expensive and are sometimes inappropriate for low budget educational institutions [1, 7].

On the other hand I-V measurements are made by many various types of customers. Many engineering and physics research facilities use these systems regularly to further measurement theory and component theory. Almost all educational facilities use these systems as a tool to further understanding and research. The applications of these tests are almost limitless. It may seem strange that something seemingly insignificant like the slope of a curve can lead to such incredible conclusions, but the interpretation of the test results is the key to relating this information to something useful. Traditionally, these measurements have been dominated by large rack-mounted measurement systems. These instruments came with a several thousand dollars price tag [7]. The measurement parameters were not nearly as flexible as a computer-based solution. The measurement industry sold these high cost systems each year by providing only small upgrades, but still raising the price with each new version. Today, it is common to see LabVIEW used to automate all these expensive systems. With the combination of low cost hardware and graphical software, a PC-based measurement system could be built that could compete with many of the traditional measurement systems.

The present paper describes an automated system for optoelectronic devices characterization. The system encompasses a personal computer, a set of inexpensive hardware tools for data acquisition and control and graphical programming environment in order to make learning for students attractive and innovative.

The purpose of the laboratory is to invoke student interest by exposing them to various modern tools in comparison to very conventional as well as boring existing optics laboratories. The use of this scientific graphical programming software will help in performing various measurements and calculations with ease. The automation of the experiments will also save great amount of experimentation time and procurement of costly equipments dedicated to each experiment thus providing an efficient way to carry out studies with reduced financial constraints.

II. LED BASIC ELECTRICAL PROPERTIES

Light Emitting Diodes (LEDs) are the most widely used optical sources for optical communication and relevant studies. A light emitting diode works on the principle of spontaneous emission. The regions of operation of LEDs can be identified by studying the two most important characteristics, device current vs. device voltage (I-V) characteristics and optical output power vs. device current (P-I) characteristics.

LEDs are monochromatic devices. The color is determined by the bandgap of the semiconductor used to make them. Red, green, yellow and blue LEDs are fairly common. White light contains all colors and cannot be directly created by a single LED. The most common form of white LED really isn't white. It's an Indium Gallium Nitride blue LED coated with a phosphor that, when excited by the blue LED light, emits a broad range spectrum that in addition to the blue emission makes a fairly white light.

The current-voltage (I-V) characteristic of a p-n junction was first developed by Shockley and the equation describing the I-V curve of a p-n junction diode is therefore referred to as the Shockley equation [4]. The Shockley equation for a diode with cross-sectional area A is given by:

$$I = eA\left(\sqrt{\frac{D_P}{\tau_P}}\frac{n_i^2}{N_D} + \sqrt{\frac{D_N}{\tau_N}}\frac{n_i^2}{N_A}\right)\left(e^{\frac{eV}{kT}} - 1\right) \quad (1)$$

[1]Georgi T. Nikolov is from the Faculty of Electronics and Technologies, Technical University of Sofia, Bulgaria, e-mail: gnikolov@tu-sofia.bg

[2]Todor S. Djamiykov is from the Faculty of Electronics and Technologies, Technical University of Sofia, Bulgaria, e-mail: tsd@tu-sofia.bg

where e is elementary charge, k is Boltzmann constant, T is absolute temperature in K, N_A and N_D are the acceptor and donor concentration, respectively, n_i is the intrinsic carrier concentration of the semiconductor, $D_{n,p}$ and $\tau_{n,p}$ are the electron and hole diffusion constants and the electron and hole minority carrier lifetimes, respectively.

Under reverse bias conditions, the diode current saturates and the saturation current is given by the factor preceding the exponential function in the Shockley equation. Under typical forward bias conditions, the diode voltage is $V >> k\,T / e$, and thus [4]:

$$I = I_S e^{\frac{eV}{MkT}} \tag{2}$$

where M is the ideality factor of the diode. For a perfect diode, the ideality factor has a value of unity. For real diodes, the ideality factor assumes values of typically M = 1.1 – 1.5. However, values as high as M = 2.0 have been found for III–V arsenide and phosphide diodes. Values as high as M = 6.0 have been found for GaN/GaInN diodes. [4]

In addition a diode has unwanted or parasitic resistances. A series resistance can be caused by excessive contact resistance or by the resistance of the neutral regions. A parallel resistance can be caused by any channel that bypasses the p-n junction. This bypass can be caused by damaged regions of the p-n junction or by surface imperfections [2, 5].

III. Hardware Set-up of The Virtual System

The automated characterization system is built using LabVIEW and its compatible Data Acquisition (DAQ) card, NI USB-6008 for interfacing. The experimental set-up and its working for the different characteristic studies are explained in following section.

The National Instruments USB-6008 provides basic data acquisition functionality for applications such as simple data logging, portable measurements and academic lab experiments. It is affordable for student use, but powerful enough for more sophisticated measurement applications [6].

The NI USB 6008 uses last achievement in technology to deliver high-performance and reliable data acquisition capabilities to meet a wide range of application requirements. This device has 12-bit performance on 8 singled-end analogue inputs. The module features digital triggering capability, as well as one 24-bit, 20 MHz counter/timers, and 12 digital I/O lines. It also features two 12-bit analogue outputs.

The experimental set-up is shown in Fig. 1. This schematic shows the test circuit for the two terminal I-V measurement. There are only two components, the device under test (LED in this example) and the load resistor.

Analog Output Channel (AO0) of the NI USB 6008 module is used to supply voltage to the DUT in user defined steps. Volts and the corresponding device voltage is read using Analog Input Channel 0 (AI0). The difference in the two voltages is then used to calculate the current flowing through the series resistor, R which is same as the device current in this case. This current is plotted with respect to the voltage read from AI0 thus generating the I-V characteristic curve for the DUT. The value of the load resistor should be known as accurately as possible for accurate measurements.

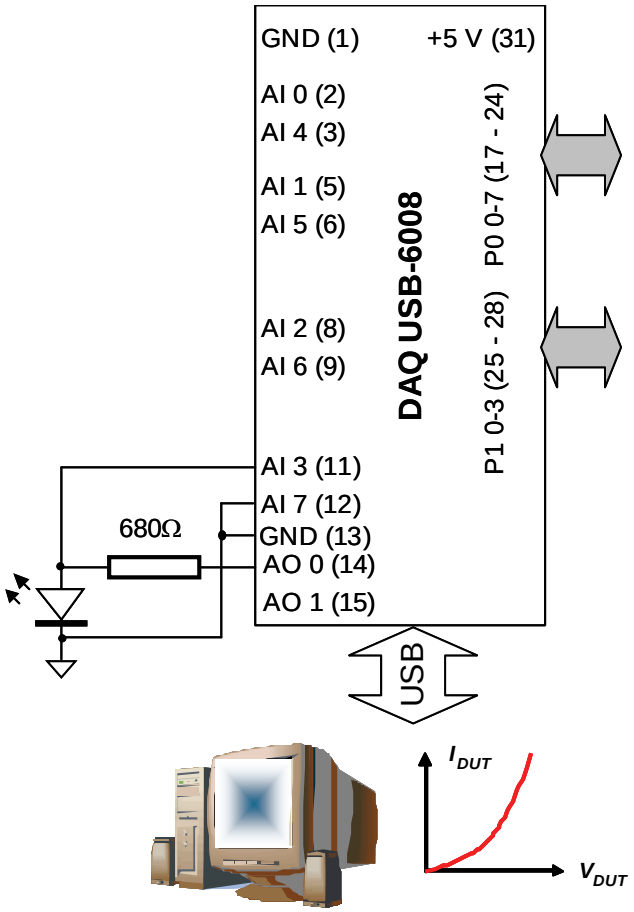


Fig. 1. The components of virtual system

It is known that LabVIEW is a highly productive graphical programming environment that combines easy-to-use graphical development with the flexibility of a powerful programming language. It offers an intuitive environment, tightly integrated with measurement hardware, for engineers and scientists to quickly produce solutions for data acquisition, data analysis and data presentation [3].

One of the more power features of LabVIEW is the build in connection with DAQmx data acquisition device drivers. It is a new type of drivers which architectural changes and new features allow DAQmx to deliver increased ease-of-use and improved performance over the traditional driver. This approach saves development time and improves the performance of data acquisition applications [3]. One of the ways DAQmx saves development time is by providing only a small number of functions to expose. This means that all of the functionality of a multifunction device is programmed with the same set of functions. In LabVIEW, this is possible because polymorphic software component accepts multiple data types for one or more input and output terminals. Another significant feature of the DAQmx architecture is measurement multithreading which means that multiple data acquisition operations can occur simultaneously.

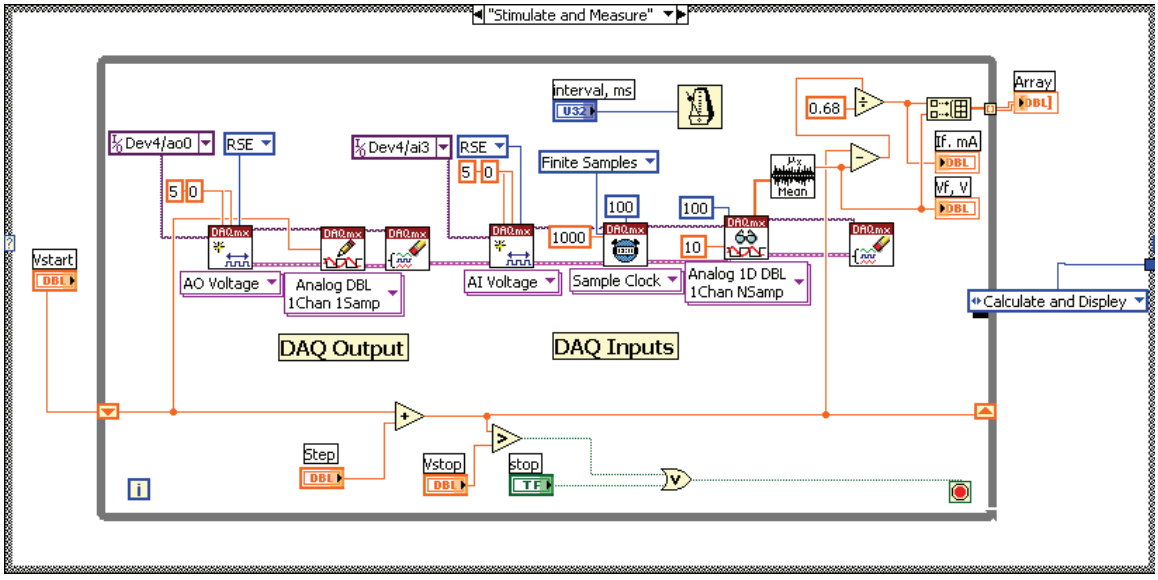


Fig. 2. The block diagram of created virtual system

IV. Software Design and Experimental Results

The programming code in LabVIEW is called block diagram and contains graphical source code. Front panel objects appear as terminals, on the block diagram. Block diagram objects include terminals, subVIs and functions.

The designed block diagram that controls hardware set-up for measuring I-V characteristic of LEDs is shown in Fig. 2. As base design pattern a Standard State Machine is implemented [3]. As can be seen from the Fig. 2. the DAQmx functions is placed in centre of the while loop. There are function controlling DAQ output and DAQ input.

Write function is configured to write one samples of analogue output data for iterations of while loop. A Read function read samples from the specified acquisition task. The voltage difference is calculated and by dividing to load resistance the current is obtained.

In order to illustrate presented approach experimental results acquired from a number of light emitting diodes with different lighting colours as a DUTs were obtained.

The I-V characteristic of white LED is shown in Fig. 3. This is a screenshot of the virtual system for two terminal I-V measurement. The vertical (*Y*) axis is current, and the horizontal (*X*) axis is voltage. The white light emitting diode turns on at approximately 2.7 V. It is possible to determine these values more accurately by using markers. As can be seen from the figure, the measured data follow the theoretical prediction of Eq. (2) quite well for moderate currents, over a range of 0.100 mA to 3 mA. There are some disturbances in the I-V curve caused by noise and limited resolution of ADC. Better results can be achieved with high resolution DAQ.

As can be seen on the Front panel in Fig. 3, there is possibility to store measured data to file. Once stored, this data can be mathematically manipulated further.

In order to determine important parameters such as reverse saturation current, I_S and the ideality factor M the Eq. (2) can be used (e/kT is the equivalent thermal voltage - 22.4 mV for room temperature). These parameters can be obtained using LabVIEW fitting functions more accurately Exponential Fit.vi.

This VI uses the iterative general Least Square method and the Levenberg-Marquardt method to fit data to an exponential curve of the general form described by the following equation:

$$f = ae^{bx} + c \qquad (3)$$

where x is the input sequence X, a is amplitude, b is damping and c is offset. This VI finds the values of a, b, and c that best fit the observations (X, Y).

Front panel of virtual instrument that use this function and measured I-V data from red LED is shown in Fig. 4. The dashed line in Fig. 4 represents the measured data and solid line prediction of Equation in box with parameters I_S = =1.346E-9. These were obtained with a "least squares" fit to the measured data points over that region. With additional mathematical manipulations ideality factor M is achieved. As can be seen from the Fig. 4 a good agreement with fitted equation and measured data is achieved.

V. Conclusion

The presented virtual system is designed for educational purpose and can be used for static characterization of optoelectronic devices, sensors and other semiconductors. In the paper are considered some topics concerning the connectivity between data acquisition board and investigated device and design of programming code. Measured data from the various LEDs showed good agreement with the exponential model of forward voltage and current. The designed system is capable of performing a more of the tests required for the laboratory experiments. Future hardware improvements are expected to increase the accuracy.

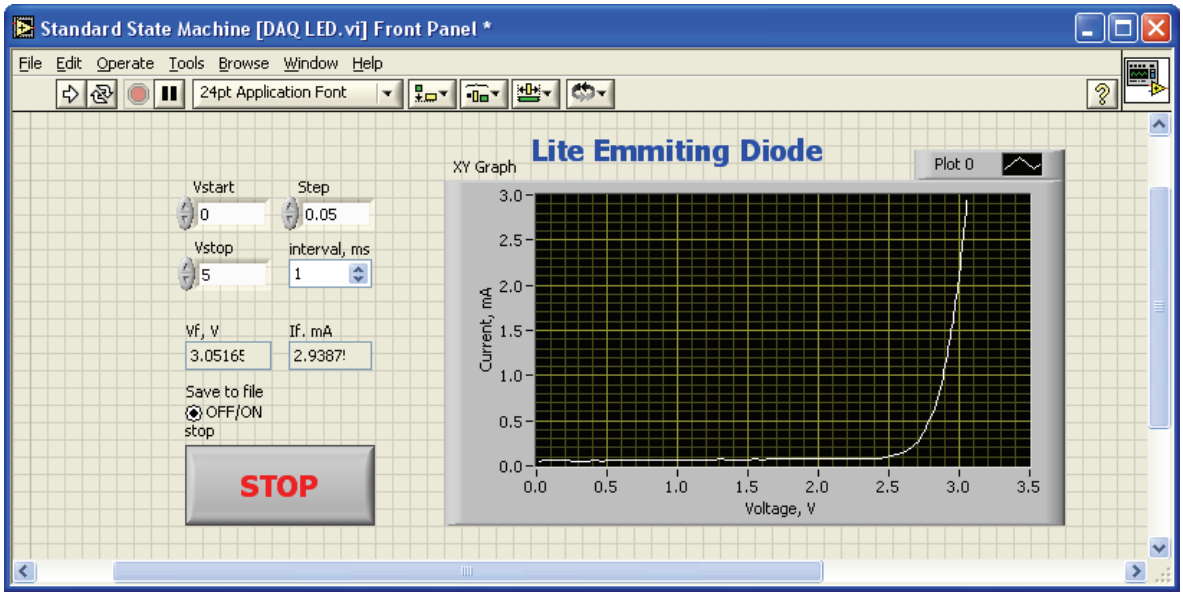


Fig. 3. The I-V characteristic of white LED

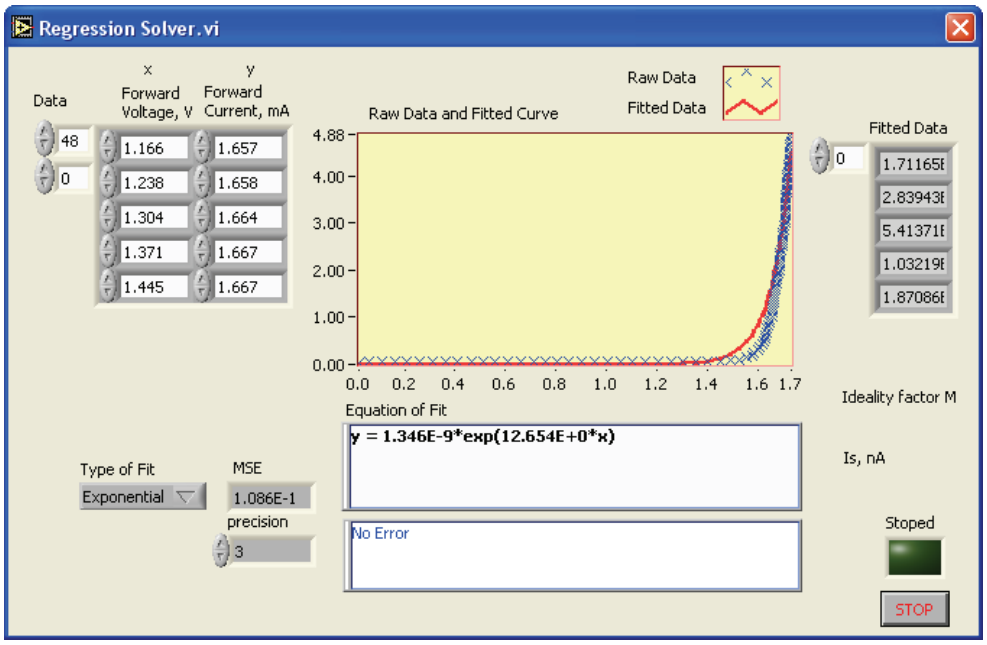


Fig. 4. Front panel of virtual instrument for exponential regression

ACKNOWLEDGEMENT

This investigation has been carried out within the framework of the research project 102ни197-3.

REFERENCES

[1] T. Djamiykov, “Designing an Optoelectronic Characterization System for an Educational Laboratory”, Proceedings of the Technical University – Sofia, Volume 58, book 2, pp 99 - 104, 2008

[2] Sam Mollet, "Analyze LED characteristics with PSpice", GE Harris Harmon Railway Technology, Grain Valley, MO, pp.150-152, EDN, January 2001.

[3] J. Travis, J. Kring, *LabVIEW for Everyone: Graphical Programming Made Easy and Fun*, Prentice Hall, 2006.

[4] E. Fred Schubert, *Light-Emitting Diodes*, Cambridge University Press, ISBN-13-978-0-521-82330-5, 2003.

[5] A. Garg, R. Sharma, V. Dhingra, Development of an Automated Modern Undergraduate Optics Laboratory using LabVIEW, Department of Electronics Acharya Narendra Dev College, University of Delhi Govindpuri, Kalkaji, New Delhi-110019, India, 2008.

[6] National Instruments, USB-6008/6009 User Guide and Specifications, July 2005.

[7] John A. McNeill, Light Emitting Diode Characteristics (Sample Lab Writeup), ECE Box 000, January 19, 1997.

Experimental Galvanomagnetic Transdusers of Linear Offset – Part I

Nikola D. Draganov[1] and Anatolii T. Aleksandrov[2]

***Abstract* – Instrumentation and automatic control need transducers of mechanical quantities which should have high precision, sensitivity and reability fundamentally defined by the magnetosensitive element parameters. An experimental investigations of various sensors constructive solutions of linear offset with open magnetic circuit have been presented and described in the report. Conversion characteristics have been obtained and theoretical-experimental models developed, which illustrate their properties.**

***Keywords* – Hall elements, Offset transducers, Transducers characteristics modelling.**

I. INTRODUCTION

Various areas (spheres) of instrumentation and automatic control need transducers of mechanical quantities which should have high precision, sensitivity and reliability fundamentally defined by the magnetosensitive element parameters [2-4, 6].

The object to this elaboration is to realize, investigate and describe different variants of galvanomagnetic transducers for linear offset based on Hall element of the type VHE101 and to premise their future projection and application.

II. PRESENTATION

Different constructive variants of the magnetosensitive transducers of linear offset with open magnetic circuit based on the synthesized and investigated galvanomagnetic transducer [5] have been developed. They consist of magnetic system, magnetosensitive element (Hall element of type VHE101) and signal processing circuit [5]. They transform the linear alteration L into magnetic field value B alteration which magnetosensitive element converts into electrical signal, i.e. there is a double energy transformation. The magnetic field course of action in different constructions is defined by the permanent magnet disposition according to a Hall element. The investigations have been fulfilled at constant temperature T_O = 25°C on especial installation equipped with standardized watch with precision 0,001mm. The measuring amplifier gain is so picked out that output voltage working range should be closed from – 10V to +10V.

[1]Nikola D. Draganov is with the Technical University, H. Dimitar, str. 4, 5300 Gabrovo, Bulgaria, E-mail: nikola_draganov@mail.bg

[2]Anatolii T. Aleksandrov is with the Technical University, H. Dimitar, str. 4, 5300 Gabrovo, Bulgaria, E-mail: alex@tugab.bg

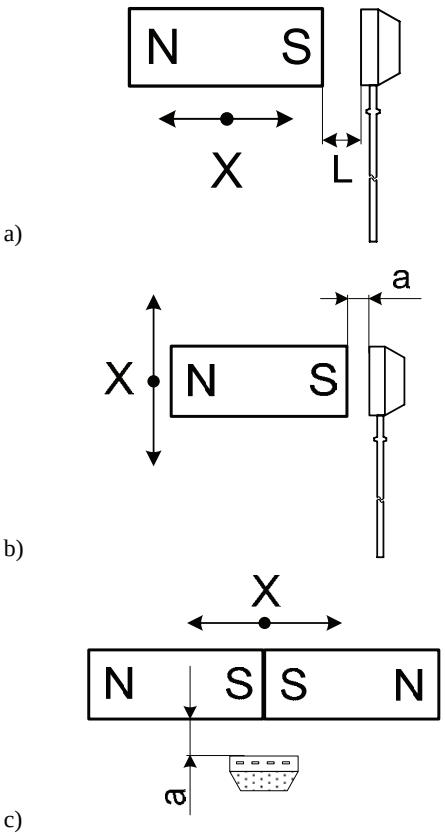


Fig. 1. Galvanomagnetic transducers for linear offset

In case of open magnetic circuit sensor the magnetic field action is one-sided (unilateral) in relation to a Hall element sensitive surface (Fig.1a, b, c). It is necessary to know there are only linear output characteristics if the inductance B is in a sensor linear range. The first constructive variant is depicted in Fig.1a. The permanent magnet is so arranged that its magnetic lines pass across sensitive surface of Hall element which is immobile. When the magnet is moved to a direction of X the distance L between it and Hall element is changed. It is registered by Hall voltage U_O alteration as a result of magnetic field B change effect. Conversion characteristics U_O= f(L) are shown in Fig. 2. They have been measured at constant temperature (T_O = const) at three values of control current. Their analysis shows there is output voltage U_O nonlinear alternating when a distance L is increased. The characteristics slope is not equal in whole investigated range L. For offsets from 2 mm to 6 mm the slope is 0,11; 0,46; 0,6V/mm at a control current I_S=1; 5; 10mA respectively. In

case of offsets from 15mm to 20mm it is 0,04; 0,045; 0,5V/mm at I_S = 1; 5; 10mA respectively.

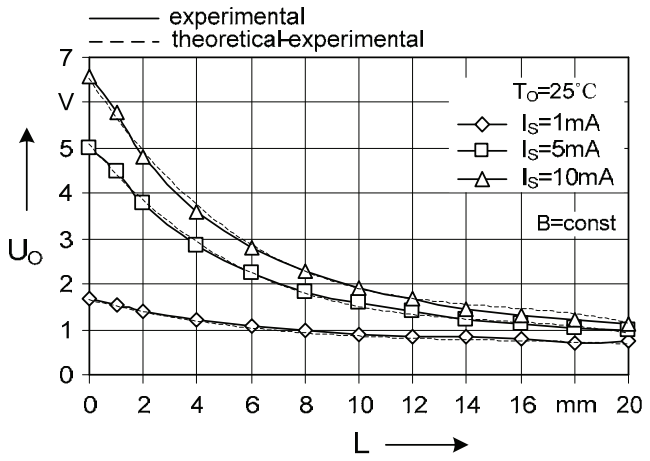


Fig. 2. Conversion characteristics U_O= f(L) for constructive variant in Fig. 1a.

The control current I_S influence on output level U_O decreases when the distance increases. At L=0 output level is U_O = 1,69 to 6,6V at control current I_S = 1 to 10mA respectively. At L = 20mm there is U_O = 0,76 to 1,1V at control currents I_S = 1 to 10mA respectively, i.e. it is fourteen times less in comparison with a case L = 0.

The obtained characteristics have been described by the equation:

$$U_O = a.L^3 + b.L^2 + c.L + d \qquad (1)$$

The model coefficients have been defined by the least squares method [1]. The theoretical-experimenral equations are valid for distance changing from 0 to 20mm. They are:

$$\begin{aligned} U_O &= -0{,}0013.L^3 + 0{,}0589.L^2 - 0{,}9329.L + 6{,}5696, \quad \text{at } I_S\text{=1mA} \\ U_O &= -0{,}0009.L^3 + 0{,}0426.L^2 + 0{,}6851.L + 5{,}0379, \quad \text{at } I_S\text{=5mA} \\ U_O &= -0{,}0002.L^3 + 0{,}0088.L^2 - 0{,}149.L + 1{,}6878, \quad \text{at } I_S\text{=10mA.} \end{aligned} \qquad (2)$$

Constructed on a basis of (2) theoretical-experimental characteristics are depicted in Fig. 2. They present experimental results with ±5% accurasy.

It has been realised and investigated a constructive variant which enables offset watching of magnetic object in parallel to a Hall element sensitive surface direction (Fig. 1b). The Hall element is stationary arranged at a constant distance (a=1mm) from a permanent magnet. The experimental characteristics are depicted in Fig. 3. They are obtained at room temperature (T_O= const), three values of control current I_S = 1; 5; 10mA and distancce changind ΔL=16mm. Their analysis shows that the magnetic object provokes an outpput voltage U_O decrease. It is typically of this construction which generates a same output voltage when there is a offset at the same distances L in both direction of X. Therefore this sensor is not suitable to determination of linear offset direction. The sensor sensitivity is comparable to that determinated for the circuit in [5]. There is a higher sensitivity for offset alterating L from 0 to 2 mm (0,15V/mm at I_S = 1mA; 0,58V/mm at I_S=5mA and 1V/mm at I_S = 10mA respectively) and smaller one for L = 6 ÷ 8 mm (0,05; 0,2; 0,25V/mm at I_S=1; 5; 10mA

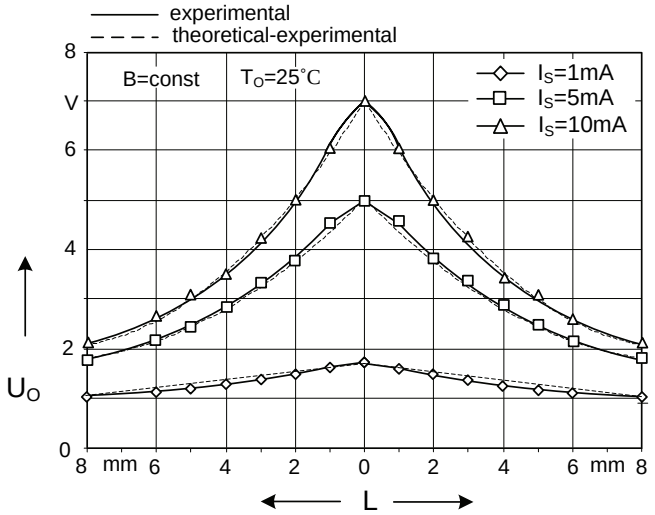


Fig. 3. Conversion characteristics U_O= f(L) for constructive variant in Fig. 1b.

respectively). This trend is valid for both direction of X.

The regressive equation (3) have been obtained by the least squares method [1]. They describe experimental characteristics and results at alteration L from 0 to 8 mm in both directons of X.

$$\begin{aligned} U_O &= -0{,}084.L + 1{,}696, \quad \text{at } I_S\text{=1mA} \\ U_O &= 0{,}038.L^2 - 0{,}707.L + 5.051 \quad \text{at } I_S\text{=5mA} \\ U_O &= 0{,}061.L^2 - 1{,}091.L + 6{,}951, \quad \text{at } I_S\text{=10mA.} \end{aligned} \qquad (3)$$

Theoretical-experimental characteristics are depicted in Fig. 3. They repeat the experimental characteristics with accuracies ±2; ±1,1; ±1% at I_S = 1; 5; 10mA respectively.

The constructed sensor for linear offset depicted in Fig 1c is fulfilled by means of two permanent magnets brought together with homonymous poles. Hall element is fixed immovably at constant distance (a = 1mm) to the magnets. The analysis of the obtained at constant control current (I_S = 1mA), constant temperature (T_O=const) and constant magnetic field (B=const) conversion characteristics (Fig. 4) shows raised output voltage values at small transfers (L = 0 ÷ 1mm). Characteristics are arranged in four quadrants of the coordinates and are

symmetric in first to second and in third to fourth quadrants respectively. Constructed magnetomodulating system defines a conversion chracteristics type and output voltage values.

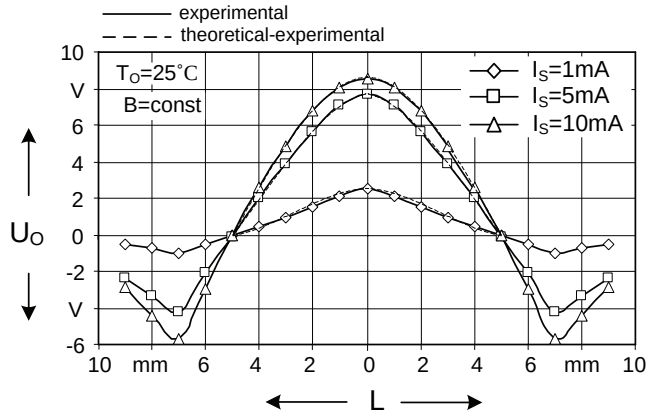


Fig. 4. Conversion characteristics U_O= f(L) for constructive variant in Fig. 1c.

There is a magnetic induction B falsery raisind where homonymoys magnetic poles are in contact. Therefore at L=0 there are the bigest output voltage values (U_O = 8,5V at I_S=10mA). But output voltage U_O is zero when the magnet is moved in distance L = 5mm in both direction of X. In this case as a result of a neutralization of positive and negative magnetic lines the magnetic induction B is zero. Magnetic lines direction and magnetic field action are changed when a distance is increased and became 5 to 7mm. As a result the output voltage changes its sign and has a maximum value U_O=-5,8V at L = 7mm and I_S = 10mA. At L>7mm output voltage U_O decreases again since the magnet is moved away and magnetic field action on Hall element is grown weak. The output voltage alteration is kept for a whole investigated control current range. The characteristics slope is not constant and is dependet on control current value. The highest slope value 2,5V/mm is obtained at control current I_S=10mA and L= 4 ÷ 5mm until the smallest one 0,4V/mm at I_S = 1mA and L= 8 ÷ 9mm. These values are valid in both directions of X. The obtained experimental characteristics can be described by means of equations of type (1). They are a mathematical presentation of output voltage alteration dependet on a transfer L = 0 ÷ 5 mm in both direction of X. The least squares method [1] is used for regressive equations coefficient finding. Then the equations are:

$$U_O = 0{,}023.L^3 - 0{,}195.L^2 - 0{,}051.L + 3{,}121, \quad \text{at } I_S=1\text{mA}$$
$$U_O = 0{,}039.L^3 - 0{,}465.L^2 - 0{,}21.L + 8{,}534, \quad \text{at } I_S=5\text{mA} \qquad (4)$$
$$U_O = 0{,}028.L^3 - 0{,}475.L^2 - 0{,}035.L + 9{,}401, \quad \text{at } I_S=10\text{mA}.$$

Constructed theoretical-experimental characteristics which look like experimental ones with ±1.48% accurasy at the all investigated control currents are depicted in Fig. 4.

III. Conclusion

1. Three constructive variants of galvanomagnetic transducers of linear offset have been realized and investigated.

2. Their experimental conversion characteristics have been obtained and investigated.

It has been established:

- obtained characteristics are symmetrical and disposed in first and second quadrants of co-ordinates;

-constructions for linear offset depicted in Fig.1a and Fig.1c have conversion characteristics with higher slope which is 1V/mm at L = 2 mm and I_S = 10 mA.

3. Theoretical-experimental models of sensor characteristics have been developed which aid their modeling and elaboration.

4. Created consructive sensor variants for linear offset can be applied in instrumentation and autimation. Mathematical equations obtained on the basis of the experimental characteristics aid the sensor modeling and elaboration.

References

[1] Aleksandrova, I.S. Basis of Engineering Investigations. Technical University Issue “Vasil Aprilov”, Gabrovo, 2003.

[2] Takov, T. Semiconductor Sensors. Issue “Techniques”, Sofia, 1986.

[3] Aleksandrov, A., N. Draganov. Monitoring System of Pulsation Processes in a Milking Machine. Proceedings of papers – ICEST, p.711-714, Ohrid, Macedonia, 24-27 June 2007.

[4] Draganov, N., A. Aleksandrov. Galvanomagnetic Antilock Braking System. ELMA, p.134-136, Sofia, 15-16 September 2005.

[5] Draganov, N., A. Aleksandrov. Galvanomagnetic Transducer with Analogue Output. Conference ELEKTRONICA`10, Sofia, to print.

[6] Roumenin, C. S. Solid State Magnetic Sensors, Amsterdam, Elsevier, 1994.

This page intentionally left blank.

Experimental Galvanomagnetic Transducers of Linear and Angular Offset – Part II

Anatolii T. Aleksandrov[1] and Nikola D. Draganov[2]

Abstract **- Instrumentation and automatic control need trancducers of mechanical quantities which should have high precision, sensitivity and reliability fundamentally defined by the magnetosensitive element parameters.**

Conversion characteristics of the sensor for linear and angular offsets have been presented and described. Their theoretical-experimental models have been developed.

Keywords **– Hall elements, Offset and Angular transducers, Transducers characteristics modelling.**

I. INTRODUCTION

Transducers of mechanical quantities are widely applied in instrumentation and automatic control. They should have high precision, sensitivity and reliability fundamentally defined by the magnetosensitive element parameters [2-4, 6].

The object and aim of the present elaboration is to realize and investigate experimental constructive variants of galvanomagnetic transducers of linear and angular offset on a Hall element of the type VHE101 basis and to enable their future project.

II. PRESENTATION

On the basis of the synthesized and investigated galvanomagnetic transducers [5] different constructive variants of galvanomagnetic transducers of linear and angular offset have been elaborated and investigated. They consist of a magnetic system, magnetosensitive element (Hall element of the type VHE101) and signal processing system. Their action is based on energy double conversion. The linear offset L alteration is converted into magnetic field B values change which in a Hall element is transformed into a electrical signal. Three constructive variants created by means of a permanent magnet have been investigated. A magnet arrangement in relation to a Hall element defines a magnetic field influence in different constrictions. The investigations have been fulfilled at constant temperature T_O = 25°C on an especial standard installation equipped with indicative watch with precision 0,001 mm. The measuring amplifier gain is picked out so that an output voltage working interval should be closed from -10V to +10 V.

The investigated sensor constructions are:

- Sensors with closed magnetic systems. In the first version the Hall element is fixed stationary at a permanent magnet pole (Fig. 1a). In the second version it is fixed (Fig. 1b) between homonymous poles of two permanent magnets. The magnetic field action on the Hall element sensitive surface is bilateral (orthogonal);
- Sensor for angular offset with open magnetic system (Fig.1b). A Hall element and control part is immovable. An angle of a magnetic field action is changed.

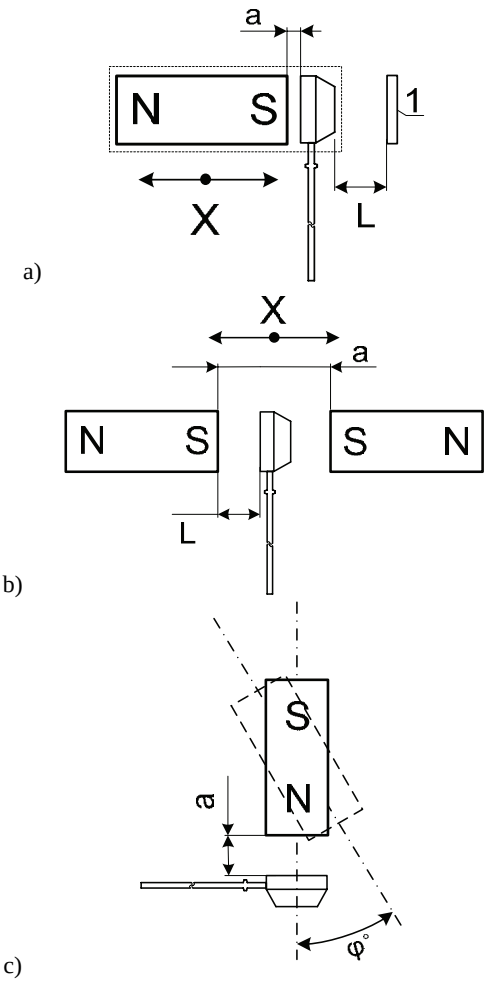


Fig. 1. Galvanomagnetic transducers for linear and angular offset.

It is necessary to have in mind there are linear output characteristics only if an inductance B is in a sensor linear range.

In first constructive variant the ferromagnetic object 1 in Fig. 1a is moved. It has a concentrative role. When this object is moved to a Hall element the magnetic field action is

[1]Anatolii T. Aleksandrov is with the Technical University, H. Dimitar, str. 4, 5300 Gabrovo, Bulgaria, E-mail: alex@tugab.bg

[2]Nikola D. Draganov is with the Technical University, H. Dimitar, str. 4, 5300 Gabrovo, Bulgaria, E-mail: nikola_draganov@mail.bg

increased and is decreased if it is gone away. Experimental conversion characteristics are shown in Fig. 2. Their shape is defined by the magnetic induction value and ferromagnetic substance properties from which the mobile object is produced. The control current I_S influence is equal in whole distance L alteration. The deduced theoretical-experimental equation is:

$$U_O = a.L^3 + b.L^2 + c.L + d \qquad (1)$$

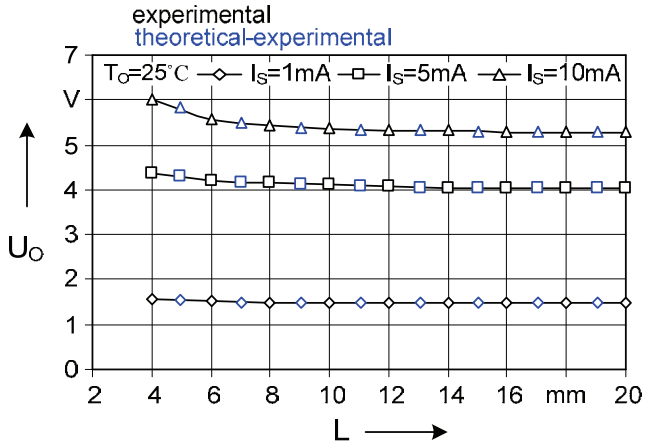


Fig. 2. Conversion characteristics U_O= f (L) for constructive variant in Fig. 1a.

Model coefficients are defined by means of the least squares method. Theoretical-experimental characteristics coincide with experimental ones. The obtained dependence (2) describe with high accuracy the experimental characteristics in the limits of a investigated linear offset L and control current I_S alterations at a temperature T_O = 25°C.

$$U_O = -(0{,}1.L^3 - 4{,}1.L^2 + 56{,}7.L - 1743).10^{-3}, \quad \text{at } I_S\text{=1mA}$$

$$U_O = -(0{,}2.L^3 - 7{,}5.L^2 + 122{,}7.L - 4727).10^{-3}, \quad \text{at } I_S\text{=5mA} \qquad (2)$$

$$U_O = -(0{,}6.L^3 - 25{,}5.L^2 + 370.L - 7071).10^{-3}, \quad \text{at } I_S\text{=10mA.}$$

Another sensor construction for linear offset is depicted in Fig. 1b. A Hall element is arranged between homonymous poles of two permanent magnets. They are immovable each to other at a distance a = 40mm. The permanent magnets disposition provokes with a same polarity and magnitude of magnetic fields simultaneously action on a Hall element both opposite surfaces. When Hall element is placed an equal distance to both magnets (L = 0) the output voltage will be U_O=0 because of the both equal magnetic fields synchronously action. The measured on Hall electrodes voltages separately towards ground will not be zero but will have equal absolute magnitudes |UH2| = |UH4| ≠ 0. When Hall element is moved to one of both directions of X the magnetic field influence will be amplified at another one expense. This will cause a change of the Hall voltages absolute magnitudes |UH2| ≠ |UH4|. Their difference will be amplified by the amplifier.

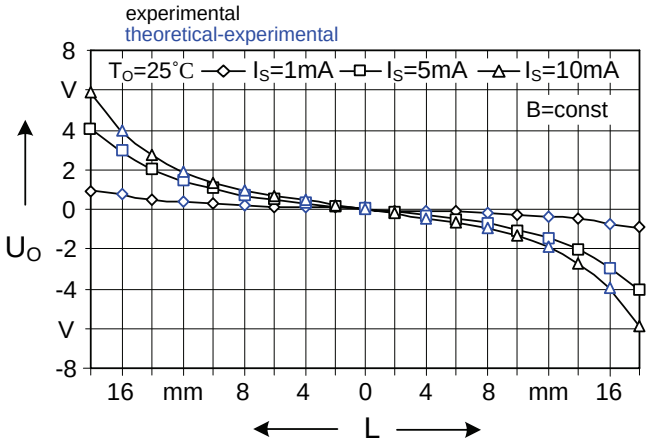


Fig. 3. Conversion characteristics U_O= f(L) for constructive variant in Fig. 1b.

Experimental conversion characteristics U_O = f (L) at I_S=const, B=const and T_O=const are depicted in Fig. 3.

Their analysis shows a characteristics symmetry in second and fourth quadrants of co-ordinates and nonlinearity in whole investigated offset range (L = 0÷20) mm. The obtained experimental characteristics U_O = (L) can be described by means of regressive cubic equations of type (1). Their coefficient are defined by the least squares method. The output voltage U_O alteration in the interval L = (0÷18) mm is described by the following models:

$$U_O = -(0{,}2.L^3 - 8{,}6.L^2 + 144.L).10^{-3} + 1{,}607, \quad \text{at } I_S\text{=1mA}$$

$$U_O = -(1{,}3.L^3 - 47{,}2.L^2 + 676{,}9.L).10^{-3} + 4{,}707, \quad \text{at } I_S\text{=5mA} \qquad (3)$$

$$U_O = -(2{,}3.L^3 - 80{,}8.L^2 + 1058{,}5.L).10^{-3} + 6{,}47, \quad \text{at } I_S\text{=10mA.}$$

Theoretical-experimental characteristics are drawn on the equations (3) basis. Their analysis shows they are fully analogical to the experimental ones.

A constructive variant intended for rotation angle reading, for little slope e. t. is depicted in Fig. 1c. A Hall element is arranged stationary at distance a = 1mm towards the permanent magnet. In according to a sensor purpose the in the center fastened permanent magnet can rotate or fastened in opposite end towards a Hall element can incline. The proposed constructive variant is investigated at a permanent magnet rotation to an angle φ = ±(0÷90)deg. The conversion experimental characteristics U_O = f(φ) obtained at constant

control current (I_S = const) and constant temperature (T_O=const) are depicted in Fig. 4.

Their analysis shows they are nonlinear in whole investigated range φ = ±(0÷90)deg but symmetrical in first

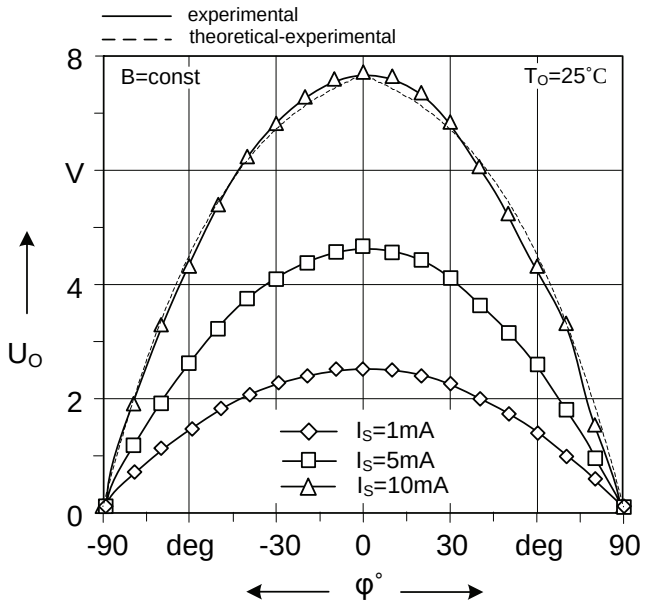


Fig. 4. Conversion characteristics U_O= f(L) for constructive variant in Fig. 1c.

and second quadrants of co-ordinates. The highest slope is received at big angles. For example, the slopes are S_A=0,03; 0,09; 0,065V/deg at I_S = 1; 5; 10mA respectively. At little angular offset φ = (0÷10)deg the slope is S_A = 0,01V/deg at I_S=10mA i.e. it is 6 times less than at big angles.

Regressive equations (4) which describe constructed experimental characteristics at angular rotation φ = (0÷90)deg have been obtained.

$$U_O = -(0{,}2.10^{-3}.\varphi^3 + 90.10^{-3}.\varphi^2 + 3.9.\varphi).10^{-3} + 2{,}621, \quad \text{at } I_S=1\text{mA}$$

$$U_O = -(3.10^{-3}.\varphi^3 + 0{,}2.\varphi^2 + 8{,}5.\varphi).10^{-3} + 4{,}681, \quad \text{at } I_S=5\text{mA} \qquad (4)$$

$$U_O = -(6.10^{-3}.\varphi^3 + 0{,}2.\varphi^2 + 22{,}2.\varphi).10^{-3} + 7{,}759, \quad \text{at } I_S=10\text{mA}.$$

Theoretical-experimental characteristics are depicted in Fig. 4 by doted line. Their analysis shows at control current I_S=1; 5mA these characteristics fully concur to experimental ones and at I_S = 10mA approach with accuracy ±5%.

III. CONCLUSION

1. Three constructive variants of galvanomagnetic sensors for linear and angular offset have been realized and investigated.

2. Experimental conversion characteristics U_O = f (L) of the synthesised galvanomagnetic sensor for linear offset illustrated an output voltage alteration towards a distance and an angular rotation have been investigated.

3. Theoretical-experimental models of these sensors have been developed which aid their modeling and elaboration.

4. Created constructive sensor variants for linear and angular offset can be practically applied in different spheres of science, instrumentation and automatic control and mathematical equations obtained on their experimental characteristics basis aid its modeling and elaboration.

REFERENCES

[1] Aleksandrova, I.S. Basis of Engineering Investigations. Technical University Issue "Vasil Aprilov", Gabrovo, 2003.

[2] Takov, T. Semiconductor Sensors. Issue "Techniques", Sofia, 1986.

[3] Aleksandrov, A., N. Draganov. Monitoring System of Pulsation Processes in a Milking Machine. Proceedings of papers – ICEST, p.711-714, Ohrid, Macedonia, 24-27 June 2007.

[4] Draganov, N., A. Aleksandrov. Galvanomagnetic Antilock Braking System. ELMA, p.134-136, Sofia, 15-16 September 2005.

[5] Draganov, N., A. Aleksandrov. Galvanomagnetic Transducer with Analogue Output. Conference ELEKTRONICA`10, Sofia, to print.

[6] Roumenin, C. S. Solid State Magnetic Sensors, Amsterdam, Elsevier, 1994.

This page intentionally left blank.

Modeling FPGA Logic Architecture

Petar Minev[1] and Valentina Kukenska[2]

***Abstract* – Designing FPGA architectures is a process related to numerous experiments involving different combinations of architectural parameters. The development of analytical models, indicative of the relation between architecture parameters and their impact on the working area, power consumption and FPGA speed, would narrow the range of researched architectures and make designing new FPGA IC families faster.**

***Keywords* – Field-Programmable Gate Arrays, Modelling, Logic Architecture, Logic Blocks, CAD.**

I. INTRODUCTION

A single empirical approach of research and comparison has dominated the practice of developing new or improving already existing FPGA architectures. With it, a number of standard schemes, corresponding to the target schemes that the FPGA family is designed for, are synthesized in different architectures with the help of CAD tools. They allow changing the given architecture parameters. This approach is similar to the one used in the development and research of computer architectures, wherein a standard myriad of software applications are compiled with different processor architectures in order to measure their output.

An alternative to this empirical approach is the use of theoretical methodic, wherein the applications are modeled through statistical and graphical theories, while redesigning and attaching an application to a given architecture is modeled using probability calculations and another theoretical apparatus. The creation of such a methodic has been the object of different studies, usually focusing on a narrow range of architectures and CAD tools. These studies mainly review the tracking architecture and only provide a small amount of information about the logic architecture. The issue for creating and implementing a fundamental and applicable theory in studying FPGA architectures still remains open.

The purely experimental methodic for FPGA architecture research requires a large number of experiments with many different combinations of architecture parameters. The development of analytical models, indicative of the relation between architecture parameters and their impact on the working area, power consumption and FPGA speed, would reduce the number of studied architectures. Once the count of possible solutions is reduced, we can use the traditional experimental methodic to pinpoint the architecture parameters. This would considerably accelerate the while process of FPGA design.

[1]Petar Minev, Department of Computer Systems and Technologies, Technical University of Gabrovo, Phone: +359 66 827 411, E-mail: pminev@tugab.bg.

[2]Valentina Kukenska, PhD, Department of Computer Systems and Technologies, Technical University of Gabrovo, Phone: +359 66 827 456, E-mail: vally@tugab.bg.

A. Architecture of FPGA devices

The architecture of modern FPGA devices is shown on fig.1. It consists of logic blocks, configured in a specific way as to provide for the links among the LBs. The very LBs are also configured in order to perform a given task. Configuration is done by storing configuration data in SRAM cells, which are part of the structure of FPGA IS.

The logic block structure, also called logic architecture is shown on fig.2. It consists of *N* number of basic logic elements with k number of inputs each.

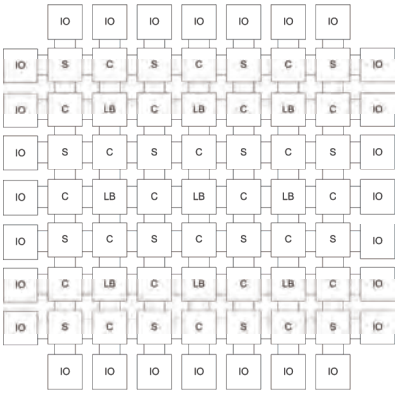


Fig.1. Overview of FPGA architecture

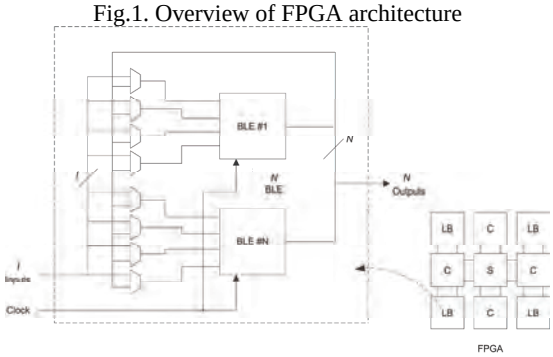


Fig.2. Logic architecture of FPGA

B. Specifying the number of I inputs for a logic block

The number of inputs I can be either limited, i.e. $I < k.N$ or unlimited - $I = k.N$. In most of the FPGA IC families, produced by Xilinx Inc., the number of LB inputs is unlimited and equal to the total number of the BLE elements' inputs in the LB. In the case of the other major manufacturer - Altera Inc., the FPGA devices feature a smaller number of inputs than the overall count of BLE elements' inputs in the LB. Table 1 shows some of the main FPGA IC families of Xilinx and Altera, and their logical architecture parameters.

In Ahmed work, [1] has derived the following analytical

model, specifying the right number of inputs (I) for a single LB, given k and N.

$$I = \frac{k}{2}(N+1) \quad (1)$$

The model has been derived from an attempt to find out the minimal number of inputs (I), necessary to provide for a 98% of LB usage in the FPGA architecture. The average error percentage, as reported by the authors of the model, is 10.1% against a standard deviation of 7.6%.

TABLE I
LOGIC ARCHITECTURE OF FPGA IC FAMILIES, MANUFACTURED BY XILINX AND ALTERA

	Xilinx FPGA Families					Altera FPGA Families					
Parameter	Spartan 3	Virtex	Virtex II	Virtex 4	Virtex 5	Cyclone	Cyclone II	Cyclone III	Stratix	Stratix II	Stratix III
N	8	4	8	8	8	10	16	16	10	8	10
k	4	4	4	4	6	4	4	4	4	6	6
I	32	16	32	32	48	23	36	36	26	38	44

In Betz & Rose works [3] and [4] derived the following dependency between the necessary number of inputs (I) and N if k = 4:

$$I = 2 \cdot N + 2 \quad (2)$$

The authors show this is the right number of inputs I to get a 98% usage of the logic resources in the logic block given that k=4. What's interesting in this case is that you can get a 100% usage when I ranges from 50% to 60% of the overall number of BLE elements' inputs. Similar results are reported in the work of Fang & Rose [7], who explore the dependency between the average number of used inputs per LB (λ) out of the overall number of BLEs (N) in a single LB. They derived the following dependency:

$$\lambda = 0.88N + 3{,}2 \quad (3)$$

This dependency is derived from a regression analysis with the implementation of a 20x20 multiplier in a logic architecture, wherein N varies from 1 to 17 and k = 4. The square of the deduced correlation coefficient for regression dependency (3) is 0,994, which confirms its linear nature. The authors believe this survey is representative and covers the needs of a wide circle of schemes implemented in FPGA logic architecture.

The three models presented above are based on experimental research with different extent of representation. Most representative is the study of Ahmed [1], wherein k is included, apart from N. In this particular research N varies from 1 to 10, and k varies from 2 to 7, which makes a total of 60 different architectures. 28 standard test schemes are implemented in each of these 60 architectures. The parameter I is also configured (changes) in the range from 1 to $k \times N$. The purpose is to find what I can get you 98% of usage of the resources in the logic block.

The other two studies do not aim at deriving a representative model for specifying the necessary number of inputs I per LB and the proposed models cannot be used as universally applicable (for all logic architectures and all schemes implemented in them).

In the study of Lam [9], a theoretical approach is used to deduce an analytical model, which defines the necessary average number of inputs per LB given certain parameters of the logic infrastructure and the FPGA-implemented schemes.

$$i = \frac{(k+1-\gamma)N^p}{1+\frac{1}{f}}, \quad (4)$$

Where γ is the average number of unused BLE inputs and is defined in a table with regard to k, and f is the average number of pins connected to a given output of the entire scheme. The value p is known as Rent's constant. The parameter f is a function of i, N, k and p, which makes model (4) recursive and hard to calculate manually. The model is validated with two algorithms for grouping TV-Pack and iRAC. The results from the model are similar to those experimentally derived from iRAC and are quite different to the ones deduced with TV-PACK. This is understandable given the fact that TV-PACK tries to minimize the number of used BLE elements, whereas iRAC – the number of used LB pins.

C. Rent's rule

Rent's rule was empirically derived for digital IC with average extent of integration that IBM produced back in 1960 [5]. It represents the relation between the number of pins (external connections) P for an area of B number of logic blocks, where each logic block has C number of pins – equation (5).

$$P = CB^p, \quad (5)$$

where p is Rent's constant. For the different types of IC, p is derived experimentally. This rule is successfully applied for specifying the necessary number of pins for different types of IC. Thus, for instance, it has been used to find the right count of pins for the all generations of the Intel's Pentium processors [6].

Though Rent's rule was not derived for FPGA schemes, it represents an interconnection among parameters, which are compatible with those in the FPGA logic architecture. At that stage it looks a proper base in the search of dependency among the parameters in the FPGA logic architecture and of a model that describes it in the right way.

The analysis of the existing models for FPGA logic architecture gives us grounds to separate them in two categories: linear and nonlinear.

The linear models are easy to apply and use, but they do not indicate the extent of complexity of the schemes implemented in the architecture. This is no issue for the nonlinear models, but they are recursive and hard for manual calculation. Further research is needed to find a proper, easy to apply model to indicate the impact of the complexity of the schemes implemented in the logic architecture. A possible solution could be a model, based on Rent's rule, which can be applied to the logic architecture and its parameters. Since the functional dependency in Rent's rule is exponential for proving the hypothesis that the rule is appropriate to apply to the FPGA architecture, it is necessary to check whether the

interconnection among the parameters in the FPGA logic architecture is linear or nonlinear (exponential), and then continue with building a model and its verification.

II. Modeling The Relation Among The Parameters In The FPGA Logic Architecture

A. Making a hypothesis about the relation among the parameters in the FPGA logic architecture

If the relation among the FPGA architecture parameters proves nonlinear, we could model their dependency using Rent's rule (6):

$$P = CB^{p}, \tag{6}$$

where $P=I_{au}+N_{au,}$ $C=k_{au}+1$ и $B=N_{au}$, and p is Rent's contant.

Once we apply logarithm to equation (6), we derive a linear equation of the type:

$$\log(P) = \log(C) + p\log(B) \tag{7}$$

If we explore the relation among parameters *I*, *N*, *k* and it turns out it is of the type (7), we could then continue with deducing a model and verifying it.

Exploring the relation among the FPGA architecture parameters

The methodic used to explore the relation among the parameters in FPGA logic architectures is shown on fig. 3. It is based on the methodic for designing FPGA devices, also used for testing the qualities of new FPGA architectures.

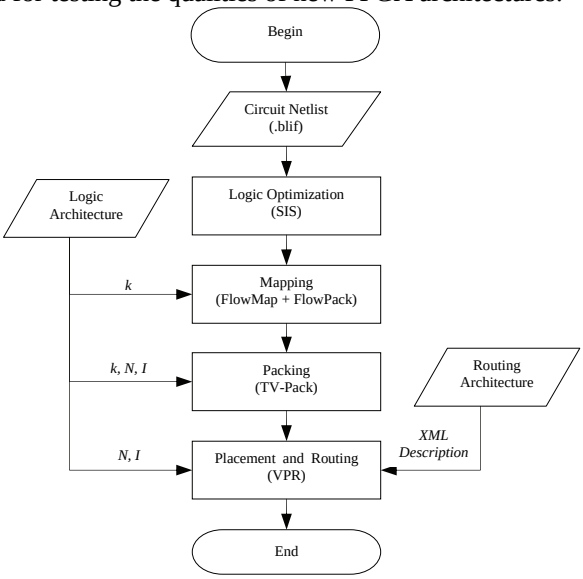


Fig. 3.Methodic for design and testing of FPGA devices

Firstly, the standard test scheme in the shape of a netlist with logic elements and triggers is synthesized and optimized with the help of SIS algorithm [10] in order to derive an optimized netlist with logic element and triggers, too. Once optimized the netlist undergoes decomposition and grouping using the FlowMap and FlowPack algorithms, thus obtaining a number of sub-schemes, each featuring a BLE element from the FPGA architecture. During the packing stage, the TV-Pack algorithm groups the BLE elements in clusters of *N* elements up until the netlist runs out of BLE elements for grouping.

Parameter *k* is input to algorithms FlowMap in order to show the size of the BLE elements and the size of the targeted subschemes. Parameters *k*, *N* and *I* are input to the packing algorithm so that the BLE elements from the decomposed netlist can form clusters of N elements, while observing the maximum number of inputs I in a cluster. An algorithm makes it possible for the factor k to vary in the range of 2 to 7, while N ranges from 1 to 20. This is the range of the current research, too.

The input factor *I* may vary from 1 to a maximum value of *k.N*. In this research no limits for *I* were set and all experiments featured its maximum value of *k.N*. Thus, it is not influencing the output parameters for the examined object. The packing ends with a netlist of LB (clusters), due to be positioned and tracked in the architecture of the FPGA device. The last two stages in the implementing process are performed by the VPR [2] program.

This study takes into account the statistics on the average number of used BLE (N_{au}) and the average number of used LB inputs (I_{au}). These are statistical reports, derived from the tasks of TV-Pack. The research used 10 standard schemes from the test pool MCNC [11], implemented in 120 different logic architectures.

Following the statistical analysis of the accumulated experimental data, as conducted by [8], several conclusions are drawn up:

The input factor *k* influences the output parameter k_{au}, but it does not affect the output parameters N_{au}.

The dispersion analysis also shows that the input factor *N* does not affect the output parameters k_{au}, but it seriously affects N_{au}. The interaction between *k* and *N* also impacts N_{au}. The two input parameters k and *N* also tangibly impact I_{au}. The same applies to the impact between them.

To find out whether the impact of the input factors on the output parameters is linear or nonlinear, [8] included further three output parameters in their research: $\log(N_{au})$, $\log(N_{au} + I_{au})$ и $\log(k_{au} + 1)$, which represent respectively the number of BLE in a single LB, the number of LB pins and the number of pins in a single BLE. The conclusions: *k* and *N* impact $\log(N_{au})$, while their relation is not influencing it; $\log(N_{au} + I_{au})$ is affected by both input factors and their relation; parameter $\log(k_{au} + 1)$ is only influenced by factor *k*.

The regression analysis that followed revealed that using the logarithmic form of output parameters means more tangible functional dependencies (larger R) and considerably reduced percentage of standard errors (fig. 4 ÷ 7). This allows the assumption that the dependency of the output parameters on the input factors in the FPGA logic architecture is nonlinear and their functional dependencies get linear in the process of applying logarithms.

As fig. 6 and fig. 7 show, the regression equations of the derived dependencies are of the type:

$$y = b_0 x + b_1 \tag{8}$$

and correspond exactly to equation (7). This gives us grounds to accept the hypothesis that the number of LB pins in FPGA can be defined using Rent's rule (5). Once accuracy

is improved and the percentage of errors drops, Rent's constant is defined with regard to the number of inputs k in a BLE.

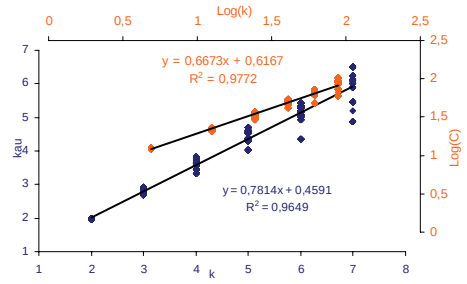


Fig.4. Comparing the dependencies of k_{au} on k and log(C) on log(k).

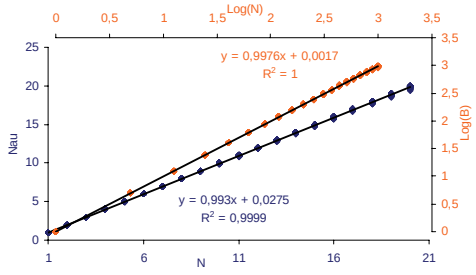


Fig.5. Comparing the dependencies of N_{au} on N and log(B) on log(N).

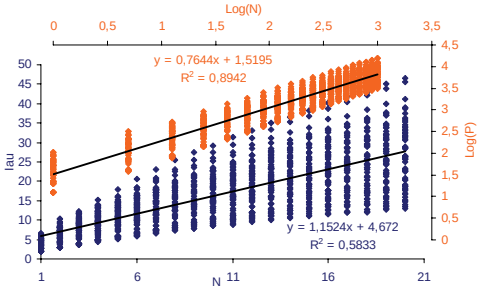


Fig. 6. Comparing the dependencies of I_{au} on N and log(P) on log(N).

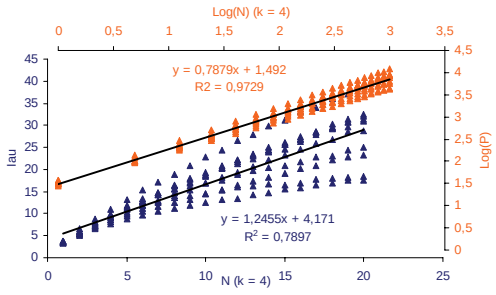


Fig.5. Comparing the dependencies of I_{au} on N and log(P) on log(N), if k = 4.

B. Modelling the relation among the parameters in the FPGA logic architecture

The confirmation of the hypothesis that the number of pins used in a LB of FPGA can be accurately defined with the help of Rent's rule, allows us to specify the necessary number of inputs in a logic block as a function of input factors k and N.

Once we apply algorithms to equation (7) and replace P, B and C with their equivalent parameters from the FPGA logic architecture, we can write down:

$$I_{au} = \left(k_{au} + \frac{N_{au}}{N}\right) N_{au}^{\,p_k} - N_{au}\,, \qquad (9)$$

where k_{au} = f(k); N_{au} = f(N) и p_k = f(k).

III. Conclusions And Future Work

The paper presents an analytical model, describing the FPGA logic architecture. Based on Rent's rule, this model specifies the necessary number of inputs per LB, given the number of BLE in the block and the number of inputs for BLE, as well as Rent's constant, which gives the scheme implemented in the logic architecture.

The model should undergo verification to confirm its reliability and the benefits from its usage.

Acknowledgement

This paper is financed by project: Creative Development Support of Doctoral Students, Post-Doctoral and Young Researches in the Field of Computer Science, BG 051PO001-3.3.04/13, EUROPEAN SOCIAL FUND 2007–2013г. OPERATIONAL PROGRAMME "HUMAN RESOURCES DEVELOPMENT"

References

[1] Ahmed E., "The Effect of LUT and Cluster Size on Deep-Submicron FPGA Performance and Density", University of Toronto, 2001

[2] Betz V. et All, VPR and T-VPack1 User's Manual, University of Toronto 2008.

[3] Betz V. and Rose J., "Cluster-Based Logic Blocks for FPGAs: Area-Efficiency vs. Input Sharing and Size", IEEE Custom Integrated Circuits Conference, Santa Clara, CA, 1997, pp. 551-554.

[4] Betz V. and Rose J., "How Much Logic Should Go in an FPGA Logic Block?", IEEE Design and Test Magazine, Spring 1998, pp. 10-15.

[5] Christie P. and D. Stroobandt, "The interpretation and application of Rent's rule," IEEE Trans. VLSI Syst. (Special Issue on System- Level Interconnect Prediction), vol. 8, no. 6, pp. 639–648, Dec. 2000.

[6] Davis J., V. K. De, and J. D. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI)—Part I: Derivation and validation", IEEE Trans. Electron Dev., vol. 45, no. 3, pp. 580–589, 1998.

[7] Fang W., A Modeling Routing Demand for Early-Stage FPGA Architecture, Master Thesis, University of Toronto, 2007.

[8] Kukenska V., P. Minev, "A Study into the Interconnections of Parameters in the Logic Architecture of FPGA Devices", ETRAN2010, Serbia.

[9] Lam A. et All, "An Analitical Model Describing the Relationships between Logic Architecture and FPGA Density",

[10] Sentovich E.M. et All., "SIS: A System for Sequential Circuit Analysis", Tech. Report No. UCB/ERL M92/41, University of California, Berkeley, 1990.

[11] Yang S., "Logic Synthesis and Optimization Benchmarks, Version 3.0", Tech. Report, Microelectronics Centre of North Carolina, 1991.

Behavioural VHDL-AMS Model for the Current-Feedback Operational Amplifier

Ivailo M. Pandiev[1]

***Abstract* – In this paper linear, frequency-dependent VHDL-AMS model for the Current-Feedback Operational Amplifier (CFOA) is introduced. The model simulates the actual performance of typical CFOAs for a wide range of frequencies. For creating the model, simplification and build-up techniques known from macromodelling operational amplifiers have been adapted. Model parameters are extracted for the CFOA AD8001 from Analog Devices as an example. To confirm the validity, the simulation results are compared with the manufacturer's data and with the results obtained by simulation of the PSpice macromodel supplied from Analog Devices.**

***Keywords* – Analog circuits, Current-feedback operational amplifier, Analog simulation, VHDL-AMS.**

I. INTRODUCTION

The CFOAs, as a variety of operational amplifiers (op amps) family, have been realized to overcome the finite gain-bandwidth product of VFOAs. The CFOAs can be represented as a cascade structure of positive second-generation current conveyor (CCII+) and an additional voltage buffer. Nowadays the CFOAs have emerged as one of the most important active building block in developing many analog and mixed-signal circuits and systems [1-3]. Testing the workability of the developed electronic circuits is usually done using SPICE (Simulation Program with Integrated Circuit Emphasis) with a company specifying macromodel for the CFOA. A variety of SPICE macromodels for the CFOAs, are available in the literature [4-10]. The majority of published SPICE based macromodels are presented as *device-* (*transistor-*) *level models*, called micromodels. Typically the input stage of the micromodels is a *device-level model* and other stages consist of an extensive number of passive elements, ideal controlled sources and ideal diodes.

Testing a complete analog system via transistor-level simulation is an extremely difficult process and can often become infeasible due to the limitation of simulation capacity. A similar difficulty is encountered when high-level design is performed for the whole system. For these reasons, compact models of analog blocks, and in particular, op amps are desired which can be substituted in place of the actual transistor-level netlist to speed up the simulation without sacrificing any of the required accuracy.

One method to decrease simulation time and improve the convergence, without a significant loose of information, is by using behavioural modelling technique. Behavioural modelling is a way of providing macroscopic models of the corresponding microscopic (microelectronic) circuits. The use of behavioural modelling for op amps has been well known for a few years. Behavioural models are realized by using structural macromodelling, the C code modelling, Analog Behavioural Modelling (ABM) available in PSpice A/D simulators and finally the behavioural modelling with VHDL-AMS.

Nowadays one of the most effective techniques for behavioural modelling of analog and mixed electronic circuits is by using VHDL-AMS. The conveyed literature survey reveals that behavioural models for the CFOAs have not been available yet. Without any doubt behavioural models for basic types of CFOAs, are necessary for simulating complex circuits. The goal of this paper therefore is to develop a simple behavioural VHDL-AMS model that accurately simulates the frequency performance of most common CFOAs.

II. CFOAS

The most common CFOA is equivalent to a CCII+ plus an output voltage buffer. These op amps have a high impedance non-inverting input y, a low-impedance inverting input x, a current output z and the voltage output o. In some of the CFOAs the port z, between the first stage (CCII+) and the second stage (voltage buffer), is defined as an external pin. This allows the usage of CFOAs in some specific selective amplifier circuits, sinusoidal oscillators and multivibrators. The port o is the output of the voltage buffer, where the resistance is very low (magnitude of several ohms). The simplified model of the CFOA is presented in Fig. 1. This model contains three parameters to be determined by measurements, i.e. α_{yx}, the y to x voltage gain, α_{xz}, the x to z voltage gain and α_{zo}, the z to o voltage gain. All these parameters are ideally equal to 1. For this model the general relation between input and output voltages and currents can be given as $i_y = 0$, $v_x = \alpha_{yx} v_y$, $i_z = \alpha_{xz} i_x$ and $v_z = \alpha_{zo} v_o$. If pin z is defined as an external port α_{yx} is measured by app-

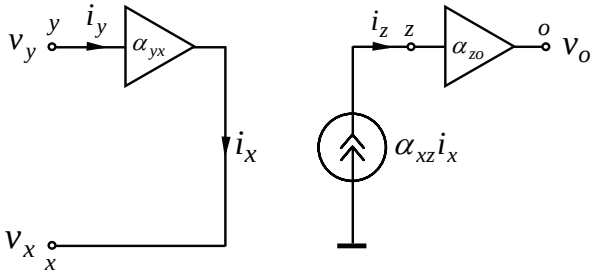


Fig. 1. A simple model of the real CFOA.

[1]Ivailo M. Pandiev is with the Faculty of Electronics, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ipandiev@tu-sofia.bg

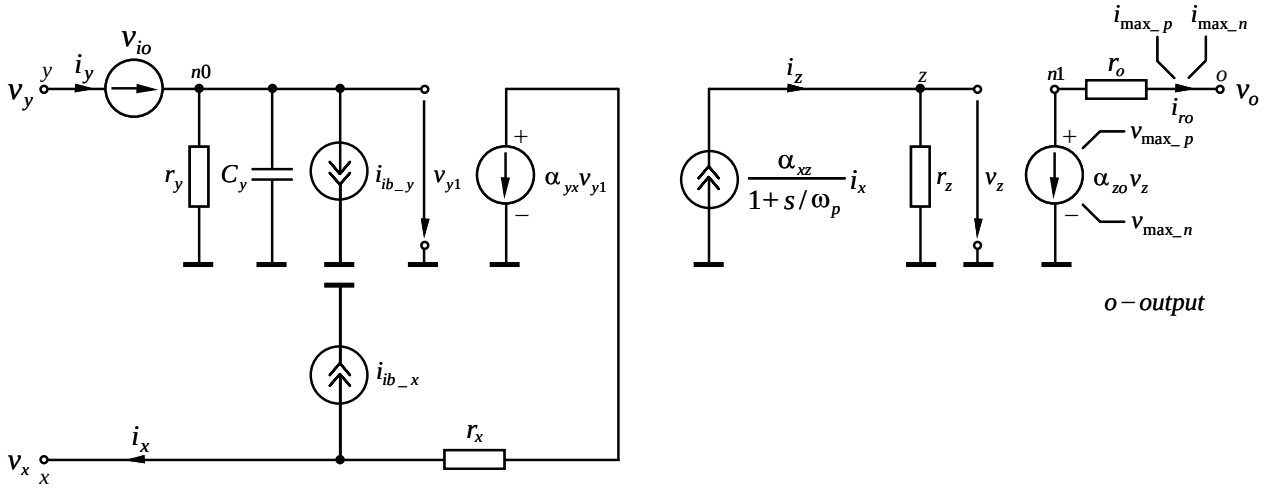


Fig. 2. Circuit diagram of the Current-Feedback Operational Amplifier (CFOA) behavioural model.

lying a voltage signal at port y and measuring the resulting voltage at port x. α_{xz} is measured by injecting a current signal at port x and measuring the output current at port z. Whereas α_{zo} is measured the same way as α_{yx}.

III. Behavioural Modelling with VHDL-AMS

The technical requirement for effective models is generally agreed when the simplest model possible is developed. Simple models have a number of advantages. They can be developed faster, are more flexible, require less data, run faster and it is easier to interpret the results since the structure of the model is better understood. As the complexity increases, these advantages are lost [11]. The proposed behavioural model of the CFOA is developed following the design method based on a Top-Down analysis approach and applying simplification and build-up technique. The process of model building and testing can be broken down into three basic steps: *structure the model*, *build the model* and *validate the model* [12].

A. A behavioural language: VHDL-AMS

VHDL-AMS is a comparatively new standard 1076.1 of VHDL that support hierarchical description and simulation of analog, digital and mixed-signal applications with conservative and non-conservative equations [13, 14]. On the analog side, a variety of abstraction levels is supported. The VHDL-AMS modelling is not restricted to electronic circuits, but also supports thermal, mechatronic, optical and other systems.

B. A CFOA behavioural VHDL-AMS model

The behavioural model of CFOA is built using the electrical characteristic curves obtained by analyses of the PSpice macromodel for AD8001AN [15]. In fact, the monolithic op amp AD8001AN is a typical representative of the commercially available CFOAs. The circuit diagram of the new CFOA model is shown in Fig. 2, where the different stages are presented with controlled sources and passive components.

The proposed model including effects such as (1) accurate inverting and non-inverting input impedance, (2) offset voltage and input bias currents, (3) dc transfer characteristic, (4) ac small-signal frequency response, (5) transient response under a wide range of conditions, (6) input and output voltage and current limitation and (7) output resistance.

The mathematical equations that describe the model can be given as

$$i_y = \left(v_y - v_{io}\right)\left(\frac{1}{r_y} + sC_y\right) \tag{1}$$

$$i_{ib\_y} = i_{ib} + i_{io}/2 \tag{2}$$

$$v_x = \alpha_{yx}v_{y1} - r_x i_x \tag{3}$$

$$i_{ib\_x} = i_{ib} - i_{io}/2 \tag{4}$$

$$i_z = \frac{\alpha_{xz}}{1+s/\omega_p}i_x \tag{5}$$

$$v_z = i_z r_z \tag{6}$$

$$v_o = \begin{cases} v_{max\_p} - r_o i_{ro}, & v_z \geq v_{max\_p} \\ \alpha_{zo}v_z - r_o i_{ro}, & v_{max\_n} < v_z < v_{max\_p} \\ v_{max\_n} - r_o i_{ro}, & v_z \leq v_{max\_n} \end{cases} \tag{7}$$

where $v_{y1} = v_y + v_{io}$ and ω_p is the $-3dB$ radian frequency for the dominant pole of a typical CFOA.

Short-circuit current limiting is simulated with model parameters i_{max_p} and i_{max_n}. In short-circuit mode the output current i_{ro} is limited to one of the values i_{max_p} or i_{max_n}.

Fig. 3 shows the behavioural VHDL-AMS model of CFOA. The library clause and the use clause make all declarations in the packages `electrical_systems` and `math_real` visible in the model. This is necessary, because the model uses nature `electrical` from package `electrical_system`

```
library IEEE;
use IEEE.math_real.all;
use IEEE.electrical_systems.all;
entity      CFOA is
generic (
  iib   : CURRENT := -2.0e-6;       --input bias current
  iio   : CURRENT := 4.0e-6;        --input offset current
  vio   : VOLTAGE := 2.0e-3;        --input offset voltage
  ry    : RESISTANCE := 10.0e6;     --positive input resistance
  cy    : CAPACITANCE := 1.5e-12;   --positive input capacitance
  rx    : RESISTANCE := 50.0;       --negative input resistance
  rz    : RESISTANCE := 900.0e3;    --transresistance
  ro    : RESISTANCE := 50.0;       --output resistance
  ayx   : REAL := 1.0;              --the Y to X voltage gain
  axz   : REAL:= 1.0;               --the X to Z current gain
  fp    : REAL:= 700.0e3;           --fp of gain
  azo   : REAL:= 1.0;               --the Z to O voltage gain
  v_max_p : VOLTAGE := 5.0;         --max pos. output voltage
  v_max_n : VOLTAGE := -5.0;        --max neg. output voltage
  i_max_p : CURRENT := 110.0e-3;    --max pos. output current
  i_max_n : CURRENT := -110.0e-3); --max neg. output current
  port (terminal y, x, output : electrical);
end entity CFOA;
architecture arch_CFOA of CFOA is
-- inner terminals
  terminal n0, n1, z: electrical;
-- inner branch quantities and free quantities
  quantity vy across          y to electrical_ref;
  quantity v_io across i2 through y to n0;
  quantity vy1 across iy, icy, iib_y through n0 to electrical_ref;
  quantity vx across ix, iib_x through x to electrical_ref;
  quantity iz through electrical_ref to z;
  quantity irz through z to electrical_ref;
  quantity vz across          z to electrical_ref;
  quantity vo1 across io1 through n1 to electrical_ref;
  quantity vro across iro through n1 to output;
  quantity voutput across output to electrical_ref;
  quantity iro_h : current;
-- constants
  constant wp: REAL := fp * math_2_pi; -- -3db frequency in radians
  constant num: REAL_VECTOR := (0 => axz);
  constant den: REAL_VECTOR := (1.0, 1.0/wp);
begin
--***input stage***--
  v_io == vio;
  iy  == vy1/ry;
  icy == cy * vy1'dot;
  iib_y == iib + iio/2.0;
  vx == ayx * vy1 + ix * rx;
  iib_x == iib - iio/2.0;
--***transfer stage***--
  iz == ix'ltf(num, den);
  irz == vz/rz;
--***output stage***--
  iro_h == vro/ro;
--limitation of the output voltage
  if vz'above(v_max_p) use vo1 == v_max_p;
      elsif not vz'above(v_max_n) use vo1 == v_max_n;
    else vo1 == azo * vz;
end use;
--limitation of the output current
if iro_h'above(i_max_p) use iro == i_max_p;
      elsif not iro_h'above(i_max_n) use iro == i_max_n;
    else iro == iro_h;
 end use;
end architecture arch_CFOA;
```

Fig. 3. A CFOA behavioural VHDL-AMS model.

and constant `math_2_pi` for the value of π from package `math_real`. The proposed CFOA model is composed by an *entity* and an *architecture*, where bold text indicates reserved words and upper-case text indicates predefined concepts. The entity declares the generic model parameters and specifies three interface terminals of nature `electrical`. The parameters are given with concrete numerical values for the CFOA AD8001AN. The proposed CFOA model includes the following electrical terminals: non-inverting input – `y`, inverting input – `x` and output – `o` (or `output`). Furthermore, the model has three inner terminals: `n0`, `n1` and `z`. They are used to specify the voltages `vy1`, `vz` and `vo1`, respectively.

The architecture of the model is subdivided into three parts: *input stage*, *transfer stage* and *output stage*. It contains the implementation of the model. The architecture is coded by using a style combining structural and behavioural elements. The structural description is the netlist of the model and the behavioural description consists of simultaneous statements to describe continuous behaviour.

IV. MODEL PERFORMANCE

The verification of the proposed behavioural model, shown Fig. 3, is performed by comparing simulation results with the manufacturer's data and with the results obtained by simulation of the PSpice macromodel supplied from Analog Devices. The test circuits for simulation modelling are created following the test conditions given in the semiconductor data books of the corresponding CFOA. The power supply voltages of the circuits are chosen to be ±5V. The model parameters α_{yx}, α_{xz} and α_{zo} are approximately equal to 1.

The simulation modelling of the proposed behavioural model of the CFOA was implemented within simulation program System Vision 5.5 (from Mentor Graphics) and simulation modelling of the macromodel AD8001AN/AD was performed with OrCAD PSpice.

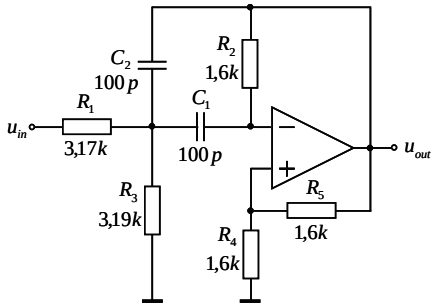
In Table 1 are presented the simulation results and the data sheet parameters for the CFOA AD8001AN. As a result, the proposed behavioural CFOA model corresponds with the parameters of the AD8001AN and it is more accurate at DC and low frequencies than the AD8001AN/AD.

TABLE I
COMPARISON BETWEEN SIMULATION RESULTS AND DATA SHEET

Parameter	AD8001AN Data sheet	AD8001AN/AD macromodel	The proposed CFOA model
DC parameters			
Input offset voltage v_{io}	$2mV$	$5mV$	$2mV$
– Input bias current – i_{ib_x}	$5\mu A$	$6\mu A$	$5{,}01\mu A$
+ Input bias current – i_{ib_y}	$3\mu A$	$25\mu A$	$3\mu A$
AC parameters			
Transresistance – r_z	$900k\Omega$	$316k\Omega$	$900k\Omega$
Input resistance:			
+Input – r_y	$10M\Omega$	$3{,}3M\Omega$	$10M\Omega$
–Input – r_x	50Ω	51Ω	50Ω
+ Input capacitance – C_y	$1{,}5\,pF$	$1{,}5\,pF$	$1{,}5\,pF$
–3 dB Small Signal Bandwidth	$880MHz$ (1) $260MHz$ (2)	$370MHz$ (1) $427MHz$ (2)	$870MHz$ (1) $270MHz$ (2)
Dominant pole frequency– f_p	$700kHz$	$896kHz$	$700{,}01kHz$
Slope (open loop)	$-20dB/dec$	$-20dB/dec$	$-20dB/dec$
Output voltage swing – V_{max}	$\pm 3{,}1V$ (3)	$\pm 2{,}68V$ (3)	$\pm 3{,}09V$ (3)
Short circuit current – I_{max}	$100mA$	$85{,}64mA$	$100{,}01mA$
Output resistance – r_o	50Ω	51Ω	50Ω

Notes: (1) $A_U = +1$ ($R_F = 649\Omega$ [15]); (2) $A_U = +10$ ($R_F = 470\Omega$ and $R_N = 51\Omega$ [15]); (3) $R_L = 100\Omega$.

To validate the proposed CFOA model, simulation of Deliyannis-Frend biquad bandpass filter shown in Fig. 3 was carried out for both the aforementioned model and the macromodel given by Analog Devices in the PSpice library. The results are shown in Fig. 4a and Fig. 4b.



Bandpass filter with parameters:

$f_o = 1MHz$

$Q = 1$

$H_{oBP} = 1$

$n = 2$

Fig. 3. The Deliyannis-Frend biquad filter used in testing the different CFOA models.

The result of the simulations performed shows that the presented model corresponds to the AD8001AN/AD for frequencies from 0 up to $100MHz$. At higher frequencies the macromodel supplied from Analog Devices is superior to the behavioural one, with the expense of a much more complicated set of mathematical equations that describe the proposed model.

The purpose of developing that CFOA model was to introduce a behavioural model that is suitable for analyses of the mixed-signal applications with VHDL-AMS simulators yet quite accurate to a wide frequency range.

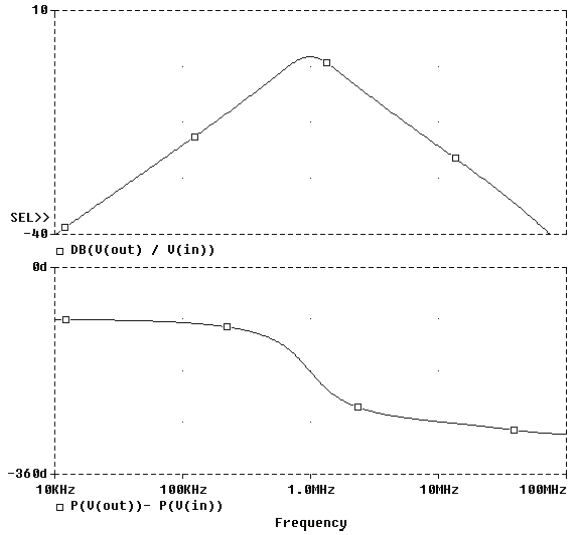


Fig. 4a. The simulated response of the bandpass filter shown in Fig. 3 for the AD8001AN/AD.

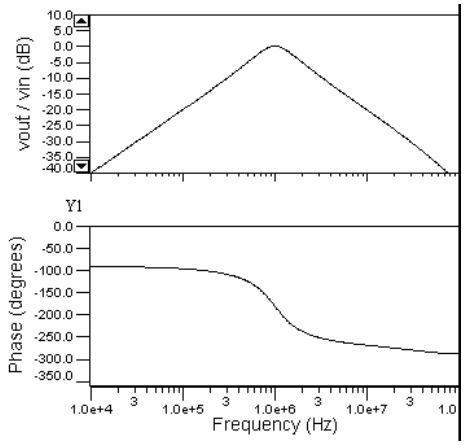


Fig. 4b. The simulated response of the bandpass filter shown in Fig. 3 for the proposed CFOA model.

V. Conclusion

In this paper behavioural VHDL-AMS model of CFOA has been presented. The proposed model accurately describes the behaviour of a typical CFOA over a wide frequency range, including electrical parameters such as inverting and non-inverting input impedance, open-loop transresistance, bandwidth, voltage and current limitations, output resistance, etc. To achieve simplicity of the mathematical equations describing the model, it neglects several second-order effects found in the CFOAs (such as the frequency dependence of the output impedance of the voltage followers, the CMRR and the temperature effects). One of the aims of the further work is to explore the possibility of modelling second-order effects of the typical CFOA, which are taken into consideration for designing high-frequency mixed-signal applications.

Acknowledgement

This paper is a part of a project which is sponsored by the research program of the Technical University of Sofia, Bulgaria.

References

[1] V. Tietze, Ch. Schenk, *Halbleiter-Schaltungstechnik. 12. Auflage*. Springer-Verlag, Berlin, New York, 2002 (in German).

[2] W. Jung, W. *Op amp applications*. Analog Devices, MA, 2002.

[3] M. Seifart, *Analoge Schaltungen. 6 Auflage*. Verlag Technik Berlin, 2003 (in German).

[4] D. Bowers, M. Alexander, J. Buxton, "A comprehensive simulation macromodel for current feedback operational amplifiers", IEE Proceedings G, vol. 137 (2), pp. 137-145, 1990.

[5] National Semiconductor, "Development of an extensive SPICE macromodel for current-feedback amplifiers", National Semiconductor application note 840, 1992.

[6] Intersil, "HA5013 SPICE Macromodel (CFA)", Intersil application note MM5013.1, 1996.

[7] M. Eltawil, A. Soliman, "New precise Spice macromodels for the current-feedback operational amplifiers", Microelectronics Journal, 30, pp. 841-849, 1999.

[8] Analog Devices, *Linear products data book*. Norwood, MA, 2000.

[9] I. Pandiev, "An enhanced simulation macromodel for current-feedback amplifiers", Proceedings of the TU-Sofia, vol. 56, pp. 9-15, 2006.

[10] I. Pandiev, "Behavioral macromodeling of current feedback amplifiers and Drive-R – Amplifiers", Proceedings of the TU-Sofia, vol. 58 (2), pp. 129-136, 2008.

[11] S. Robinson, "Conceptual modeling for simulation: Issues and research requirements", Proceedings of the 2006 Winter Simulation Conference, pp. 792-799, 2006.

[12] S. Robinson, *Successful Simulation. A Practical approach to Simulation Projects*. McGraw-Hill, 1998.

[13] E. Christen, K. Bakalar, "VHDL-AMS – A hardware description language for analog and mixed-signal applications, IEEE Transactions on circuits and systems – II, vol. 46 (10), pp. 1263-1272, 1999.

[14] *Definition of analog and mixed signal extensions to IEEE standard VHDL*, IEEE Standard 1076.1, 1999.

[15] *AD8001 current feedback amplifier – data sheet*, Analog Devices, 2003.

Voltage-mode Lowpass, Bandpass, and Bandstop Programmable Filter using Four-terminal CFOAs

Ivailo M. Pandiev[1]

***Abstract* – In this paper a voltage-mode programmable active-RC filter employing four Current-Feedback Operational Amplifiers (CFOAs), one dual CMOS digital potentiometer, seven virtually grounded resistors and two grounded capacitors is presented. The presented active-RC filter is the result of a systematic circuit synthesis and can comparatively easily be derived from the VFOA-based Fleischer-Laker (FL) Switched Capacitor (SC) biquad stage, providing lowpass, bandpass and bandstop responses. The proposed circuit offers the following advantages: (1) realization of lowpass, bandpass and bandstop (notch) responses from the same configuration, (2) orthogonal digital control of the natural frequency and the quality factor by grounded resistors, (3) the use of two grounded capacitors which is not the case in the FL circuit, (4) low active and passive sensitivities and (5) low output impedance. The workability of the new active-RC filter has been demonstrated by PSpice simulation results.**

***Keywords* – Analogue circuits, Active-RC filters, Four-terminal CFOAs, CCII+s.**

I. INTRODUCTION

The applications and advantages in the realization of various active-RC filters, using four-terminal current-feedback operational amplifiers (CFOAs), have received a considerable amount of attention [1, 2]. The four-terminal CFOAs are special type of monolithic operational amplifiers (op amps) and can be represented as a cascade structure of positive second-generation current conveyor (CCII+) and an additional voltage buffer. These op amps provide wide bandwidth, which is relatively independent of the closed-loop gain, and a very high slew rate. Moreover, the CFOAs have a low-impedance output, which makes the circuit cascadable without the use of additional buffers. A number of active-RC filter circuits, built with CFOAs, have been presented in the literature [3−7]. Each of the CFOA-based filters uses two or three active elements and a small group of resistors and capacitors. The natural frequency (ω_o) and the quality factor (Q) can be orthogonally adjustable by grounded resistors or virtually grounded resistors. When a grounded resistor is replaced by a JFET transistor, a voltage-controlled filter can be obtained. However, the electronic control of the parameters is not demonstrated in the reported circuits.

On the other hand, the conveyed literature survey reveals that recently large numbers of tunable universal filters using CCIIs are available [8−12]. The benefits of these circuits are that the bandwidth and natural frequency can be tuned electronically via the input bias currents of the used op amps.

To the author's knowledge, the use of four-terminal CFOAs in designing programmable filters has not yet been reported in the literature. It is, therefore, the purpose of this paper to present a voltage-mode lowpass, bandpass and bandstop digitally programmable filter, employing CFOAs, one dual digital potentiometer and two grounded capacitors.

II. FOUR-TERMINAL CFOAS

The four-terminal CFOA is equivalent to a CCII+ and an output voltage buffer. These op amps have a high impedance non-inverting input y, a low-impedance inverting input x, a current output z and a voltage output o. The port z is between the first stage (CCII+) and the second stage (voltage buffer), where the resistance is very high (magnitude of several mega ohms). This allows the usage of four-terminal CFOAs in selective amplifier circuits, sinusoidal oscillators and multivibrators. The port o is the output of the voltage buffer, where the resistance is very low (magnitude of several ohms). The schematic representation of a four-terminal CFOA is shown in Fig. 1a [1, 2]. The monolithic op amps AD844 (from Analog Dev.), OPA660, OPA860 (from Texas I.), MAX436 (from Maxim) and LM13700 (from National) can be used as four-terminal CFOAs. For some of the available op amps, such as OPA860, LM13700 and MAX436, the CCII+ is not connected to the output voltage buffer.

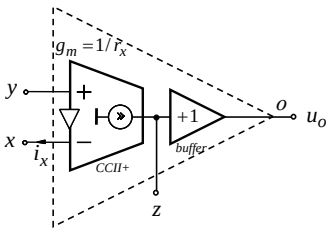


Fig. 1a. Symbol of a four-terminal CFOA.

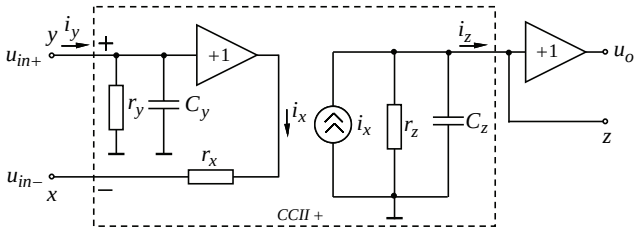


Fig. 1b. Linear macro-model of a four-terminal CFOA.

[1]Ivailo M. Pandiev is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ipandiev@tu-sofia.bg

The linear macro-model of the four-terminal CFOA, presented in Fig. 1b, reflects the small-signal behavior of the real device. The model includes the following elements: input and output buffers (voltage followers); i_x – controlled current source; r_y and C_y – input resistance and capacitance of the non-inverting input; r_x – resistance of the inverting input; r_z and C_z – output resistance and capacitance.

For this CFOA the general relation between input and output voltages and currents can be given by the following hybrid matrix

$$\begin{bmatrix} i_y \\ u_x \\ i_z \\ u_o \end{bmatrix} = \begin{bmatrix} 1/Z_y & 0 & 0 \\ 1 & r_e & 0 \\ 0 & 1 & 1/Z_z \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_y \\ i_x \\ u_z \end{bmatrix}, \qquad (1)$$

where $Z_y = r_y \,\|\, (1/sC_y)$ and $Z_z = r_z \,\|\, (1/sC_z)$.

The matrix representation given with (1) is valid only for ideal input and output voltage buffers.

The four-terminal CFOAs have several important applications. For example, these types of CFOAs are used in realization of analogue computing current-mode circuits, high-speed amplifiers, active filters, oscillators etc. The objective in this paper is to demonstrate the practicality of the four-terminal CFOAs in transforming well known Voltage-Feedback Op Amp (VFOA) based FL SC biquad stage, providing lowpass, bandpass and notch responses, to voltage-mode digitally programmable active-RC filter using CFOAs [13-14].

III. CIRCUIT DESCRIPTION

Fig. 2 represents Fleischer-Laker SC biquad stage providing notch, bandpass and lowpass responses. This circuit uses four voltage-feedback amplifiers and the natural frequency can be tuned via the clock frequency f_S. The aim here is to have an exact equivalent circuit with three outputs, notch, bandpass and lowpass, using four-terminal CFOAs. The approach is based on realizing the four basic building blocks, namely the two summers and two non-inverting integrators [1, 15], using four-terminal CFOAs plus a dual digital potentiometer, and then connecting these blocks correctly.

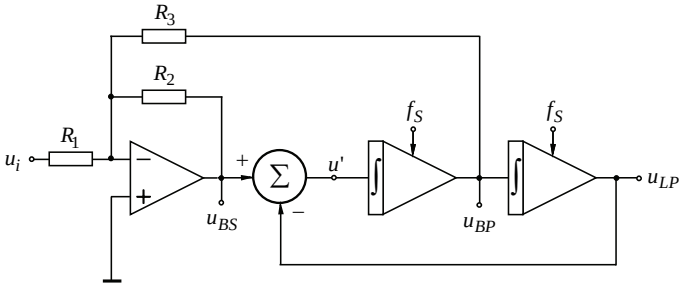


Fig. 2. VFOA based Fleischer-Laker SC biquad stage providing lowpass, bandpass and notch responses.

The first inverting summer of the circuit, shown in Fig. 2, is realized using equivalent closed-loop inverting summer circuit with a single four-terminal CFOA. For the second summer, an open-loop summing circuit with a second CFOA is used. In this circuit the output voltage from the first summer is applied directly to the y terminal of the CFOA and the output voltage from the second integrator is applied to the x terminal of the CFOA, through a resistor. The voltage summing node of the second summer is the z terminal of the op amp. Additionally, a resistor is connected between the z terminal and ground, equal to the one, connected to the x terminal of the CFOA [15]. The realization of the non-inverting active-RC integrator using a positive CCII, a grounded resistor and a grounded capacitor is well known [1]. Although summer output and the two integrators can be easily cascaded, the integrator outputs have to be buffered before being connected to the proper summer inputs. Here the two built in voltage buffers of the CFOAs are used.

The proposed voltage-mode lowpass (LP), bandpass (BP) and bandstop (BS) programmable filter with a single input and three outputs (LP, BP and BS), employing four CFOAs, is shown in Fig. 3. This circuit is a cascade structure of two non-inverting integrators, using monolithic CFOAs $U_3 - U_4$ (and the associated variable resistors $R_{AB1} = R_{AB2} = R_{AB}$ and capacitors $C_1 = C_2 = C$) and two summers, realized with CFOAs $U_1 - U_2$ and resistors $R_1 - R_5$. The capacitors and variable resistors of the two integrators are all connected to ground. This is particularly advantageous in facilitating the electronic control of the time constants. The pole frequency (or the time constant) of this filter are electronically controllable via the digital codes D (i.e. $D = D_1 = D_2$) of the dual grounded digital potentiometer – U_5. Moreover, the terminal W (wiper) of the dual potentiometer U_5 is connected to ground, which decreases the influence of the noise voltages. The programmable resistances $R_{WB1} = R_{WB2} = R_{WB}$ between W and B terminals are given by

$$R_{WB}(D) = \frac{D}{2^n} R_{AB} = qR_{AB}, \qquad (2)$$

where $q = D/2^n$, n is the resolution or "step size" of the potentiometer and R_{AB} is the nominal resistance between terminal A and terminal B.

Using the hybrid matrix given in equation (1), which characterizes the four-terminal CFOA, and using the condition $R_4 = R_5$ by routine analysis, the LP, BP and BS functions realized by this circuit are given by

$$A_{LP}(s) = \frac{u_{LP}}{u_i} = \frac{-\frac{R_2}{R_1}\frac{1}{\tau^2}}{D(s)}, \qquad (3)$$

$$A_{BP}(s) = \frac{u_{BP}}{u_i} = \frac{-\frac{R_2}{R_1}\frac{1}{\tau}s}{D(s)} \qquad (4)$$

and

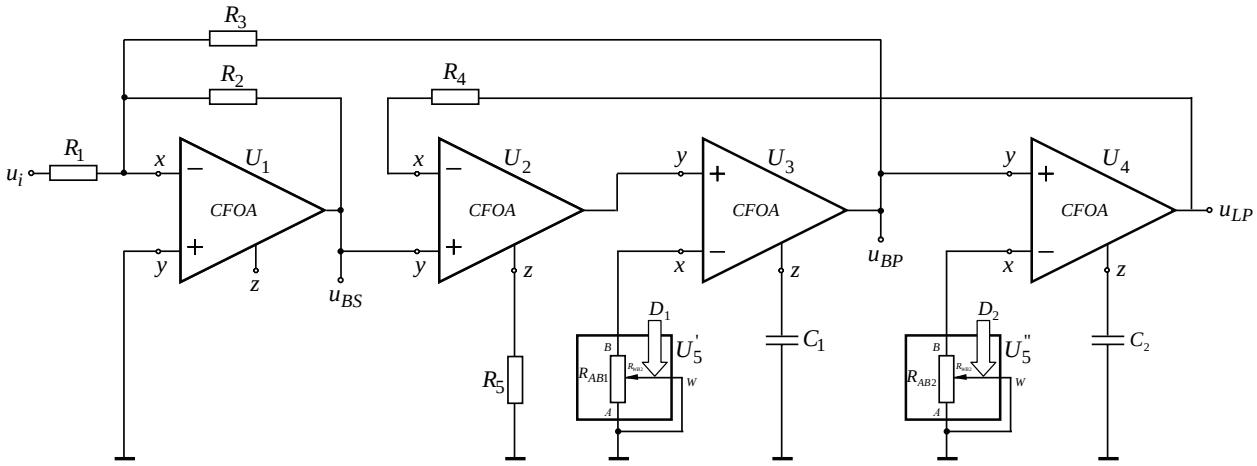


Fig. 3. Proposed digitally programmable lowpass, bandpass and bandstop filter configuration using monolithic four-terminal CFOAs.

$$A_{BS}(s) = \frac{u_{BS}}{u_i} = \frac{-\frac{R_2}{R_1}\left(s^2 + \frac{1}{\tau^2}\right)}{D(s)}, \qquad (5)$$

where

$$D(s) = s^2 + \frac{R_2}{R_3}\frac{1}{\tau}s + \frac{1}{\tau^2}, \qquad (6)$$

$$\omega_o = \frac{1}{\tau} = \frac{1}{\underbrace{(R_{WB} + r_x)C}_{\tau}} \overset{R_{WB} >> r_x}{\approx} \frac{2^n}{D}\frac{1}{R_{AB}C} \qquad (7)$$

is the pole frequency of the filters,

$$Q = \frac{R_3}{R_2} \qquad (8)$$

is the quality factor (Q-factor) and

$$H_{oLP} = -\frac{R_2}{R_1}, \quad H_{oBP} = -\frac{R_3}{R_1}, \quad H_{oBS} = -\frac{R_2}{R_1} \qquad (9)$$

are the pass-band gains.

The above equations are identical of those of the circuit of Fig. 2. From (7) and (8), note that the pole frequency can be linearly adjusted directly by the digital code D of the dual digital potentiometer without disturbing the quality factor Q, while the quality factor can be adjusted independently from the pole frequency and the pass-band gains of the LP and BS functions by varying the resistor R_3. From (9), it can be remarked that the desired pass-band gain H_O can be independently adjusted by the resistor R_1. Other advantages of the proposed circuit are that they have no sample-data effect (increased clock feedthrough noise, reduced Power Supply Rejection Ratio – PSRR, aliasing errors) and they are better suitable for high frequency applications in comparison with the SC biquad stages [13, 14].

The above analytical formulas, as a result of the theoretical analysis, are the base of the design procedure of the proposed programmable filter with monolithic CFOAs. The basic design equations are given by

$$C_1 = C_2 = C \qquad R_4 = R_5 \qquad R_{WB1} = R_{WB2} = \frac{1}{\omega_o C} \qquad (10)$$

$$R_3 = QR_2; \qquad (11)$$

$$\textit{in case of LP: } R_2 = |H_{oLP}|R_1,$$

$$\textit{in case of BP: } R_2 = |H_{oBP}|R_1/Q \textit{ or}$$

$$\textit{in case of BS: } R_2 = |H_{oBS}|R_1. \qquad (12)$$

These design equations take into account that the two digital potentiometers $U_5^{'}$ and $U_5^{''}$ are implemented with one dual potentiometer with $R_{AB1} = R_{AB2}$ and moreover the positions of the wipers (W) are equally for a given digital code D, i.e. $R_{WB1} = R_{WB2}$. Other design equations are possible due to the available degrees of freedom, but are not discussed here.

IV. Simulation results

In order to confirm the validity of the proposed programmable active-RC filter (Fig. 3), the circuit has been simulated using PSpice simulation program. The four-terminal CFOA, given in Fig. 1a, was designed using an analog IC type OPA860 (from Texas I.), biased with $\pm 5V$. The digitally controlled variable resistors were constructed using the dual digital potentiometer AD8402 (from Analog Dev.), biased with $+5V$ single power supply. The IC AD8402 is with nominal resistances $R_{AB1} = R_{AB2} = 1k\Omega$ and has 256 positions of the wipers.

The active-RC filter example was designed for an initial pole frequency $f_o = \omega_o / 2\pi = 10kHz$ and a quality factor $Q = 0{,}707$ based on the Butterworth approximation method [1]. Capacitors $C_1 = C_2 = 16nF$ and resistors $R_4 = R_5 = 1k\Omega$

were chosen. Other parameters of the circuit are: $R_1 = R_2 = 1{,}4k\Omega$, $R_3 = 1k\Omega$ and $R_{WB1} = R_{WB2} = 1k\Omega$.

For the computer simulation the OPA860 and the AD8402 macromodels, given in the PSpice libraries, were used. The simulation results for the LP, BP and BS filter characteristics are shown in Fig. 4.

To demonstrate the digital tuning of ω_o, the digital codes D (i.e. $D = D_1 = D_2$) were simultaneously adjusted for the values 255, 128, 64 and 32 respectively, while keeping the rest of passive parameters for a constant $Q = 0{,}707$ and $|H_{oBP}| = 0{,}7$. The resulting responses of the BP filter for different digital codes D when $C_1 = C_2 = 16nF$ are given in Fig. 5.

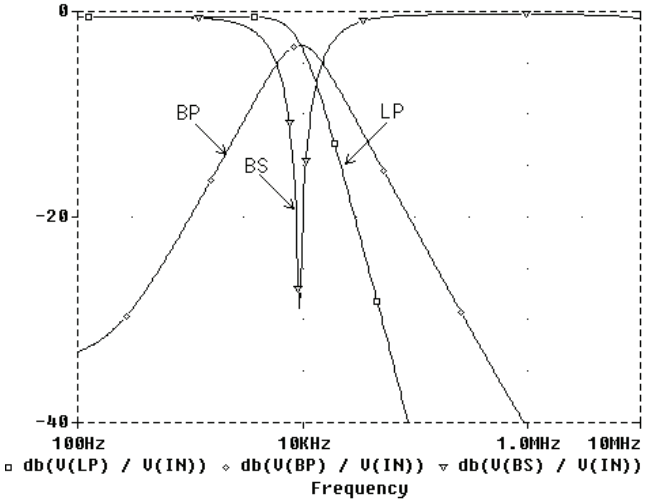


Fig. 4. Simulated frequency characteristics of LP, BP and BS of the proposed voltage-mode filter in Fig. 2.

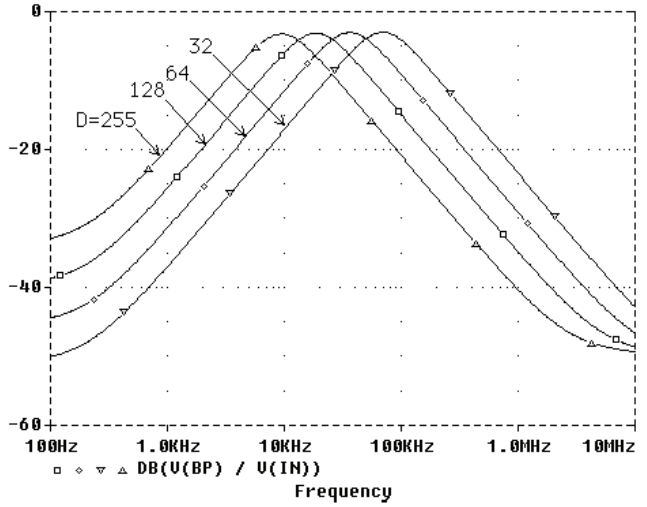


Fig. 5. Simulated frequency responses of the BP filter when the digital code D is varied.

V. Conclusion

It is seen that with the four-terminal CFOA one equivalent circuit to the Fleischer-Laker SC biquad stage providing notch, bandpass and lowpass responses is created, based on the equivalent building blocks. The proposed digitally programmable filter requires four CFOAs and one dual digital potentiometer. The circuit has two grounded capacitors, which is not the case in the Fleischer-Laker SC biquad stage. Moreover, the two variable resistors of the integrators are also connected to the ground. The created filter has been tested with monolithic op amp OPA860, used as CFOA, obtaining good behaviour for low and high frequencies.

In fact, the proposed circuit adds a new configuration to the existing group of electronically controllable active-RC filters with CFOAs and should be useful in offering some new features to the analog designers.

Acknowledgement

This paper is a part of a project which is sponsored by the research program of the Technical University of Sofia, Bulgaria.

References

[1] V. Tietze, Ch. Schenk, *Halbleiter-Schaltungstechnik. 12. Auflage*. Springer-Verlag, Berlin, New York, 2002 (in German).
[2] W. Jung, W. *Op amp applications*. Analog Dev., USA, 2002.
[3] C.-M. Chang, C.-S. Hwang, S.-H. Tu, "Voltage-mode notch, lowpass and bandpass filter using current-feedback amplifiers". Electronics Letters, vol. 30, No. 24, pp. 2022-2023, 1994.
[4] S.-I. Liu, "Universal filter using two current-feedback amplifiers". Electronics Letters, vol. 31, No. 8, pp. 629-630, 1995.
[5] S.-I. Liu, "High input impedance filters with low component spread using current-feedback amplifiers". Electronics Letters, vol. 31, No. 13, pp. 1042-1043, 1995.
[6] N. Shah, S. Iqbal, M. Rather, "Versatile voltage-mode CFA-based universal filter". Int. J. Electron. Commun. (AEÜ), vol. 59, pp. 192-194, 2005.
[7] S. Gupta, D. Bhaskar, R. Senani, A. Singh, "Inverse active filters employing CFOAs". Electrical Engineering, vol. 91, pp. 23-26, 2009.
[8] R. Senani, K. Singh, A. Singh, D. Bhaskar, "Novel electronically controllable current-mode universal biquad filter". IEICE Electronics Express, vol. 1, pp. 410-415, 2004.
[9] W. Tangsrirat, W. Surakampontorn, "Electronically tunable current-mode universal filter employing only plus-type current-controlled conveyors and grounded capacitors". Circuits, Systems and Signal Processing, vol. 25, pp. 701-713, 2006.
[10] M. Siripruchyanum, W. Jaikla, "Electronically controllable current-mode universal building filter using single DO-CCCDTA". Circuits, Systems and Signal Processing, vol. 27, pp. 113-122, 2008.
[11] J. Jiang, Y. He, "Tunable frequency versatile filters implementation using minimum number of passive elements". Analog Integrated Circuits and Signal Process, vol. 59, pp. 53-64, 2009.
[12] H. Chen, P. Chu, "Versatile universal electronically tunable current-mode filter using CCCIIs". IEICE Electronics Express, vol. 6, pp. 122-128, 2009.
[13] P. Fleischer, K. Laker, "A family of active switched capacitor biquad building blocks". Bell Labs Technical Journal, vol. 58, pp. 2235-2268, 1979.
[14] K. Laker, W. Sansen, *Design of analog integrated circuits and systems*. McGraw-Hill, New York, 1994.
[15] A. Soliman, "Kerwin-Huelsman-Newcomb circuit using current conveyors". Electronics Letters, vol. 30, pp. 2019-2020, 1994.

Recombination Process in Bilayer Polymer Electroluminescent Structures

Mariya P. Aleksandrova[1], Milka M. Rassovska[1], Georgy H. Dobrikov[1]

***Abstract* – In this article an analysis of the recombination process in bilayer polymer electroluminescent structure was presented, using simulations, made by program SimOLED Electrical. The aim of this study is to make reconstruction of the processes, arising at the junction between the hole transporting layer (HTL) and the electroluminescent layer (EL), where we could not receive information by direct measurements. The dependence of the recombination profile on the contact injection barriers at the polymer interface was estimated. The impact of the barriers height on the location of recombination peak and on the recombination rate was examined too. The mechanisms by which the recombination zone shifts from HTL side to the EL side were described in detail, according to the increases in conduction bands energy offset.**

***Keywords* – Polymer electroluminescent structure, Recombination zone, Interface barriers.**

I. INTRODUCTION

The polymer light-emitting devices (PLED) are intensively investigated as flat display technology of the future due to their several advantages including their own lightening, wide viewing angle, thin size and low cost [1]. The simplest PLED structure consists of a single polymer layer sandwiched between an anode and a cathode, but this structure suffers from poor efficiency [2]. In the most organic semiconductors inclusive conjugated polymers the hole mobility is larger than the electron mobility. This is the reason for charge carrier recombination near the cathode interface, where is more probable charge carriers trapping in the quenching centers formed during the cathode deposition. By insertion of additional buffer layer, decreasing of the anode injection barrier and recombination zone shifting away from the electrode interface could be achieved. As a result the quantum efficiency is increased.

Several studies are carried out for investigating the operation principle of bilayer polymer electroluminescent structures [3-5]. The authors explore the role of the electrode interfacial barriers height for the current balance in the electroluminescent thin film. Goliney first report data from modeling of performance characteristics of organic light-emitting device in the case of tunneling and thermionic injection through the electrodes into the emitting layer [6]. Paasch et al. conduct simulations to confirm the hypothesis for exponentially trap energy distribution and existence of localized charges in the vicinity of the anode or the cathode of organic light emitting devices [7]. Although it is known that the recombination process is dependent on the electrode contact barriers, there is no detail information about the charge carrier transport through the polymer heterojunction, according to the polymer interfacial barriers for the electrons and holes.

In the present article the radiative recombination rate and position in bilayer PLED structure ITO/PVK/PPV-D/Al is analyzed according to the polymer interface barriers. Indium-tin oxide (ITO) is transparent and conductive film, used as anode. Polyvinylcarbazole (PVK) is hole transporting buffer layer (HTL), polyphenylenevinylene derivative (PPV-D) - electroluminescent layer (EL) and aluminum film is used as cathode. The investigation aims to reconstruct the processes at the polymer layers interface, where it is not possible to receive information by direct experimental measurements.

II. EXPERIMENTAL

The analysis in this article is based on simulation results achieved with the specialized OLED/PLED simulation software SimOLED ELECTRIC [8], which considers processes like charge carrier injection and transport, radiative recombination, non-radiative decay etc. This program uses experimental measured data for input parameters, so the simulation results complement the data from the real measurement. The necessary parameters taken from the real prepared device are layer sequence and thickness d, highest occupied and lowest unoccupied molecular orbital (HOMO and LUMO) energies, electrodes work function, main charge carrier mobility μ in the used polymer materials and relative permittivity. For the concretely investigated structure these material property's values are as follow: d_{ITO} = 180 nm, d_{PVK} = d_{PPV-D} = 100 nm, d_{Al} = 200 nm; the energy levels (HOMO, LUMO, work functions) could be seen in the energy diagram on Fig. 1; $\mu_{h\ PVK} = 5{,}4.10^{-3}$ cm^2/V.s, $\mu_{h\ PPV-D} = 3.10^{-5}$ cm^2/V.s; $\varepsilon_{r\ PVK}$ = 2,9, $\varepsilon_{r\ PPV-D}$ = 3, turn on voltage 10 V (from the forward current-voltage characteristic).

The current flow at the polymer interface is modeled in SimOLED ELECTRIC by thermoionic emission, according to the approved from Crone et al. model [9]. The assumption is for exponentially distribution of trap energies, which is the most widespread case in the organic semiconductors and in particular in the polymers.

[1]Mariya P. Aleksandrova is with the Faculty of Electronic Engineering and Technology, Technical University of Sofia, "Kl. Ohridski" str, 8, Bulgaria, E-mail: meri_7@abv.bg

[1]Milka M. Rassovska is with the Faculty of Electronic Engineering and Technology, Technical University of Sofia, "Kl. Ohridski" str, 8, Bulgaria, E-mail: rasovska_m@mail.bg

[1]Georgy H. Dobrikov is with the Faculty of Electronic Engineering and Technology, Technical University of Sofia, "Kl. Ohridski" str, 8, Bulgaria, E-mail: georgy_ddobrikov@abv.bg

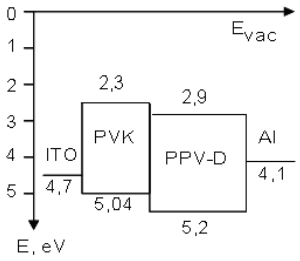


Fig. 1. Energy diagram of ITO/PVK/PPV-D/Al electroluminescent structure.

The used approach for explaining of charge carrier motion in PLED is not appropriate for applying to hybrid organic/inorganic interface. The suitable for PLED inorganic buffer materials have mainly insulating properties and for that reason they must be deposited in form of ultrathin films (under ~6-8 nm) in the structure for prevention of charge carrier blocking. Injection through such interface is by tunnelling mechanism over triangular barrier, the material transport properties are neglected and they don't influence on the charge transportation. For simplicity, constant charge mobility in the bulk of the material is accepted. In our case this supposition is valid, because the metal/polymer contact is not Ohmic, so there is no disturbing field from space charge formation.

III. Results and Discussion

A model of structure, consisting of polymer material – 1 (PVK), as hole transporting layer and polymer material – 2 (PPV-D), as electroluminescent layer, sandwiched between 2 electrodes is shown on Fig. 2a. Fig. 2b represents energy diagram with the corresponding polymer energy levels offsets relative to each other. Although at the cathode interface Schottky barrier is formed (≈ 2 eV) during the simulations value of 0,2 eV was set to avoid restraints of the charge carrier injection. Thus the influence of the injection barriers height at the polymer interface on the device performance could be successfully demonstrated.

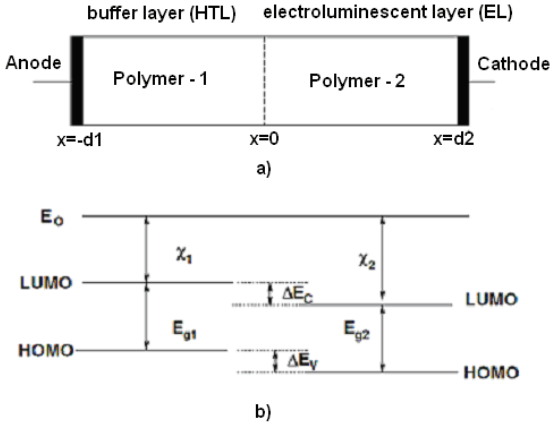


Fig. 2. a) Schematic diagram of a bilayer PLED and b) its associated energy band diagram.

Fig. 3 (up) illustrates the recombination profile in the bilayer PLED structure with ΔE_v=0,16 eV and ΔE_c=0,6 eV. It could be seen, that the recombination occurs predominantly at the polymer interface (zero position in relation to the junction). Some of the charge carriers, however, recombine in the HTL or EL layer. The situation of constant HOMO energy offset was considered first (ΔE_v=const) and LUMO energy offset varies from 0 to 0,8 eV. Fig. 3 (down) shows the recombination rate at the both side of polymer interface, depending on the electron injection barrier ΔE_c. Initially at small values of ΔE_c the charge carriers recombine in the HTL near the polymer interface. With ΔE_c increasing above certain value (~0,6 eV), the recombination zone is entirely shifted in EL layer. It seems that the holes already reach to the cathode interface, passing through the EL layer until the electrons overcome the ΔE_c barrier. The recombination rate in EL layer actually increase with electron injection barrier at the polymer interface as is shown on Fig. 3 (down). This effect probably arises, because of the accumulated electrons in the EL layer.

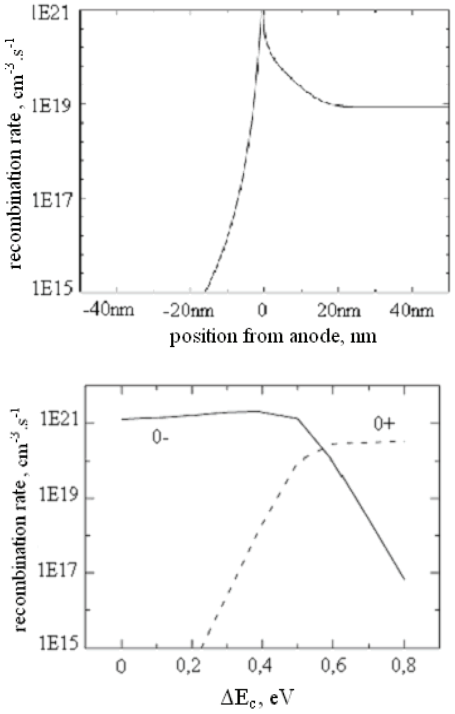


Fig. 3. A recombination profile in the bilayer structure (up); recombination rate immediately to the left (solid line) and to the right (broken line) of the polymer interface as a function of LUMO energy offset (down).

Situation of constant HOMO energy offset was considered first (ΔE_v=const) and LUMO energy offset varies from 0 to 0,8 eV. Fig. 3 down shows the recombination rate at the both side of polymer interface, depending on the electron injection barrier ΔE_c. Initially at small values of ΔE_c the charge carriers recombine in the HTL near the polymer interface. With ΔE_c increasing above certain value (~0,6 eV), the recombination zone is entirely shifted in EL layer. It seems that the holes already reach to the cathode interface, passing through the EL

layer until the electrons overcome the ΔE_c barrier. The recombination rate in EL layer actually increase with electron injection barrier at the polymer interface as is shown on Fig. 3. This effect probably arises, because of the accumulated electrons in the EL layer.

It is difficult to give a detailed explanation why the current through the structure remains almost constant by increasing of the LUMO offset (Fig. 4). Actually, all anode injected holes, which have been recombined in HTL before, recombining in the EL layer now. Consequently, the hole injection in the EL layer is expect to increase together with the electron injection barrier increasing, regardless of the unchanged applied voltage U.

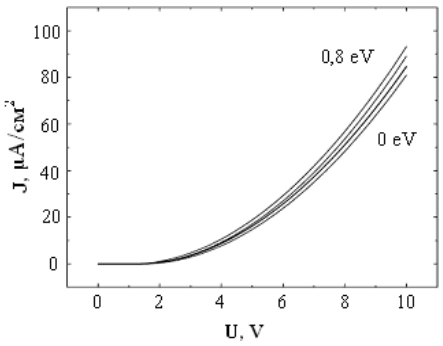


Fig. 4. Current density as a function of voltage obtained from numerical simulations for different values of LUMO energy offset (0; 0,2; 0,6; 0,8 eV) at the polymer interface.

Assuming that due to the large electric field at the interface, the drift component is predominant, the electron current in HTL at the polymer interface can be expressed as:

$$J \cong q\mu_{n1}E(0)n(0^-), \tag{1}$$

where E(0) is the electrical field at the interface, μ_{n1} – electron mobility in HTL (polymer – 1), $n(0^-)$ – electron density at the interface in HTL side. Using the relation

$$n(0^-) = n(0^+)\exp(-\frac{\Delta E_c}{kT}) \tag{2}$$

we can re-write Eq. 1 as

$$J \approx q\mu_{n1}E(0)n(0^+)\exp(-\frac{\Delta E_c}{kT}), \tag{3}$$

where k is the Boltzman's constant and T is temperature.

To estimate the increase in electric field or electron density separately, relation between both of them next to the interface is required. The results from simulation indicate that carrier density decrease exponentially away from the interface [10].

$$n(x>0) = n(0^+)\exp(-\delta_2 x), \tag{4}$$

where δ_2 is the distance from interface to bulk region in the electroluminescent film.

The electron density in EL close to the interface is

$$Q_N = qn(0^+)\delta_2^{-1} \tag{5}$$

The accumulation of electrons close to the interface means that electron drift and diffusion currents must be almost equal and in opposition to each other. The assumption for quasi-equilibrium condition together with Eq. 4 can be used to obtain an expression of the electric field at the interface.

$$E(0) = \frac{kT}{q}\delta_2 \tag{6}$$

Using this expression, Eq. 5 can be written as

$$Q_N = n(0^+)\frac{kT}{E(0)} \tag{7}$$

Assuming that the region with accumulated charges acts as "shield" for further charge carrier transport, another expression can be applied, relating electric field and interface charge density [11]:

$$E(0) = \frac{Q_N}{\varepsilon} + E_{b2}, \tag{8}$$

where E_{b2} electric field in the EL bulk near the interface and ε is relative permittivity of the same layer. Substitution of Eq. 8 in Eq. 7 leads to quadratic equation in relation to E(0), whose solution is:

$$E(0) = 0{,}5E_{b2}\left\{1+\sqrt{1+\frac{n(0^+)}{n_2}}\right\}, \quad n_2 = (\frac{\varepsilon E_{b2}^2}{4kT}), \tag{9}$$

or Eq. 9 can be rewritten in a form, which is more handy for verification

$$\underbrace{\left\{\frac{2E(0)-E_{b2}}{E_{b2}}\right\}^2}_{Y_1} = 1+\frac{n(0^+)}{n_2} \tag{10}$$

In order to verify Eq. 9, numerical simulations were carried out to obtain values of E(0) and $n(0^+)$ for electron barrier height in the range of 0–0,4 eV. Fig. 5 shows the left side of Eq. 10 as a function of the electron density, which is approximated with straight line.

Finally it is considered a bilayer PLED structure with ΔE_v=0,2 eV and ΔE_c=0 eV at the polymer interface. When forward bias is applied, holes are accumulate in the buffer layer side, because of the hole barrier. In the light-emitting layer the hole current is negligible. In the same time injected from the cathode electrons are transported to the buffer layer without barrier. The hole density in HTL at the interface is large and only a few electrons could penetrate into the buffer, so the recombination peak will be at the interface from the HTL side. Electron mobility in PPV-D is smaller than the hole mobility in PVK, so the applied voltage is distributed as voltage drop mainly over the EL layer. The current flow then

is determinate from the electrons which can reach the interface. Such behavior give us a reason for modeling of this bilayer structure as a single layer structure, where the cathode serves as first electrode and the accumulated holes serve as a second electrode. At these conditions in the organic semiconductor devices the current is described as:

$$J = \frac{9}{8}\varepsilon_o \varepsilon_r \mu \frac{U^2}{d_2^3}, \qquad (11)$$

where ε_r is relative material permittivity; ε_o is dielectric constant; μ – mobility of the main charge carriers in layer, having thickness d_2. Fig. 6 shows that the simulation results are in good agreement with the behavior according to Eq. 11.

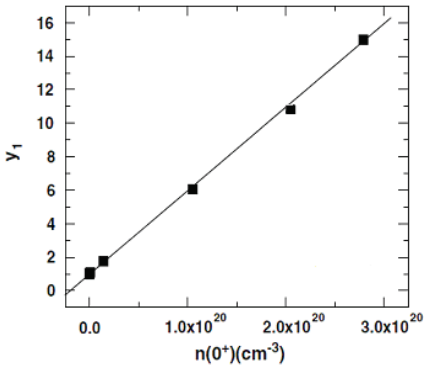


Fig. 5. Y_1 (left side of Eq. 10) plotted as a function of electron density in EL at the polymer interface.

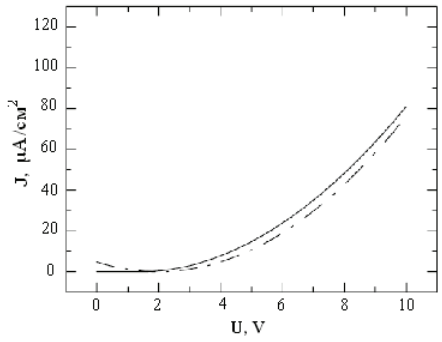


Fig. 6. Current density as a function of voltage obtained from numerical simulations (solid line) and the model (dotted line) described by Eq. 11 for ITO/PVK/PPV-D/Al bilayer structure.

IV. Conclusion

In this paper model of experimental bilayer polymer based electroluminescent structure was proposed. The real structure has configuration of ITO anode, PVK layer as hole transporting, PPV-D – as electroluminescent layer, and Al cathode. The model explains charge carrier recombination process in the electroluminescent layer. It was investigated the impact of interfacial barrier heights on the location of recombination peak and the recombination rate in such bilayer device. The mechanism by which recombination shifts from HTL side of the polymer–polymer interface to the EL side with increase in LUMO energy offset is described in detail. It was established that the total current flow in the structure is not sensitive to the electron barrier height at the polymer interface.

Acknowledgement

This work was financially supported by grant D002-358/2008 of Ministry of Education and Science, Bulgaria.

References

[1] C.W. Tang, S.A. Van Slyke, “Organic electroluminescent diodes”, Appl. Phys. Lett., vol. 51, no. 12, pp. 913-916, 1987.

[2] B.K. Crone, I.H. Campbell, P.S. Davids, D.L. Smith, C.J.Neef, J.P. Ferraris, “Device physics of single layer organic light-emitting diodes”, J. Appl. Phys., vol.86, no. 10, pp. 5767-5775, 1999.

[3] F.J.J. Janssen, J.M. Sturm, A.W. Denier van der Gon, at al., “Interface instabilities in polymer light emitting diodes due to annealing”, Organic Electronics, vol. 4, no. 4, pp. 209–218, 2003.

[4] H. Mu, W. Li, R. Jones, A. Steckl, D. Klotzkin, “A comparative study of electrode effects on the electrical and luminescent characteristics of Alq3/TPD OLED: Improvements due to conductive polymer (PEDOT) anode”, Journal of Luminescence, vol. 126, no. 1, pp. 225–229, 2007.

[5] T. Matsushima, K. Goushi, C. Adachi, “Charge-carrier injection characteristics at organic/organic heterojunction interfaces in organic light-emitting diodes”, Chemical Physics Letters vol. 435, no. 4-6, pp. 327–330, 2007.

[6] I.Goliney, “Modeling of performance characteristics of OLED”, Synthetic Metals, vol. 111–112, no. 1, pp. 359–361, 2000.

[7] G. Paasch, A. Nesterov, S. Scheinert, “Simulation study of the influence of polymer modified anodes on organic LED performance”, Synthetic Metals, vol. 130, no. 2, pp. 165–175, 2003.

[8] http://www.sim4tec.com/?Products:SimOLED_OLED_simulation_software:SimOLED_ELECTRIC.

[9] B.K. Crone, P.S. Davids, I.H. Campbell, D.L. Smith, “Device model investigation of bilayer organic light emitting diodes”, J. Appl. Phys., vol. 87, no 4, pp. 1974-1983, 2000.

[10] H. Bassler, Y. H. Tak, D. V. Khramtchenkov, V. R. Nikitenko, “Models of organic light-emitting diodes”, Synthetic Metals, vol. 91, no. 1-3, pp. 173-179, 1997.

[11] B. Mazhari, “Impact of Interfacial Barriers on Recombination Profile in Bilayer Organic Light-Emitting Diode”, Organic Electronics, vol. 6, no. 5-6, pp. 229-236, 2005.

Modeling of a Discharge Pulse in a Circuit With Two Discharge Gaps

Stefan T. Barudov[1] and Milena D. Dicheva[2]

***Abstract* –Many technological processes refer to formation of high voltage pulses in liquid medium. The high electrical conductivity of this medium imposes usage of switches in the high voltage circuit – most often controlled dischargers. The paper is dedicated to modelling of the electrical processes in a high voltage circuit with two discharge gaps and defining of the formed discharge pulse parameters. The analytically obtained results are compared to experimental ones from a prototype.**

***Keywords* – high voltage circuit, discharge gap, pulsed discharge in liquid medium.**

I. INTRODUCTION

The application of different technologies for treatment of liquid medium suggests forming of discharge pulses with certain parameters – energy, duration, amplitude of the current, etc. Often, when the electrical conductivity of the liquid medium is high it is necessary to be used high voltage switching element [2,3,4,5].

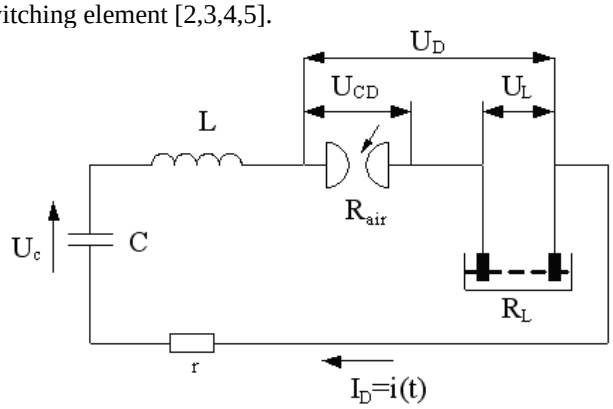


Fig.1 Discharge circuit with two discharge gaps

A similar discharge circuit is shown on Fig.1, where:

C – work capacitor battery (accumulating capacitive element), charged to voltage U_C;

L – inductivity of the capacitor battery and the connecting conductors;

CD – controlled discharger with resistance R_{IN};

R_L – resistance of the load (liquid medium);

r - resistance of the terminals of the capacitor battery and the connecting conductors;

R_{air} – resistance of the air gap in the trigatron.

The circuit on Fig.1 is non-linear because both of the discharge gaps are nonlinear:

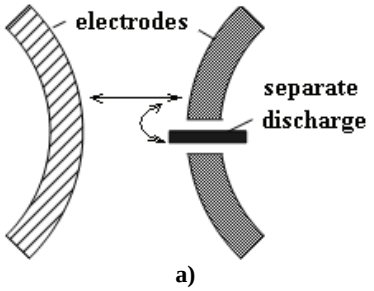
- for controlled discharger are valid:
  $U_{CD}=U_{CD}(t)$
  $I_{CD}=I_{CD}(t)$
  $U_{CD}=U_{CD}(I_{CD})$

- for the load (liquid medium) are valid:
  $U_L=U_L(t)$
  $I_L=I_L(t)$
  $U_L=U_L(I_L)$

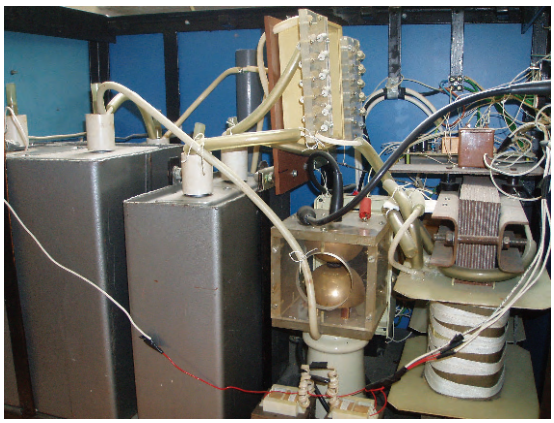
II. ANALYSES

The purpose of the current work is modeling of a discharge circuit with two discharge gaps and analytical and experimental study of the formed discharge pulse in the liquid medium.

If the controlled discharger is a trigatron – Fig.2, i.e. discharge in air is realized, the processes description according to fig.3 is as follows:



a)



b)

Fig. 2. a) Trigatron and b) experimental prototype

[1]Stefan T. Barudov is with the Faculty of Electrical Faculty, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: sbarudov@abv.bg

[2]Milena D. Dicheva is with the Faculty of Electrical Faculty, Technical University of Varna, 1 Studentska Str., Varna 9010, Bulgaria, E-mail: m.dicheva@tu-varna.bg

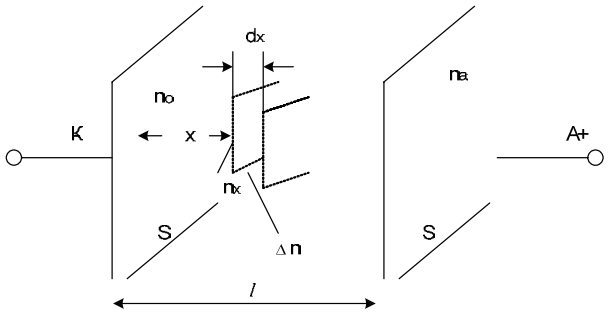


Fig. 3

If we suggest that near the cathode at the expense of external factors appeared n_o- free electrons. We accept that at a distance x from the cathode due to the avalanche process the electrons reach concentration $n_x > n_o$. Then, the increasing of the electrons dn for a distance dx is valid Eq.1:

$$dn = (\alpha - \zeta) n_x dx \tag{1}$$

- α - α - shock ionization coefficient
- ζ - coefficient, considering the reduction of electrons due to the recombination and their acceptance by the atoms, which become negative ions.

After integrating Eq.1:

$$n = n_0 e^{(\alpha - \zeta)x} \tag{2}$$

For $x=l$ the number of additional electrons formed by the avalanche process and hit the anode is n - Eq.3:

$$n = n_0 \left[e^{(\alpha-\zeta)l} - 1 \right] \tag{3}$$

If γ_{ph} is photo-ionization coefficient, then the number of new generated electrons n'_0 is:

$$n'_0 = \gamma_{ph} n_0 \left[e^{(\alpha-\zeta)l} - 1 \right] \tag{4}$$

In that case for the electrons near to the cathode is valid Eq.5:

$$n_1 = \frac{n_0}{1 - \gamma_{ph} \left[e^{(\alpha-\zeta)l} - 1 \right]} \tag{5}$$

And for the electrons near the anode (n_{an}) is valid Eq.6:

$$n_{an} = \frac{n_0 e^{(\alpha-\zeta)l}}{1 - \gamma_{ph} \left[e^{(\alpha-\zeta)l} - 1 \right]} \tag{6}$$

Then, the amplification of the gas discharge medium $k = \frac{n_{an}}{n_0}$ can be presented with Eq.7:

$$k = \frac{e^{(\alpha-\zeta)l}}{1 - \gamma_{ph} \left[e^{(\alpha-\zeta)l} - 1 \right]} \tag{7}$$

For creation of independent discharge is necessary $k \to \infty$ i.e. Eq.7 for an independent discharge becomes Eq.8:

$$(\alpha - \zeta) l = \ln \left(1 + \frac{1}{\gamma_{ph}} \right) \tag{8}$$

If we suggest for the α - shock ionization coefficient Eq.9:

$$\alpha = P_g A e^{-\frac{A U_i P_g l}{U}} \tag{9}$$

where:

- A – constant characterizing the gas
- P_g – pressure of the gas
- U_i – ionization voltage
- U - voltage applied between the electrodes, which create the field

From Eq.8 and Eq.9 can be defined the minimum between the electrodes, which guarantees inducing of independent discharge- Eq.10:

$$U = \frac{A U_i P_g l}{\ln P_g A l - \ln \left[\zeta l + \ln \left(1 + \frac{1}{\gamma_{ph}} \right) \right]} \tag{10}$$

The specific electrical conductivity γ is an integral characteristic of the discharge and generally can be defined from Eq.11:

$$\gamma = \frac{n_e e^2 \lambda_{av}}{2 m_e \upsilon_h} \tag{11}$$

where:

- n_e – concentration of the electrons
- e - charge of the electrons
- λ_{av}- free running average length of the electrons $\lambda_{av} = \frac{1}{\pi n_a r^2}$

(n_a – concentration of the atoms; r – radius of the gas atoms)

- m_e - electron weight
- υ_h - heat velocity of the electrons for gases $\upsilon_h = \sqrt{3kT_e m_e^{-1}}$ (k - Boltzmann constant; T_e – temperature of the electron gas)

After transforming Eq.11 becomes Eq.12:

$$\gamma = \frac{n_e}{n_a T_e^{0,5}} \cdot \frac{e^2}{2\pi\sqrt{3}k^{0,5} m_e^{0,5} r^2} \qquad (12)$$

For a discharge channel with length l_{ch} and radius r_{ch} the discharge voltage U_{CD} as a function of the discharge current I_{CD} – $U_{CD} = f_1 (I_{CD})$ – Eq.13 can be received after transforming Eq.12:

$$U_{CD} = I_{CD} \frac{n_a T_e^{0,5}}{n_e} \cdot \frac{2\sqrt{3}k^{0,5} m_e^{0,5} r^2 l_{ch}}{r_{ch}^2 e^2} \qquad (13)$$

$$\frac{dU_{CD}}{dI_{CD}} = U_{CD} \left(\begin{array}{l} \frac{1}{I_{CD}} + \frac{1}{n_a} \cdot \frac{dn_a}{dI_{CD}} + \\ \frac{0,5}{T_e} \cdot \frac{dT_e}{dI_{CD}} - \frac{1}{n_e} \cdot \frac{dn_e}{dI_{CD}} \end{array} \right) \qquad (14)$$

Eq.14 assumes a principle possibility for descending $\left(\frac{dU_{CD}}{dI_{CD}} < 0\right)$ or ascending character $\left(\frac{dU_{CD}}{dI_{CD}} > 0\right)$ of the volt-ampere characteristic or the presence of such sections on it. Most often, at the high current values of I_{CD} forms a ascending section, which can be presented by Eq.15:

$$U_{CD} = \sqrt{\lambda_1 I_{CD}} \qquad (15)$$

where λ_1 is a proportion coefficient - λ_1, Ω^2.A.

We accept that the liquid medium is technically clean i.e. there are dissolved impurities (gases and others) in it. A triggering pulse is applied to the trigatron and electrical discharge occurs. The resistance of the trigatron rapidly decreases and the voltage of the charged work capacitor C is fed to the electrode system, which leads to [1]:

- ionization of the gas in the dissolved gas inclusions. They can originate from local overheating under the influence of applied voltage due to the presence of impurities in the technically clean liquid medium.
- shifting of free charges in the ionized gas inclusions and concentration in the border regions under the influence of external electrical field. The ionized gas inclusions are turned into electrical dipoles.
- at the expense of Coulomb forces – chaining together of the created electrical dipoles and creating a channel with high electrical conductivity.

Therefore, according to above stated the VAC of the liquid medium can also be described by Eq.16:

$$U_L = \sqrt{\lambda_2 I_L} \qquad (16)$$

On Fig.4 are shown the dependencies $U_{CD}=U_{CD}$ (t), $I_{CD}=I_{CD}$ (t), $U_D=U_D$ (t) and $I_D=I_D$ (t). U* can be defined from Eq.10 [1].

We assume that in the discharge circuit is met the condition for aperiodic discharge - $r + R_{air} + R_L > 2\sqrt{LC^{-1}}$

For the discharge circuit on Fig.1 is valid Eq.17:

$$u_C(t) + L\frac{di(t)}{dt} + \sqrt{\lambda_1 i(t)} + \sqrt{\lambda_2 i(t)} + r.i(t) = 0 \qquad (17)$$

Eq.17 can be also presented as Eq.18:

$$\frac{1}{C}\int i(t)dt + L\frac{di(t)}{dt} + \sqrt{\lambda_1 i(t)} + \sqrt{\lambda_2 i(t)} + r.i(t) = 0 \qquad (18)$$

After differentiating and transforming is received Eq.19:

$$\frac{d^2 i(t)}{dt^2} + \left[\frac{\sqrt{\lambda_1} + \sqrt{\lambda_2}}{2L\sqrt{i(t)}} + \frac{r}{L} \right] \frac{di(t)}{dt} + \frac{1}{LC} i(t) = 0 \qquad (19)$$

Eq.19 is a nonlinear parametrical equation. It represents a model of the discharge circuit, considering the presence of two discharge gaps. Up to the present moment, the performed check shows that Eq.19 doesn't have analytical solution.

The numerical solution of Eq.19 is done on the basis of the direct method of Euler, using MATLAB. The results are shown on Fig.5 for U_C=10 kV, C=1 μF, L=0,1 μH, r =0,5 Ω.

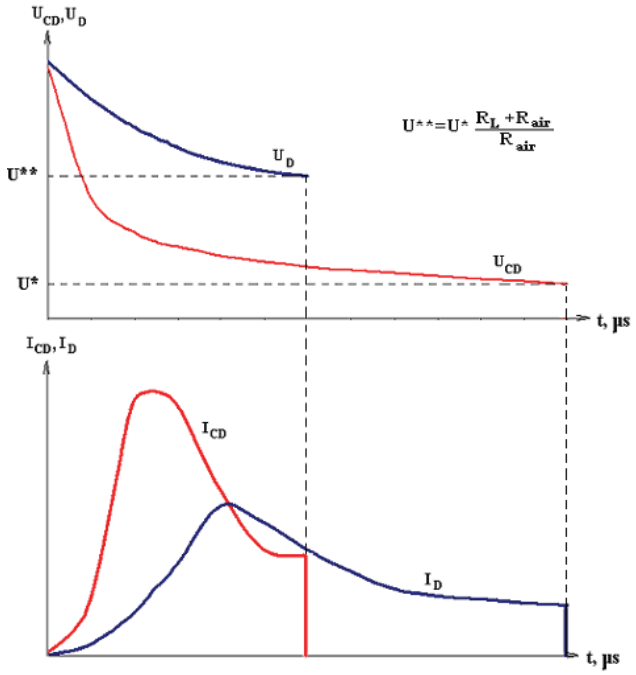


Fig.4

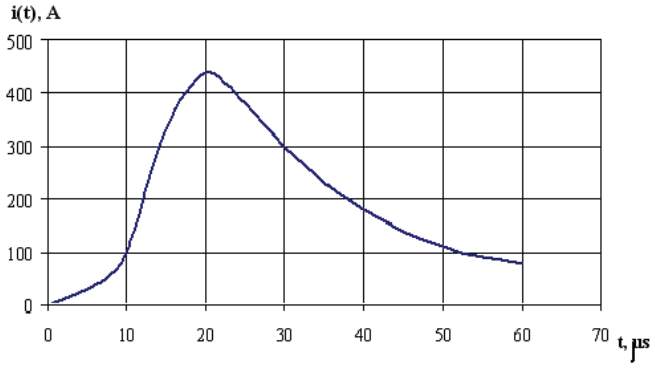


Fig.5

Experimental analyses in a discharge circuit with the configuration from Fig.1 and the above mentioned parameters show that:

- U* (Fig.4) is 700V and it is changing by ±15% as a function of the frequency of the discharge pulses. The experiments are conducted for single discharge pulses and a series with repetition frequency up to 5Hz.
- U** (Fig4) is about 2,2 kV and should be considered when defining the pulse energy.
- The experimental curve $i(t)=f_2(t)$ is above the analytically received one with approximately 27%. When commenting the accuracy, it should be mentioned that the parasitic inductivities and resistances of the capacitor battery for the calculations are taken from the catalogue data. And the accepted presentation of VAC describes well the region of the high currents.

III. Conclusion

A nonlinear model of a discharge circuit with two discharge gaps is suggested. The numerical solution of the nonlinear parametrical differential equation can be used for defining the parameters of the discharge pulse and their conformity with the requirements for certain technological application.

References

[1] Barudov S.T. Design methods and practice for gas discharge control facilities, PGTU – Perm, Russia, 2004, ISBN 954-20-02-79-3

[2] Imada G., M.Suzuki, W.Masuda, Observation of Discharge Instability Induced by Shockwave in Self-Sustained High-Pressure Pulsed Glow Discharge, IEEE Transactions on Plasma Science, vol.35, №4, August 2007, pp.1126-1134

[3] Yusupaliev U., Model of Pulsed Electric Discharge Expansion in Dense Gas Taking into Account Electron and Radiative Thermal Conductivities. I. Expansion Mechanism, Bulletin of the Lebedev Physics Institute, 2009, Vol. 36, No. 8, pp. 239–244, ISSN 1068-3356.

[4] Remy J., B. Broks, W. Brok, J. Mullen, A. Benidar, L. Biennier, F. Salama, Modeling the influence of interelectrode spacing in the Pulsed Discharge Nozzle American Physical Society, 49th Annual Meeting of the Division of Plasma Physics, November 12-16, 2007.

[5] Elizondo, J.M., K.R. Prestwich, J. Corley, D.L. Johnson, P. Corcoran, J. Woodworth, J. Lehr, K. Struve, Modeling of High Current Water Discharges for High Energy Transfer Pulsed Power Systems to Determine Steady State Conduction Resistance, Plasma Science, 2004, IEEE Conference Record – Abstracts, The 31st IEEE International Conference on ICOPS, 28 June-1 July 2004, ISBN: 0-7803-8334-6.

Image Quality of Planar InP Detector

Maja Đekić[1] and Hasna Šamić[2]

Abstract **- In this paper we consider InP for solid state X-ray imaging detectors in the photon energy range of medical applications. Taking into account its physical properties, we calculate spatial resolution and contrast of the pixellated InP detector and examine the effect of the fluorescence on image quality. Finally, we compare InP with materials that are currently used for medical imaging.**

Keywords **- radiation detectors, InP, fluorescence, image quality**

I. Introduction

The goal of digital radiography with efficient measurement of absorption and X-ray scattering is to provide a high quality image with good spatial and energy resolution while decreasing patient radiation dose. Semiconductors have proved to be suitable for construction of planar detectors for medical imaging. Commercially available semiconductors can be divided into two categories: elementary and compound semiconductors. Both categories have pros and cons when used for detection of X-rays and creation of medical image.
The most developed production technology is for elementary semiconductors Ge and Si [1], [2], but both of them show high resolution only while cooled. They are suitable for detection of low energy radiation, Ge because of its small gap and Si because of its high bulk leakage current.
Consequently, compound semiconductors that have wider gap and higher atomic number Z have been studied as good candidates for high-resolution detectors operating at room temperatures. When, beside detection efficiency, quality of retrieved image is also considered, one has to take into account fluorescence that degrades spatial resolution and contrast. The yield and energy of fluorescence photons increase with atomic number Z [3], so when choosing suitable semiconductor for detector construction, a compromise should be made between absorption efficiency (high Z) and decrease of Z in order to obtain good contrast and spatial resolution.

Hence, this paper deals with spatial resolution, signal-to-noise ratio (SNR) and contrast for planar pixellated detector made from InP. We compare this material with other compound semiconductors.

[1] Maja Đekić is with the Faculty of Science, Zmaja od Bosne 33-35, 71 000 Sarajevo, Bosnia and Herzegovina, Email: mdjekic@pmf.unsa.ba
[2]Hasna Šamić is with the Faculty of Electrical Engineering, Zmaja od Bosne bb, 71 000 Sarajevo, Bosnia and Herzegovina, Email: hsamic@etf.unsa.ba

II. Material supply

Performances of semiconductor detectors are limited with production technologies of quality large-area monocrystals, free of defects and with uniform electric properties.
CdTe, a dominant material for compound radiation detectors is still not available in large areas and its production is limited to a small number of companies [4]. GaAs and InP offer photon attenuation coefficients between those of Si and CdTe and are being developed principally for X-ray imaging applications. GaAs in particular has relatively mature contact technologies, but studies of SI bulk GaAs have not progressed beyond prototype demonstrators due to the inhomogeneous nature of the materials and the high concentration of the charge trapping EL2 defects which decreases $\mu\tau$ product. In recent years, there has been some progress in production of epitaxial large-area GaAs [5] with uniform electric properties and lowered defect concentration. InP and HgI_2 seem to be good candidates for high quality imaging detectors because of their atomic number and energy gap.

When X-ray photon interacts with semiconductor, it generates e-h pairs. When external electric field is applied, electrons and holes move into opposite directions causing generation of electric current. In order to achieve full charge collection on both contacts, semiconductor must have a large carrier mobility ($\mu_{e,h}$) and long trapping lifetime ($\tau_{e,h}$). It means that good charge collection efficiency requires that mean drift length $\lambda_{e,h} = \mu_{e,h}\tau_{e,h}E$ be longer than semiconductor crystal thickness. On the other hand, incident photon absorption probability increases with semiconductor's thickness, thus a compromise is usually made between high photon detection and high charge collection efficiency whereas semiconductor thickness is of the order of magnitude of carriers' mean free path. Some physical properties of semiconductors considered in this paper are given in Table I.

TABLE I
PHYSICAL PROPERTIES OF COMPOUND SEMICONDUCTORS

SC	Z	Eg (eV)	$\mu_{e,h}\tau_{e,h}$ (cm^2/V)	k_α (eV)	ω_k (%)	α_k (cm^{-1})
InP	49 15	1.30	$4{,}8 \times 10^{-6}$, $\leq 10^{-7}$	13	46	252
HgI_2	80 53	2.13	10^{-4} 10^{-5}	42.26	92	100
CdTe	48 52	1.44	2.0×10^{-3}, $\leq 4.010^{-4}$	74.7	86	74.7

III. Fluorescence

When a detector of active area A and thickness t absorbs in a unit of time a photon of flux Φ and energy E_0, the expected free charge N_0 to be generated equals

$$N_0 = \phi A E_0 \varepsilon^{-1}\left(1-e^{-\alpha(E_0)t}\right) \quad (1)$$

where t is semiconductor thickness, α is linear absorption coefficient and ε is energy of e-h pair formation. Because of fluorescence, total charge induced by one absorbed photon in a sphere of radius r around incident point amounts to [7]

$$N(r) = E_0\varepsilon^{-1}\left(1-e^{-\alpha(E_0)t}\right) - \omega_k \alpha_k \varepsilon^{-1} e^{-\alpha(k_\alpha)r} \quad (2)$$

where ω_k is fluorescence yield, and $\alpha(k_\alpha)$ absorption coefficient of fluorescence photons.

The total charge induced in a sphere of radius r around the incident photon is presented in Fig. 1. InP reaches maximum efficiency of charge creation for energies around 40 keV and CdTe and HgI_2 around 50-60 keV. InP is more efficient than CdTe and HgI_2 in the energy range of medical applications.

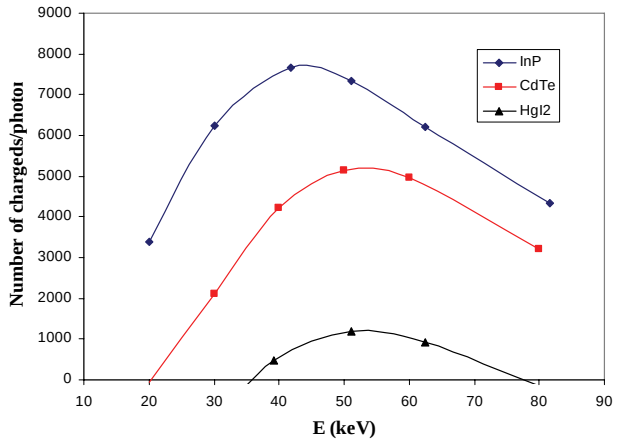


Fig.1 Number of charges per photon versus E(keV)

Considering that fluorescence around the incident photon degrades spatial resolution, we calculated distance R for which 90% of k_α photons absorption will occur. The results shown in Table II indicate that the impact of fluorescence on spatial resolution can be neglected for distances around R=90 μm for InP, but not for CdTe and HgI_2.

TABLE II
DISTANCE R FOR WHICH WE CAN NEGLECT FLUORESCENCE FOR InP, HgI_2, CdTe

Semiconductor	InP	HgI_2	CdTe
R (μm)	91	230	308

IV. Image Quality

Spatial resolution, noise and contrast determine image quality [8] and they are inter-related as follows:

$$SNR = C\sqrt{\phi A} \quad (3)$$

where SNR is signal-to-noise ratio, C contrast, Φ incident photon flux and A is object area.

Signal-to-noise ratio can be calculated as

$$SNR = \frac{N_{ON} - N_{OFF}}{\sqrt{N_{ON}}} \quad (4)$$

and contrast as

$$C = \frac{N_{ON} - N_{OFF}}{N_{ON}} \quad (5)$$

where N_{OFF} is signal inside a pixel with object, and N_{ON} is signal in adjacent radiated pixels.

In order to evaluate detector's SNR, we considered a model of pixellated detector as shown in Fig. 3, calculating signal as a number of generated e-h pairs inside the pixel, taking into account fluorescence in adjacent pixels.

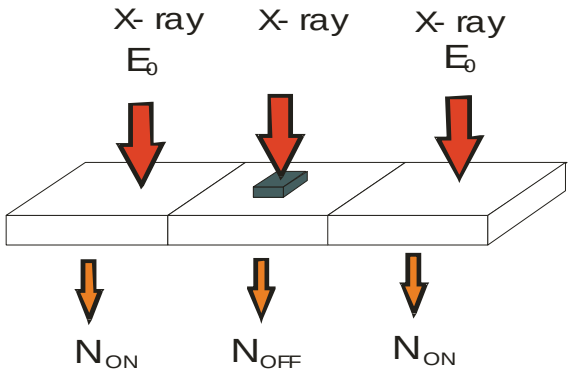


Fig 3. Detector (three pixels and object)

Fig 4. shows the results of SNR calculations depending on object size for a 100 μm pixel, photon flux ~10^8 photons/cm^2, detector thickness 200 μm and incident photon energy 30keV.

Fig 5. shows SNR as a function of incident radiation energy for a 100 μm pixel and 5μm object.

It is evident that the best SNR is achieved for InP, which means that radiation energy for CdTe and HgI_2 should be increased in order to get the same result.

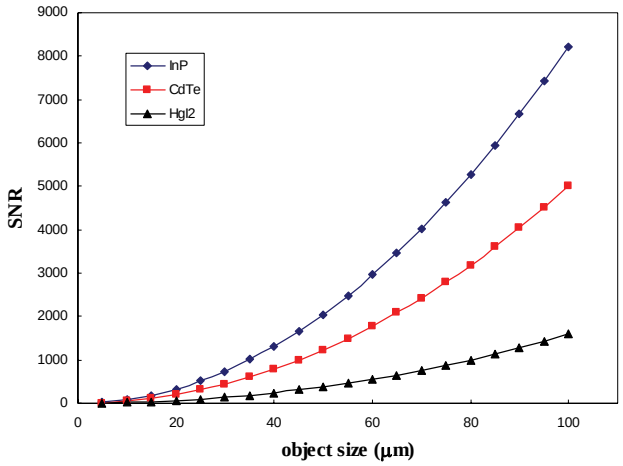


Fig 4. SNR versus object size for a 100 μm pixel and 30 keV energy for InP, CdTe and HgI_2

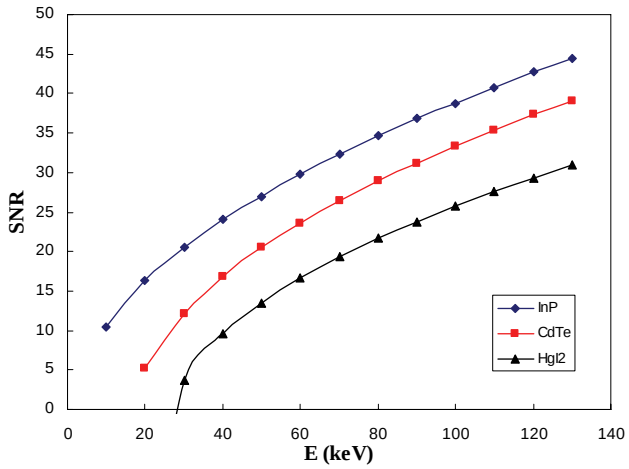


Fig 5. SNR versus incident photon energy for InP,CdTe and HgI_2

Using the same detector model, with an object smaller than a pixel that completely absorbs radiation, we have calculated contrast for different materials depending on the energy of incident X-rays.

The results show (Fig. 6) that InP has a very good contrast (about 1) for the radiation energies considered and much better contrast than CdTe for the lower radiation energies. CdTe has considerably smaller contrast for lower radiation energies, but as the incident photon energy increases, its contrast improves. Both intermediate-Z semiconductors have much better contrast than high-Z HgI_2 in the energy range of medical imaging.

As the aim of medical imaging is detection of early changes inside the tissue, i.e. detection of objects as small as possible, we calculated minimum object size that could be visible in conditions of mammographic imaging, in order to choose the best material for detector construction. During mammographic imaging, a problem is to differentiate tumorous tissue from surrounding tissue because of very similar absorption coefficients. According to Rose model [8], human eye is able to differentiate these changes if SNR is equal to 5.

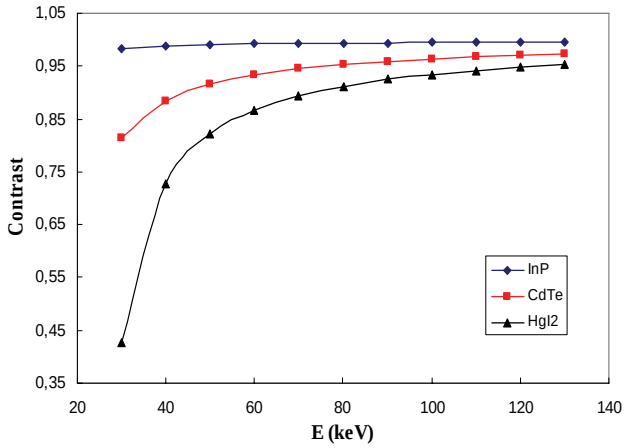


Fig 6. Contrast versus energy for InP, CdTe and HgI_2

Starting from this point and taking for E = 30 keV, absorption coefficient of tumorous tissue [9] (carcinoma) α_c =0,5 cm^{-1}, and of surrounding tissue [9] of thickness 5cm , α_t = 0,25 cm^{-1}, we have calculated minimum size of visible carcinoma in conditions typical for mammography, radiation dose of 100μGy and pixel size 150 μm .

TABLE III
MINIMUM SIZE OF VISIBLE OBJECT FOR 100 μGy RADIATION DOSE FOR InP, CdTe AND CsI

Semiconductor	InP	CdTe	HgI_2
Object size (μm)	30	47	92

The results show that minimum visible size of cubic object for radiation dose of 100μGy is 30μm for InP while for HgI_2 is 92μm. That means that we would have to drastically increase radiation dose in order to see the same size carcinoma with HgI_2 instead of InP.

Finally, we compared contrast, visible object size and radiation dose for considered materials for energy of 30 keV, and pixel size 150μm .

TABLE IV
RELATIVE VALUES OF CONTRAST, MINIMUM OBJECT SIZE AND DOSE FOR InP, CdTe AND HgI_2 FOR LOW ENERGY (30 KEV) AND PIXEL SIZE 150 μm

Relative values	InP	CdTe	HgI_2
Contrast	1	0.83	0.59
Minimum size	1	1.5	3
Dose	1	2	10

It is obvious from these results given in Table IV, that for medical imaging for low energies around 30 keV, there is a considerable reduction of radiation dose with the same contrast and minimum visible carcinoma size for InP instead of CdTe. Comparing InP and CdTe with HgI_2, it follows that intermediate-Z compound semiconductors are more suitable for X-ray medical imaging than high-Z HgI_2 because of improvement of contrast and reduction of radiation dose.

V. CONCLUSION

Our results show that for low X-ray energies and pixel size of 150 μm in conditions of low contrast according to Rose, minimum visible object size is 3 times smaller for InP, and a dose needed for detection of the object around 10 times smaller than for detectors based on HgI_2. HgI_2 offers good detection efficiency at higher photon energies and is suitable for hard X-ray and nuclear medicine imaging.

According to our evaluations, InP has better contrast than CdTe that is currently used for medical imaging. In order to see the same-size of carcinoma with the same contrast, dose must be 2 times higher for CdTe than for InP. Unfortunately, the detectors based on InP, which is a very soft material, are limited with production technology. The lack of rectifying contacts on SI InP produces high leakage currents in the devices, which must be reduced by cooling. Future improvements in InP detector performance depend on the development of the rectifying contacts. We can conclude that in the group of considered materials, InP seems to be the best choice for obtaining good response to X-ray radiation of energies 20-60 keV, good contrast and lower radiation dose.

REFERENCES

[1] G.Bertolini and A. Coche, *Semiconductor detectors*, Amsterdam, Norh-Holland Publ.Co., 1968.

[2] S.M.Sze, *Physics of Semiconductor Devices*, New York, Wiley 1981.

[3] S.A. Moszkowski in K.Siegbahn ed.,*Alpha-,Beta-,and Gamma-Ray Spectroscopy*, Amsterdam,Norh- Holland Publ.Co., 1965.

[4] P.J.Sellin, "Recent advances in compound semiconductor radiation detectors" Nucl.Instr. and Meth. A,vol. 513, pp. 332-339, 2003.

[5] H.Samic ,et al.,"Characterization of Thick Epitaxial GaAs Layers for X-ray Detection", Nucl.Instr.and Meth. A,vol. 487, pp. 107-112, 2002.

[6] A. Owens, et al., Nucl.Instr. and Meth. A, "The X-ray response of InP",vol. 487, pp. 435-440, 2002.

[7] N. Mañez, at al., Nucl.Instr. and Meth. A, "Material optimization for X-ray imaging detectors",vol. 567, pp. 281. 2006.

[8] A.Rose, *Vision: Human and Electronic*, New York,Plenum, 1973.

[9] National Institute of Standards and Technology, http://physics.-nist.gov.

Investigations of Power Losses in Off-Chip and On-Chip Inductors

Tihomir S. Brusev[1] and Boyanka M. Nikolova[2]

***Abstract* – This paper includes investigations of power losses in off-chip and on-chip filter inductors in monolithic dc-dc converter designed on CMOS 0.35 µm technology. Chip coils of the company Murata are used for off-chip inductors. For the extraction of the model's parameters of integrated inductors of CMOS 0.35 µm process is used one of the Cadence tools named "Virtuoso Passive Component Designer". Comparison between power losses in off-chip and on-chip inductors is made.**

***Keywords* – power losses, off-chip and on-chip inductors, dc-dc converters.**

I. INTRODUCTION

The fully monolithic dc-dc converters can reduce the cost and size of the battery powered portable electronic devices. One of the most important parameters in the dc-converters is efficiency:

$$\eta = \frac{P_{OUT}}{P_{IN}}, \qquad (1)$$

where P_{OUT} is the average output power, P_{IN} is the average input power of the converter. The losses in power stage of buck converter are much higher compare to the energy dissipation in the feedback control system [1]. They dominate and determine the efficiency of the system. The power losses in the filter inductor can seriously decrease the overall efficiency of the converter's system. It's very important to be investigated and compared power dissipations when off-chip and on-chip filter components are used. Integrated inductors occupied huge silicon area. Therefore is necessary to be estimated which type of inductors is more suitable for the designed dc-dc converter.

This paper presents the investigations results of power losses in filter inductor of monolithic dc-dc converter designed on CMOS 0.35 µm technology in buck converter implemented on standard CMOS 0.35-µm technology. In Section II are presented the received simulations results with Cadence. This section is divided on two parts. In Section II *A* are shown the received results for power dissipations in the off-chip filter inductor. For the investigations are used chip coils of the company Murata. In Section II *B* are presented power losses when on-chip filter inductors of CMOS 0.35 µm process are used. Influence of the low Q-factor of integrated filter inductor over the dc-dc converter efficiency is evaluated.

II. POWER LOSSES IN OFF-CHIP AND ON-CHIP INDUCTORS

In Fig. 1 is shown the circuit of a buck converter implemented on standard CMOS 0.35 µm technology. For the power losses investigation of off-chip and on-chip filter inductors an input voltage of 3.6V is chosen for the converter, since this is the normal voltage for Lithium-Ion battery cell that is typically used in battery-powered devices. The output voltage V_{out} is regulated to 1.2V, which on the other hand is determined by the standard supply voltages of advanced CMOS processes.

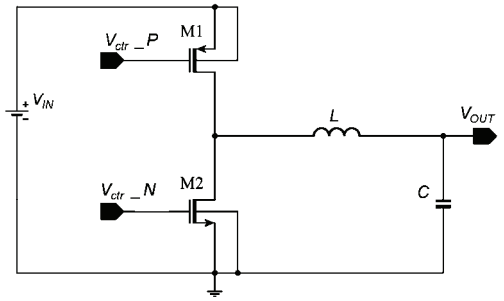


Fig. 1. Synchronous buck converter schematic.

Part of the total power losses in the monolithic switching mode buck dc-dc converter is due to the series resistance and parasitic capacitance of the filter inductor. Integrated inductors have high series resistance, which grows up with increasing of the inductor's value. This lead to the low quality factor *Q* of these components [2]. Integration of a spiral inductor with sufficient inductance is not feasible, because they occupy huge silicon area and have low *Q* factor. The total power dissipated in filter inductor, assuming that the inductor parasitic impedance scale linearly with the inductance is equal to [3]:

$$P_{ind} = b\left(\frac{I^2}{\Delta i_L f_s} + \frac{\Delta i_L}{3 f_s} + \frac{C_{L0} V_{DD1}^2}{R_{L0} \Delta i_L}\right), \qquad (2)$$

where b is a coefficient depending from the parasitic capacitance and parasitic series resistance of the filter inductor, C_{L0} and R_{L0} are respectively the parasitic stray capacitance and parasitic series resistance per 1 nH inductance, V_{DD1} is power supply.

[1]Tihomir S. 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]Boyanka M. Nikolova is with the Faculty of Telecommunications, Technical University of Sofia, Kl. Ohridski 8, 1797 Sofia, Bulgaria, E-mail: bnikol@tu-sofia.bg

The influence of switching frequency f_s and inductor current ripple Δi_L over the power losses in the filter inductor are investigated. Integrated passive inductors occupied huge silicon area, which make more expensive the electronic devices. It's necessary to be estimated energy dissipation in off-chip and on-chip filter inductors.

Power losses in the filter inductor of the synchronous buck converter designed on CMOS 0.35 µm technology are analyzed with Cadence, as a function of switching frequency f_s and inductor current ripple Δi_L. Evaluated and compared are power losses in off-chip and on-chip filter inductor of monolithic buck dc-dc converter.

A. Power losses in the off-chip filter inductor

In this section is evaluated power losses in off-chip filter inductors of monolithic dc-dc converter designed on CMOS 0.35-µm technology. The investigated inductors are chip coils for high frequency horizontal wire wound of the company Murata. They have broad range of inductance and high self-resonant frequency. This realizes high Q-factor and stable inductance at high frequency. Their low dc resistance is ideal for low power losses. These off-chip inductors are often used in telecommunication applications.

The received results of power losses in the off-chip filter inductor of buck dc-dc converter are presented below. For the simulations with Cadence on CMOS 0.35-µm technology are used inductor's models of the company Murata [4].

In Table I are presented power losses in the off-chip filter inductor P_{ind} as a function of switching frequency f_s and inductor current ripple Δi_L. Average output current of the buck converter $I_{out\ (av)}$ equal to 20 mA and Δi_L=0.5x$I_{out\ (av)}$.

TABLE I

POWER LOSSES IN THE OFF-CHIP INDUCTORS OF BUCK CONVERTER; $\Delta I_L/0.5xI_{OUT\ (AV)}$;

	L=10 nH	L=50 nH	L=100 nH	L=200 nH
$I_{out\ (av)}$ [mA]	20	20	20	20
f_s [MHz]	5000	2000	1500	1000
$\Delta i_L/0.5xI_{out\ (av)}$	0.5	0.5	0.5	0.5
P_{ind} [mW]	0.0579	0.148	0.316	1.02

In Table II are presented power losses in the off-chip filter inductor P_{ind} as a function of switching frequency f_s and inductor current ripple Δi_L. Average output current of the buck converter $I_{out\ (av)}$ equal to 20 mA and Δi_L=$I_{out\ (av)}$.

In Table III are presented power losses in the off-chip filter inductor P_{ind} as a function of switching frequency f_s and inductor current ripple Δi_L. Average output current of the buck converter $I_{out\ (av)}$ equal to 20 mA and Δi_L=2x$I_{out\ (av)}$. Received simulations results for power losses in the off-chip filter inductor P_{ind} of buck converter implemented on CMOS 0.35-µm technology as a function of f_s and Δi_L, for $I_{out\ (av)}$=20 mA given in Table I, Table II and Table III, are graphically presented in Fig.2. For simulations the inductor's models of the company Murata are used.

TABLE II

POWER LOSSES IN THE OFF-CHIP INDUCTORS OF BUCK CONVERTER; $\Delta I_L/I_{OUT\ (AV)}$;

	L=10 nH	L=50 nH	L=100 nH	L=200 nH
$I_{out\ (av)}$ [mA]	20	20	20	20
f_s [MHz]	2500	1000	750	500
$\Delta i_L/I_{out\ (av)}$	1	1	1	1
P_{ind} [mW]	0.056	0.183	0.333	1.02

TABLE III

POWER LOSSES IN THE OFF-CHIP INDUCTORS OF BUCK CONVERTER; $\Delta I_L/2xI_{OUT\ (AV)}$;

	L=10 nH	L=50 nH	L=100 nH	L=200 nH
$I_{out\ (av)}$ [mA]	20	20	20	20
f_s [MHz]	1250	500	350	250
$\Delta i_L/2xI_{out\ (av)}$	2	2	2	2
P_{ind} [mW]	0.062	0.139	0.291	0.97

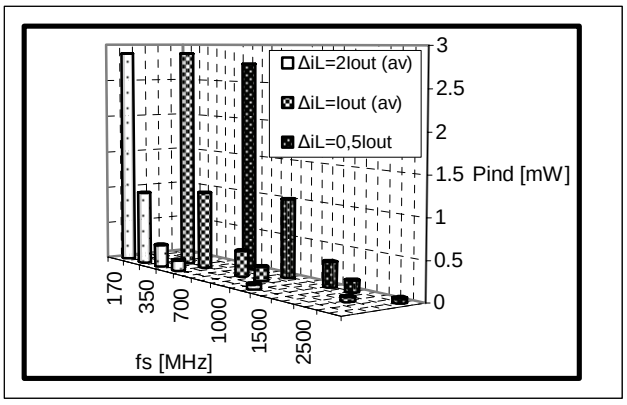


Fig. 2. Power losses in the off-chip filter inductor of buck converter implemented on CMOS 0.35-µm technology as a function of f_s and Δi_L, for $I_{out\ (av)}$=20 mA.

Power losses in the off-chip filter inductor of buck dc-dc converter decreased with increasing of switching frequency f_s and inductor current ripple Δi_L.

B. Power losses in the on-chip filter inductor

The dependence of the power dissipation in on-chip filter inductors of CMOS 0.35-µm technology are investigated and evaluated. Standard monolithic inductors are not optimized for the specific applications. In the other hand theirs number of values are limited. That's way is necessary to be designed integrated inductors with desired value and geometry. One of the Cadence tools named "Virtuoso Passive Component

Designer" is used for the extraction of inductor's model parameters [5].

Fig. 3. Window from CADENCE tool "Virtuoso Passive Component Designer".

The model of on-chip inductors used in the investigations is shown in Fig. 4. The parameters presented in Fig. 4 are: L_s – series inductance; R_s – series resistance; C_p – parallel capacitance; $C_{OX1/2}$ – oxide capacitance; $C_{s1/2}$ – substrate capacitance; $R_{s1/2}$ – substrate resistance.

The values of the extracted parameters are used for simulations. The received results of power losses in the on-chip filter inductor of buck dc-dc converter are presented below. Investigations are made for average output current of the buck converter $I_{out\ (av)}$ equal to 20 mA.

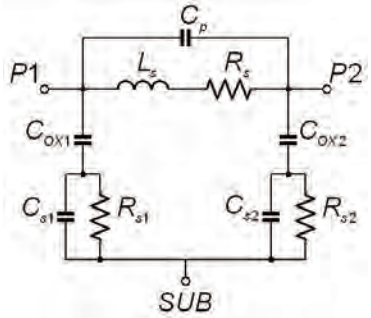


Fig. 4. The model of on-chip inductors.

In Table IV are presented power losses in the on-chip filter inductor P_{ind} as a function of switchnig frequency f_s and inductor current ripple Δi_L, for Δi_L=0.5x$I_{out\ (av)}$.

In Table V are presented power losses in the on-chip filter inductor P_{ind} as a function of switching frequency f_s and inductor current ripple Δi_L. Average output current of the buck converter $I_{out\ (av)}$ equal to 20 mA and Δi_L=$I_{out\ (av)}$.

In Table VI are presented power losses in the on-chip filter inductor P_{ind} as a function of switching frequency f_s and inductor current ripple Δi_L. Average output current of the buck converter $I_{out\ (av)}$ equal to 20 mA and Δi_L=2x$I_{out\ (av)}$.

Received simulations results for power losses in the on-chip filter inductor P_{ind} of buck converter implemented on CMOS 0.35-µm technology as a function of f_s and Δi_L, for $I_{out\ (av)}$ = 20 mA given in Table IV, Table V and Table VI, are graphically presented in Fig. 5.

TABLE IV
POWER LOSSES IN THE ON-CHIP INDUCTORS OF BUCK CONVERTER; $\Delta I_L/0.5xI_{OUT\ (AV)}$;

	L=10 nH	L=14.1 nH	L=28.2 nH	L=51 nH
$I_{out\ (av)}$ [mA]	20	20	20	20
f_s [MHz]	5000	4000	3000	2000
$\Delta i_L/0.5xI_{out\ (av)}$	0.5	0.5	0.5	0.5
P_{ind} [mW]	15.37	15.9	16.35	23.59

TABLE V
POWER LOSSES IN THE ON-CHIP INDUCTORS OF BUCK CONVERTER; $\Delta I_L/I_{OUT\ (AV)}$;

	L=10 nH	L=14.1 nH	L=28.2 nH	L=51 nH
$I_{out\ (av)}$ [mA]	20	20	20	20
f_s [MHz]	2500	2000	1500	1000
$\Delta i_L/I_{out\ (av)}$	1	1	1	1
P_{ind} [mW]	15.37	15.74	16.37	23.6

TABLE VI
POWER LOSSES IN THE ON-CHIP INDUCTORS OF BUCK CONVERTER; $\Delta I_L/2xI_{OUT\ (AV)}$;

	L=10 nH	L=14.1 nH	L=28.2 nH	L=51 nH
$I_{out\ (av)}$ [mA]	20	20	20	20
f_s [MHz]	1250	1000	750	500
$\Delta i_L/2xI_{out\ (av)}$	2	2	2	2
P_{ind} [mW]	15.37	15.5	16.3	23.59

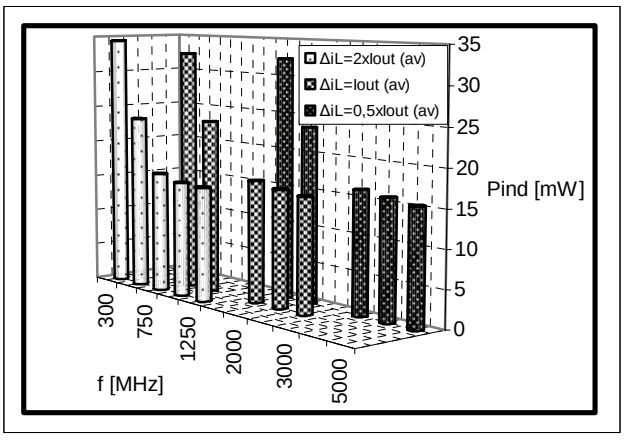


Fig. 5. Power losses in the on-chip filter inductor of buck converter implemented on CMOS 0.35-µm technology as a function of f_s and Δi_L, for $I_{out\ (av)}$=20 mA.

As can be seen from Fig. 2 and Fig. 5 power losses in the filter inductor of buck dc-dc converter goes down with increasing of switching frequency f_s and inductor current ripple Δi_L. When off-chip filter inductors of the company Murata are used, power dissipation is around 10 times lower compare to the losses in integrated inductors of CMOS 0.35 µm process (Fig. 2, Fig. 5). The reason is the high series resistance and low Q-factor of the on-chip inductors.

The efficiency of the designed monolithic dc-dc converter as a function of on-chip filter inductors is evaluated. Investigations are made at different average outputs currents ($I_{out\ (av)}$) of the circuit. In Table VII are shown received results at $I_{out\ (av)}$ = 5 mA. In Table VIII are shown received results at $I_{out\ (av)}$ = 20 mA. The simulation results shown in Table VII and Table VIII are graphically presented in Fig. 6.

TABLE VII
EFFICIENCY OF THE DC-DC CONVERTER

	L=1.4 nH	L=2.6 nH	L=4.7 nH	L=9 nH
$I_{out\ (av)}$ [mA]	5	5	5	5
η [%]	8.7	13.9	15.4	15.5

TABLE VIII
EFFICIENCY OF THE DC-DC CONVERTER

	L=1.4 nH	L=2.6 nH	L=4.7 nH	L=9 nH
$I_{out\ (av)}$ [mA]	20	20	20	20
η [%]	24.7	34.4	36.3	33.5

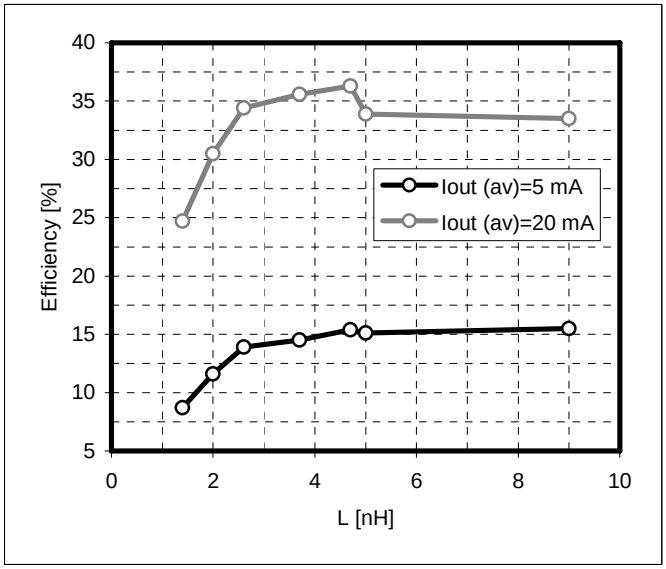


Fig. 6. Efficiency of the monolithic dc-dc converter designed on CMOS 0.35 µm technology as a function of different on-chip inductors.

The maximum value of the converter's efficiency when on-chip filter inductors are used is 36 %. One of the reasons is low Q-factor of the available integrated inductors.

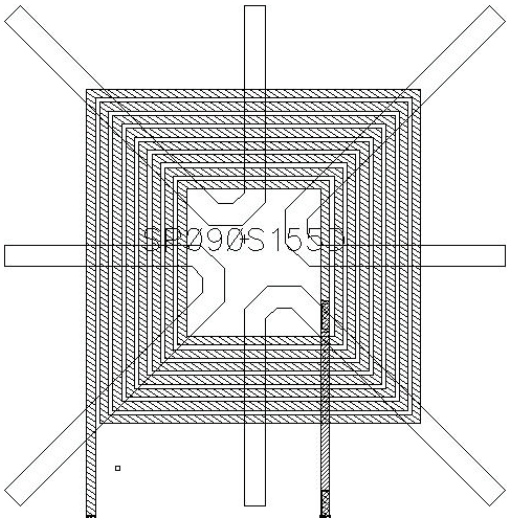


Fig. 7. The layout of on-chip inductors.

III. CONCLUSION

The power losses in the off-chip and on-chip filter inductors of monolithic buck dc-dc converter designed on CMOS 0.35 µm technology has been investigated. When off-chip filter inductors of the company Murata are used, power dissipation is around 10 times lower compare to the losses in integrated inductors of CMOS 0.35 µm process (Fig. 2, Fig. 5). The received results show that by increasing of switching frequency f_s of the circuit power dissipations in the filter inductor decreased. By increasing of the Δi_L requirements for the inductor's value decreased for fixed switching frequency f_s. This reduces the parasitic impedance of the filter inductor and the related power loss. A higher value of the Δi_L leads to the increasing of rms current through the inductor, which causes to the bigger conduction losses in the inductor.

ACKNOWLEDGEMENT

This investigation has been carried out in the framework of the research projects № Д002 – 126/15.12.2008 and № ДТК02/50 – 17.12.2009.

REFERENCES

[1] Kurson V., S. G. Narenda, V. K. De, and E. G. Friedman, "Monolithic DC-DC Converter Analysis and MOSFET Gate Voltage Optimization," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 279-284, March 2003.

[2] Gardner D., A. M. Crawford, S. Wang, "High Efficiency (GHz) and Low Resistance Integrated Inductors Using Magnetic Materials," *Proceedings of the IEEE International Interconnect Technology Conference*, pp. 101-103, June 2001.

[3] Kurson V., "Supply and Threshold Voltage Scalind Techniques in CMOS Circuits", 2004.

[4] http://www.murata.com

[5] http://www.cadence.com

Behavioural VHDL-AMS Model for Monolithic Programmable Gain Amplifiers

Dimo S. Martev[1], Ivan D. Panayotov[2] and Ivailo M. Pandiev[3]

***Abstract* – This paper presents behavioural VHDL-AMS model for monolithic programmable gain amplifiers (PGAs). The proposed model is independent from actual technical realizations and is based upon compromises regarding the representation of exact circuit structures in the model. For creating the PGA model, techniques known from modelling operational amplifiers have been adapted. The model accurately reflects input resistance, transfer functions (amplifier gain in binary steps versus controlling digital code), small-signal frequency responses, output resistance and other basic dc and ac electrical parameters of the real device. Model parameters are extracted for the integrated PGA AD526 from Analog Devices as an example. To confirm the validity, the simulation results are compared with the manufacturer's data, where a good agreement is found between simulations and performance of the actual device.**

***Keywords* – Analog circuits, Programmable gain amplifiers, Operational amplifiers, Mixed-signal simulation, VHDL-AMS.**

I. INTRODUCTION

The monolithic programmable gain amplifiers (PGAs) are an important link between analog and digital signal processing circuits. They have an analog signal input, an analog signal output and a digital control of the voltage amplification. Typical PGAs may be configured either for selectable *binary gains* such as 1, 2, 4, 8 etc, or they might also be configured for *decade gains* such as 10, 100, 1000 etc. In addition, the most of the PGAs contain input and output voltage buffer [1-3].

Testing the workability of electronic circuits with PGAs is done usually using SPICE (Simulation Program with Integrated Circuit Emphasis). A variety of SPICE macromodels for the PGAs, are available in the literature [4-7]. The majority of published macromodels contain extensive number of active and passive elements. Furthermore, testing complete mixed-signal systems with large number of elements is an extremely difficult process and can often become infeasible due to the limitation of simulation capacity.

One method to decrease simulation time and improve the convergence, without a significant loss of information, is by using behavioural modelling technique. Nowadays, one of the most effective techniques for behavioural modelling of mixed-signal electronic circuits is by using VHDL-AMS. To the authors' knowledge, behavioural models of monolithic PGAs have not yet been reported in the literature. It is, therefore, the purpose of this paper to present a behavioural VHDL-AMS model for the most common monolithic PGAs.

[1]Dimo S. Martev is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: dsmartev@yahoo.com

[2]Ivan D. Panayotov is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: idp@ecad.tu-sofia.bg

[3]Ivailo M. Pandiev is with the Faculty of Electronics, TU-Sofia, Kl. Ohridski 8, 1000 Sofia, Bulgaria, E-mail: ipandiev@tu-sofia.bg

II. MONOLITHIC PGAS

The monolithic PGAs are controllable amplifiers in which the voltage gain can be conveniently changed via digital code or dc controlling voltage (current). In fact, the electronic circuits including PGAs are very suitable in a user's view point, because the electrical parameters can be changed electronically and do not contain mechanical potentiometers or analog switches. The PGAs are widely used in mixed-signal applications, such as programmable scaling amplifiers, controllable active filters and oscillators, sonar systems, amplitude-stabilized oscillators, programmable power supplies, etc. The monolithic PGA AD526 from Analog Devices [8] is of particular interest here. The IC AD526 is a digitally programmable gain amplifier with single-ended input and output. Fig. 1 summarizes the external view of AD526. The negative feedback of the amplifier is closed by connecting pin *Sense* with pin *Force*. The controllable digital code, applied to pins

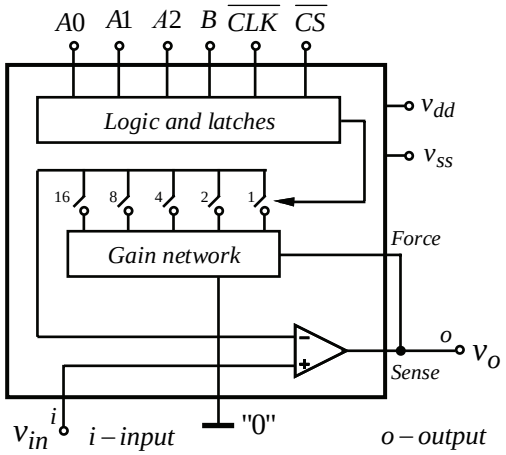


Fig. 1. PGA AD256A external view.

$A0$, $A1$, $A2$, B, $\overline{CLK}$ and $\overline{CS}$, provide voltage gains 1, 2, 4, 8 and 16. The internal structure of this amplifier generally consists of a non-inverting amplifier with a resistor network (or *gain network*), ideal voltage-controlled switches and a control logic block (or *logic and latches*). The control logic block must ensure the operation in either *latched* or *transparent mode*. In *latched mode*, the amplifier AD526 is capable of storing the gain code ($A0$, $A1$, $A2$, B) under the direction of control inputs $\overline{CLK}$ and $\overline{CS}$, which are func-

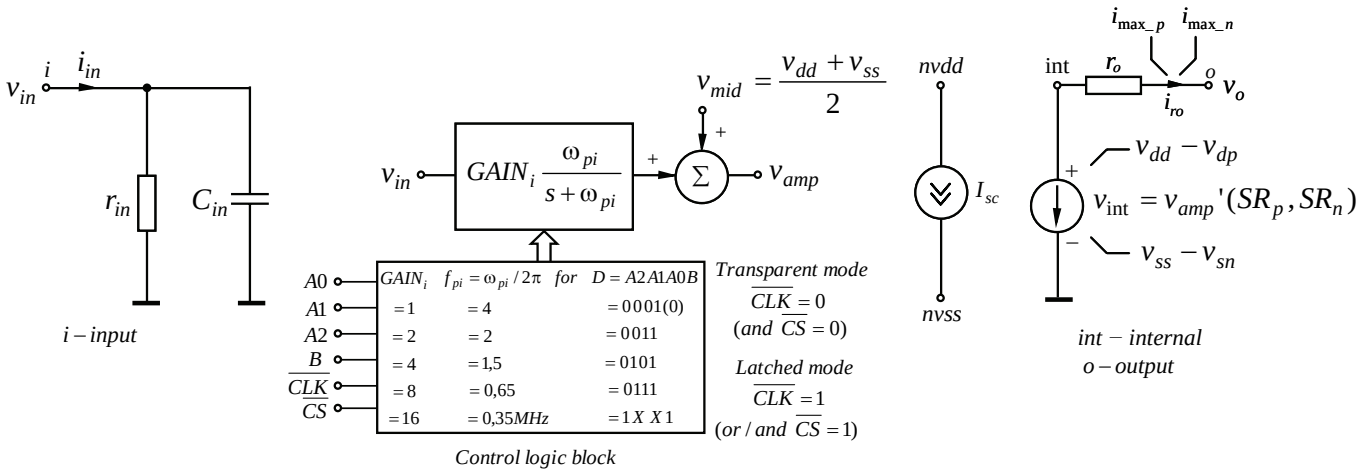


Fig. 2. Circuit diagram of the Programmable Gain Amplifier (PGA) behavioural model.

tionally and electrically equivalent. Alternatively, in *transparent mode* it can respond directly to gain code changes if the control inputs are tied low. In fact, the AD526 is a typical representative of the commercially available PGAs, which are widely used for design and analysis of mixed-signal applications.

The electrical performance of the PGA AD526 is analysed performing electrical simulations with OrCAD PSpice simulator, using PGA macromodel given in [7]. Furthermore, the obtained simulation results are compared with the manufacturer's data. The analyses of the electrical characteristic curves lead to the definition of a behavioural VHDL-AMS model for PGA.

III. Behavioural Modelling with VHDL-AMS

"The small simulation model is beautiful".
Stewart Robinson

The technical requirement for effective models is generally agreed when the simplest model possible is developed. Simple models have a number of advantages. They can be developed faster, are more flexible, require less data, run faster and it is easier to interpret the results, since the structure of the model is better understood. As the complexity increases, these advantages are lost [9]. The proposed behavioural model of the PGA is developed following the design method based on a Top-Down analysis approach and applying simplification and build-up technique. The process of model building and testing can be broken down into three basic steps: structure the model, build the model and validate the model [10].

A. A behavioural language: VHDL-AMS

VHDL-AMS is a comparatively new standard 1076.1 of VHDL that supports hierarchical description and simulation of analog, digital and mixed-signal applications with conservative and non-conservative equations [11, 12]. On the mixed-signal side, a variety of abstraction levels is supported. The VHDL-AMS modelling is not restricted to mixed-signal applications but also support thermal, mechatronic, optical and other systems.

B. A PGA behavioural VHDL-AMS model

The behavioural model of the PGA is built using the results obtained by analyses of the PGA macromodel from [7]. The circuit diagram of the PGA model is shown in Fig. 2, where the different stages are presented with controlled sources and passive components. The model includes the following elements and parameters with numerical values: $r_{in} = 10M\Omega$ and $C_{in} = 10\,pF$ – input resistance and capacitance; $I_{sc} = 10mA$ – dc supply current; $GAIN_i = 1$, 2, 4, 8 and 16, where $i = \overline{1,4}$; $f_{pi} = \omega_{pi}/2\pi = 4$, 2, 1,5, 0,65 and 0,35*MHz* the $-3dB$ frequencies for the dominant poles at gains 1, 2, 4, 8 and 16, respectively; $SR_p = -SR_n = 6V/\mu s$ – positive and negative slew rates at gains 1, 2 and 4; $SR_{pl} = -SR_{nl} = 24V/\mu s$ – positive and negative slew rates at gains 8 and 16; v_{int} – output voltage-controlled voltage source; $v_{dp} = v_{dn} = 3V$ – positive and negative voltage drops for the output voltage limitation; $i_{\max\_p} = -i_{\max\_n} = 30mA$ – maximum output currents; $r_o = 0{,}1\Omega$ – output resistance.

The proposed model includes small- and large-signal effects such as (1) accurate input impedance, (2) amplifier gain in binary steps versus controlling digital code, (3) transparent and latched mode of operation, (4) ac small-signal frequency responses, (5) slew rates, (6) dc supply current, (7) voltage and current limitations, and (8) output resistance.

The mathematical equations that describe the model can be given as

$$i_{in} = v_{in}\left(\frac{1}{r_{in}} + sC_{in}\right) \tag{1}$$

$$v_{amp} = GAIN_i \frac{\omega_{pi}}{s + \omega_{pi}} v_{in} + v_{mid} \tag{2}$$

$$I_{sc} = I_s \tag{3}$$

```
library ieee;
use   ieee.math_real.all;
use   ieee.std_logic_1164.all;
use   ieee.electrical_systems.all;

entity ad526 is
      generic(
   --generic constants here);
   port (terminal input,output,nvdd,nvss:ELECTRICAL;
   signal b,cs_n,clk_n : in STD_LOGIC;
   signal a0,a1,a2: in STD_LOGIC);
end entity ad526;

architecture behavioral of ad526 is
   --constant declarations here
   terminal internal:ELECTRICAL;
   quantity vin across iin,icin through input;
   quantity vdd across nvdd;
   quantity vss across nvss;
   quantity isc through nvdd to nvss;
   quantity vint across iintern through internal;
   quantity vro across iro through internal to
      output;
   quantity vout across output;
   quantity vamp:VOLTAGE;
   quantity iro_h:CURRENT;
   quantity vmid:VOLTAGE;
   signal gain:real:=1.0;
   signal a0_b,a1_b,a2_b:BIT;
   signal b_b,cs_n_b,clk_n_b:BIT;
   signal control_in:BIT_VECTOR(0 to 2);
begin
   isc==supply_current;
   iin==vin/rin;
   icin==cin*vin'dot;
   iro_h==vro/ro;
   vmid==(vdd+vss)/2.0;
   if  gain=1.0 use
    vamp==vin'ltf((0=>wp1*1.0),(0=>wp1,1=>1.0))+vmid;
   elsif gain=2.0 use
    vamp==vin'ltf((0=>wp2*2.0),(0=>wp2,1=>1.0))+vmid;
   elsif gain=4.0 use
    vamp==vin'ltf((0=>wp4*4.0),(0=>wp4,1=>1.0))+vmid;
   elsif gain=8.0 use
    vamp==vin'ltf((0=>wp8*8.0),(0=>wp8,1=>1.0))+vmid;
   else
    vamp==vin'ltf((0=>wp16*16.0),(0=>wp16,1=>1.0))+vmid;
   end use;
   if vamp'above(vdd-vdp) use
       vint==vdd-vdp;
   elsif not vamp'above(vss+vsn) use
       vint==vss+vsn;

   else
    if gain=8.0 or gain=16.0 use
        vint==vamp'slew(SRpl,SRnl);
       else
        vint==vamp'slew(SRp,SRn);
       end use;
      end use;
   if iro_h'above(imax_p) use
      iro==imax_p;
     elsif not iro_h'above(imax_n) use
      iro==imax_n;
     else
      iro==iro_h;
     end use;
     break on gain,
                 vamp'above(vdd-vdp),
                 vamp'above(vss+vsn);
   b_b<=to_bit(b);
   cs_n_b<=to_bit(cs_n);
   clk_n_b<=to_bit(clk_n);
   control_in<=(to_bit(a2),
      to_bit(a1), to_bit(a0));
   case_change:process is
        procedure gain_change
        (signal_in:in BIT_VECTOR;
            signal bu:in BIT;
            signal s:out REAL) is
            begin
             if bu='1' then
                  case signal_in(0 to 2) is
              when b"000" => s <= 1.0;
              when b"001" => s <= 2.0;
              when b"010" => s <= 4.0;
              when b"011" => s <= 8.0;
              when others => s <= 16.0;
                  end case;
             else s <= 1.0;
                   end if;
            end procedure gain_change;
   begin
      if clk_n_b='0' and cs_n_b='0' then
            gain_change(control_in,b_b,gain);
            else
         wait until clk_n_b='0' and cs_n_b='0';
                  gain_change(control_in,b_b,gain);
            end if;
     wait on clk_n_b,cs_n_b,control_in,b_b,gain;
end process case_change;
end architecture behavioral;
```

Fig. 3. A PGA behavioural VHDL-AMS model.

$$v_{mid} = (v_{dd} + v_{ss})/2 \tag{4}$$

$$SR = \left.\frac{dv_{amp}}{dt}\right|_{t=0} \tag{5}$$

$$v_{\text{int}} = v_{amp} \tag{6}$$

$$v_o = \begin{cases} v_{dd} - v_{dp} - r_o i_{ro}, & v_{amp} \geq v_{dd} - v_{dp} \\ v_{\text{int}} - r_o i_{ro}, & v_{dd} - v_{dp} < v_{amp} < v_{ss} + v_{sn} \\ v_{ss} + v_{sn} - r_o i_{ro}, & v_{amp} \leq v_{ss} + v_{sn} \end{cases} \tag{7}$$

where I_s is the DC supply current, $GAIN_i = 1$, 2, 4, 8 and 16 is the voltage gains ($i = \overline{1,4}$) and ω_{pi} is the $-3dB$ radian frequencies.

Short-circuit current limiting is simulated with model parameters $i_{\max\_p}$ and $i_{\max\_n}$. In short-circuit mode the output current i_{ro} is limited to one of the values $i_{\max\_p}$ or $i_{\max\_n}$.

Fig. 3 shows the behavioural VHDL-AMS model of PGA. The library clause and the use clause make all declarations in the packages `electrical_systems`, `math_real` and `std_logic_1164` visible in the model. This is necessary, because the model uses nature `electrical` from package `electrical_system` and constant `math_2_pi` for the value of π from package `math_real`. The package `std_logic_1164` declares the signals `a0`, `a1`, `a2`, `B`, `CS_N` and `CLK_N` by `std_logic` type. The proposed PGA model is composed by an *entity* and an *architecture*, where bold text indicates reserved words and upper-case text indicates predefined concepts. The *entity* declares the generic model parameters, as well as specifies interface terminals of nature `electrical` and electrical ports of `std_logic` type. The generic parameters and constants, used in the simultaneous statements, are not given with their concrete numerical values in the model description. The proposed PGA model includes the following electrical terminals: input port – `input`, output port – `output`, port for the positive supply voltage – `nvdd` and port for the negative supply voltage – `nvss`. The model has one inner terminal `internal`. It's used to specify the controlled source `vint`.

The *architecture* contains the implementation of the model. It is coded by combining structural and behavioural elements.

IV. MODEL PERFORMANCE

The verification of the proposed behavioural PGA model is performed by comparing simulation results with the manufacturer's data for the AD526A, biased with ±15V. The simulations of the model are performed within System Vision 5.5 program (from Mentor Graphics). The test circuits are created following the test conditions, given in the semiconductor data book of the corresponding PGA.

To calculate the frequency characteristics at different voltage gain ($GAIN = 1$, 2, 4, 8 and 16), *ac analyses* are performed within the frequency range from $10Hz$ to $10MHz$. Simulated ac voltage gains as a function of the frequency are shown in

Fig. 4. The comparison gives a very good correspondence between the behavioural of the proposed PGA model and the real amplifier, with a resulting error not higher than 2%.

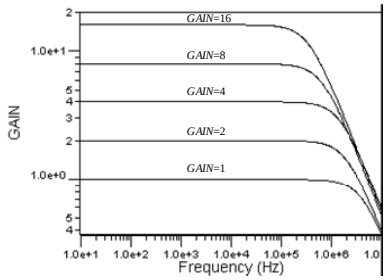


Fig. 4. The simulated frequency responses for the proposed PGA model.

To validate the proposed PGA model, simulation of 12+4-bit Floating-point analog-to-digital converter (FP-ADC) shown in Fig. 5 was carried out for the aforementioned model. This FP-ADC consists of a binary weighting circuit, realized with window comparator, a 12-bit ADC and a five-range PGA with AD526. The simulation results are shown in Fig. 6. As a result, the voltage gain of the PGA AD526 is modulating by the level of the input voltage v_{in}. The AD526, in conjunction with the comparator circuit, scales the input voltage v_o for the 12-bit ADC to a range between half scale and full scale for the maximum usable resolution.

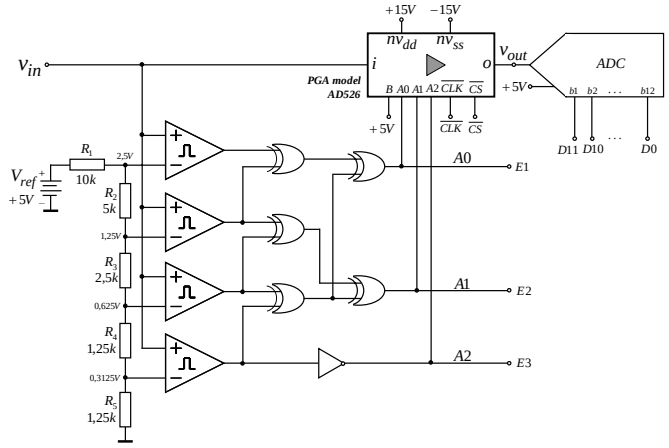


Fig. 5. 12+4-bit Floating-point analog-to-digital converter used in testing the proposed behavioural PGA model.

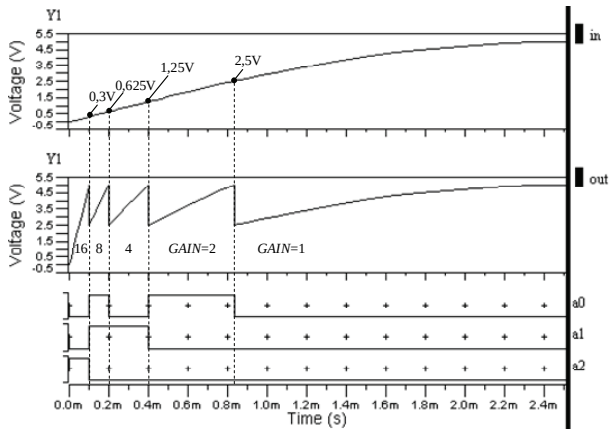


Fig. 6. The simulated response of the circuit shown in Fig. 5 for the model shown in Fig. 3.

The output data of this FP-ADC is presented as a 16-bit word, the lower 12 bits ($D_0 \ldots D_{11}$) from the ADC form the *mantissa* and the upper 4 bits ($E_1 \ldots E_3, B$) from the digital signal used to set the gain of the PGA form the *exponent*. The smallest value of the FP-ADC is $U_{LSB} = (5V / 2^{12}) / 16 \approx 76\mu V \ = 76\mu V$ and the dynamic range is $D_R = 20\lg(5V / 76\mu V) \ = 96{,}3dB$.

V. Conclusion

In this paper a generalized behavioural VHDL-AMS model of monolithic PGA, based on the data sheet characteristics, has been presented. The proposed model accurately describes the behaviour of a most common monolithic PGA with *binary voltage gains*, including electrical parameters, such as input impedance, amplifier gain versus controlling digital code, bandwidth, voltage and current limitations, slew rate, output resistance etc. The workability of the model was proved by comparison of simulation results and data sheet parameters of the monolithic PGA AD526A from Analog Devices. To achieve simplicity of the mathematical equations describing the model, it neglects several second-order effects found in the PGAs, such as the noise, the temperature effects, board parasitics, distortion (harmonic, intermodulation), etc.

One of the aims of the further work is to explore the possibility of modelling second-order effects of the typical PGA, which are taken into consideration for designing mixed-signal applications.

Acknowledgement

This paper is a part of a project which is sponsored by the research program of the Technical University of Sofia, Bulgaria.

References

[1] M. Seifart, *Analoge Schaltungen. 6 Auflage*. Verlag Technik Berlin, 2003 (in German).
[2] W. Jung, W. *Op amp applications*. Analog Devices, MA, 2002.
[3] J. Siegl, *Schaltungstechnik – Analog und gemischt analog/digital. 2. Auflage.* Springer-Verlag, 2005 (in German).
[4] *Spice models*, Analog Devices, 2000.
[5] *Spice models*, Texas Instruments, 2008.
[6] *Spice macromodels*, National Semiconductor, 2010.
[7] I. Pandiev, P. Yakimov, T. Todorov, "Macromodeling of programmable gain amplifiers". E+E, No 7-8, pp. 69-76, 2009.
[8] *AD526 software PGA – data sheet*, Analog Devices, 1999.
[9] S. Robinson, "Conceptual modeling for simulation: Issues and research requirements", Proceedings of the 2006 Winter Simulation Conference, pp. 792-799, 2006.
[10] S. Robinson, *Successful Simulation. A Practical approach to Simulation Projects.* McGraw-Hill, 1998.
[11] E. Christen, K. Bakalar, "VHDL-AMS – A hardware description language for analog and mixed-signal applications, IEEE Trans. on cir. and syst. – II, vol. 46 (10), pp. 1263-1272, 1999.
[12] *Definition of analog and mixed signal extensions to IEEE standard VHDL*, IEEE Standard 1076.1, 1999.

POSTER SESSION PO XII

PO XII - Industrial Electronics

Adjustment and Diagnostics of the System for Gas in Motor Vehicles

Milan D. Milovanović[1] and Saša J. Jovanović[2]

Abstract **– The possibilities in programming and diagnostics of the performance of the system for gas injection, applied in vehicle with bi-fuel system, are presented in the paper. Performed investigations were aimed at indicating the advantages of particular systems. The possibility for applying thermal camera in diagnostics of the system for gas performance was investigated.**

Keywords **– Gas, programming, diagnostics.**

I. INTRODUCTION

All the vehicles which use liquid fuels can be adapted to application of gas, in bi-fuel systems. Application of bi-fuel systems on petrol engines is a long-term decision. The main property of these systems is that the engine is activated either by petrol or gas, according to attuned device, driver's preference or available fuel.

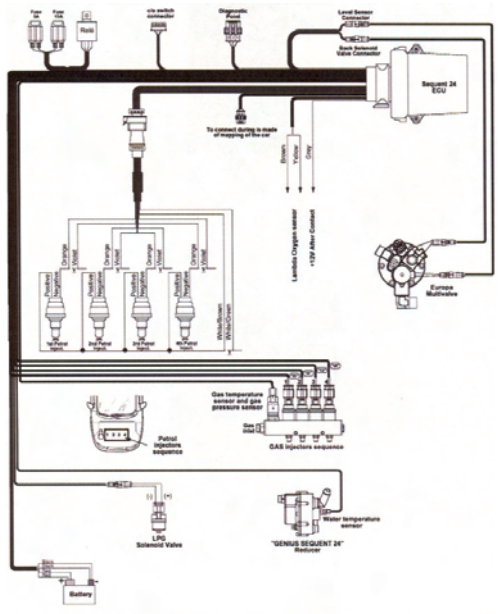


Fig. 1. Electric scheme of system for gas

[1]Milan D. Milovanović, Zastava Cars- IRC Department for Cars Development, Kosovska 4, 34000 Kragujevac, Serbia, E-mail: siljakg@sbb.rs.

[2]Saša J. Jovanović, Zastava Cars- IRC Department for Cars Development, Kosovska 4, 34000 Kragujevac, Serbia E-mail: piter@ia.kg.ac.rs

Fig.1 shows the scheme of one system for gas injection which consists of the following important elements: electronic control unit, selector of fuel type, gas filter, gas pressure regulator, unit for gas injection (injectors) and sensor of gas pressure/temperature. Another electronic unit for gas is added to the existing electronic unit of the system for petrol injection and is connected to it. System for gas injection is the system of the new generation of dual systems (petrol – gas). According to the defined procedure, the engine starts functioning with petrol at start and after a certain timing or achievement of given temperature the control unit switches the engine to functioning to gas. The electronic control unit for gas functions on the principle of calculating the gas injection timing, which is applied on gas injectors. Control of engine functioning is given to petrol control unit, while electronic control unit for gas converts such information into suitable control for gas. Some producers' electronic control units for gas calculate the gas injection timing by using specific information, such as: pressure in gas injection unit, gas temperature, cooling liquid temperature, number of engine revolutions, as well as input data of petrol electronic control unit.

II. ADJUSTMENT OF SYSTEM FOR GAS

When programming the electronic control unit of the system for gas injection, auto-calibration, recommended by the system producer, is usually used for initial adjustment of the system. The auto-calibration procedure, as well as the amount of data used thereat, can vary significantly, which is conditioned by: the system price, regulations of the country-installer, importers' demands, demands of the service provider, demands of the market, i.e. certification organization, system possibilities, i.e. device quality etc.

For investigation of the influence of programming of sequential systems for gas without auto-calibration, the system Sequent 24 of company BRC [4] was used, which does not provide the auto-calibration option; instead, adjustment of the system is realised according to the procedure defined by the producer. The initial adjustment includes adjustments at:

- idle
- idle with extra load (light, air-conditioning etc)
- at 3000o/min, with load.

The initial adjustments are realized at a standstill, after installing the system in the vehicle, i.e. after installing the system on engine in given case, with the adjustment results shown in [3]. The following interventions on the investigated system are possible in the case of system programming optimization:

- at idle
- with load.

The system for gas Sequent 24 of company BRC does not allow other options in the adjustment procedure, see Fig. 2[3]. That fact can be facilitating for a less demanding installer, but quite limiting in developmental system optimization. The results of realized investigations showed that there are not many possibilities in the optimization process of such a system. Consequently, the price of such a system is considerably higher than the price of sequential systems of other producers.

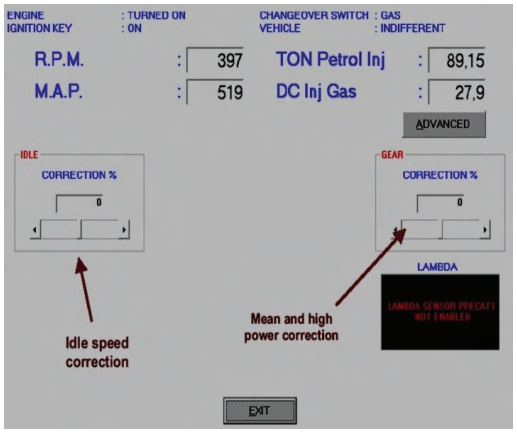


Fig. 2. Allowed adjustment of the system of company BRC

In systems with auto-calibration the initial adjustments are realized at a standstill, too, after installing the system for gas in vehicle. In general, the adjustment of system configuration is realized when the engine is not working. After that the auto-calibration procedure is started, where adjustment is performed according to the demand of the system producer. After the completion of auto-calibration procedure, each system will define its system map. The map is constructed on the basis of two pieces of data: engine rpm and petrol injection timing.

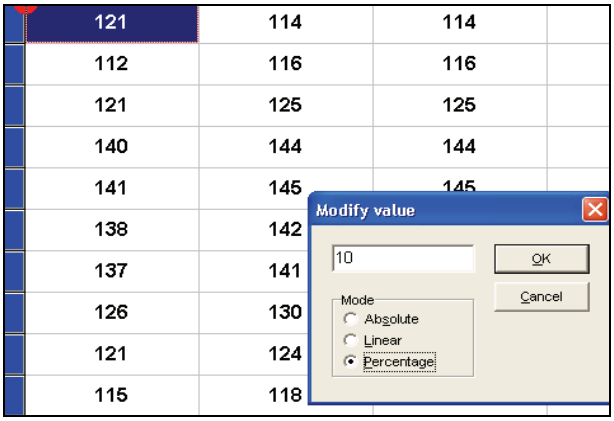


Fig. 3. System optimization

In case that you are not satisfied with auto-calibration, possible modifications can be realized by optimization of map value on gas, i.e. map coefficients loaded during automatic calibration process, both at minimum and beyond idle stroke range. In order to modify the values, you simply mark one or several fields, press ENTER on the keyboard, after which three options for correction will appear: absolute, linear and in percents (see Fig. 3). After selecting "**Switching from PC**", the new icon - "**Switch**" - will appear on the right side, which enables direct control of switch for gas, in the investigator's cabin, via computer. When testing the vehicle on the road, simply click on Switch in case that you wish to switch from petrol to gas or from gas to petrol. When you complete the testing, click on "**End switching from PC**", see Fig. 4.

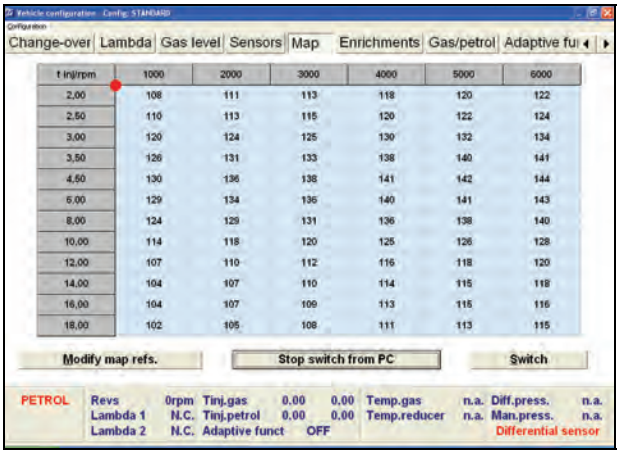


Fig. 4. Optimization – at a standstill or on the road

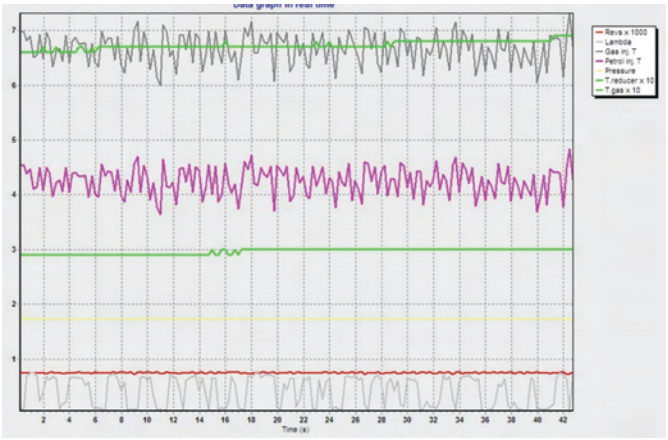


Fig. 5. Results of adjustments at idle stroke

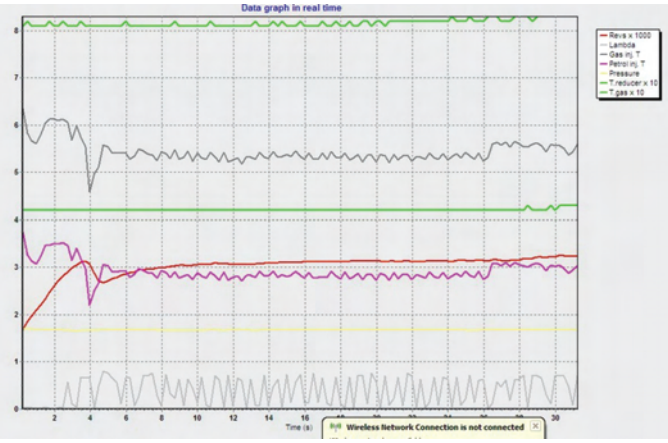


Fig. 6 Results of adjustments at 3000o/min

III. ADJUSTMENT RESULTS

When programming the ECU, in systems with auto-calibration, the auto-calibration procedure, as well as the amount of data used by ECU, varies considerably. Fig. 5 and 6 show the results of adjustments at a standstill. The effects of adjustment can be monitored on the computer for gas. Concurrent petrol injection timing and gas injection timing were obtained. In addition to the analysis of effects of adjustments via computer, attention must be given to the functioning of engine in these conditions, whose running must be constant and uninterrupted. After that, the correction effects are checked on the track in the following conditions:

- vehicle acceleration (I-IV gear)
- vehicle elasticity.

The results of vehicle behaviour check are shown in Fig. 7 and 8. In the course of testing, there must be no interruptions of engine running.

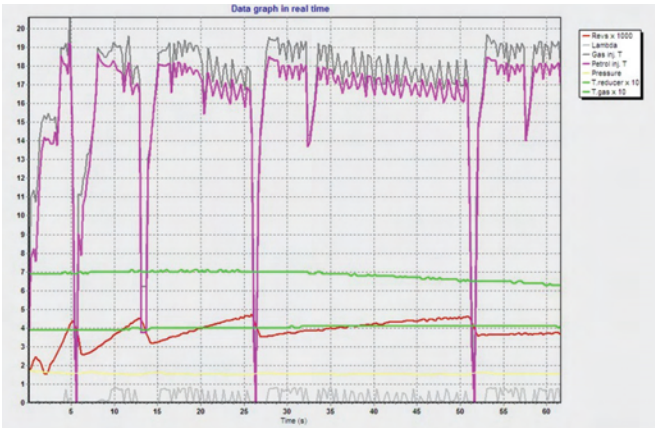


Fig. 7 Vehicle acceleration

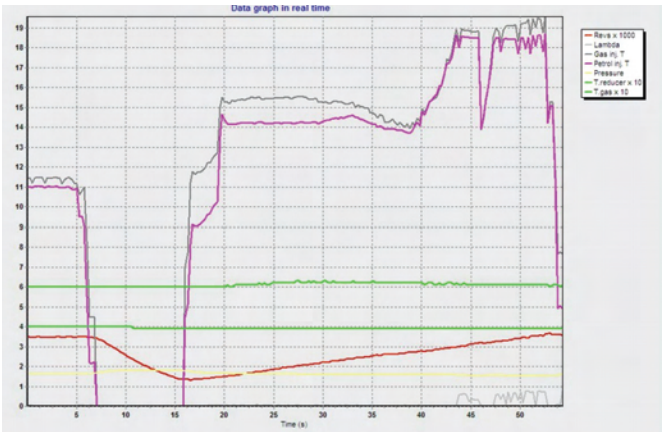


Fig. 8 Vehicle elasticity

Additional adjustments require a lot of time even for the experienced adjuster. In serial installing of the system for gas it is not enough to monitor the effects of adjustments only on the computer for gas. Fig. 9 shows the results of measured power on wheel, petrol-liquid petroleum gas comparatively, in one well adjusted system. Small degradation of vehicle power is obtained when applying gas as fuel. Fig. 10 shows the results of comparative petrol-LPG acceleration measuring. Also, insignificant deviation of vehicle properties was obtained at gas application. Adjustment effects are different for different systems.

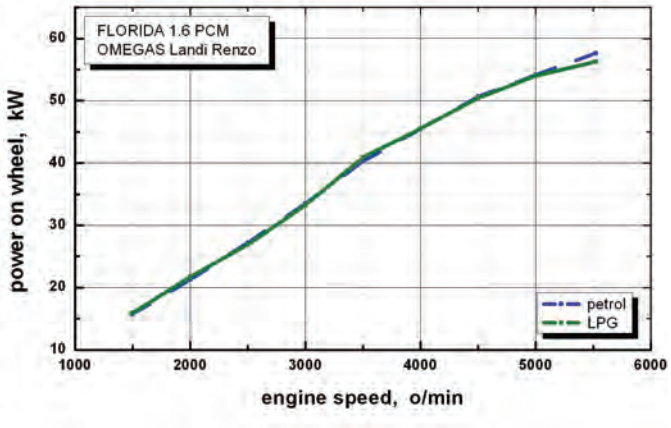


Fig. 9 Comparative presentation of power measuring results

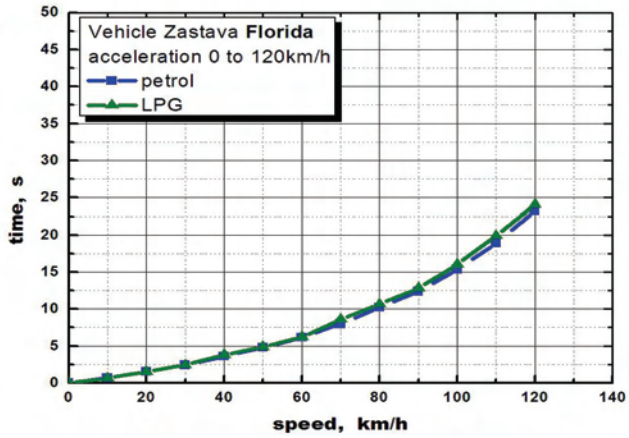


Fig. 10 Comparative presentation of vehicle acceleration measuring

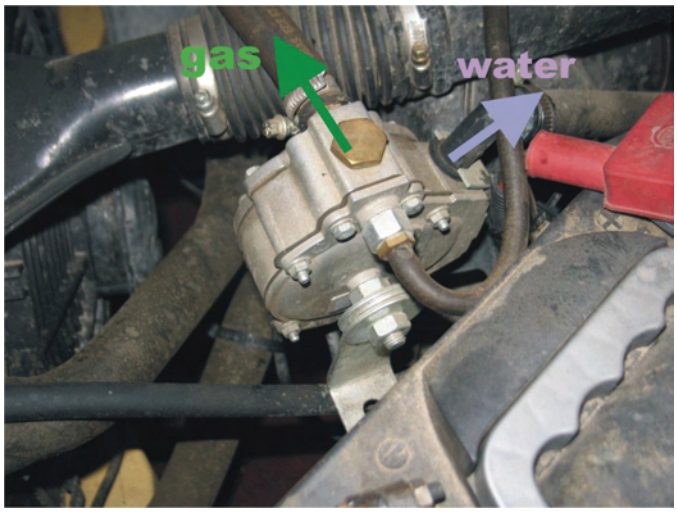


Fig. 11 Position of the pressure reducer in engine compartment

IV. APPLICATION OF THERMAL CAMERA IN DIAGNOSTICS

In most sequential systems, on of the parameters which can be monitored on the computer is gas temperature, see Fig. 6 and 7. Gas temperature is highly influenced by the position of pressure reducer, both in relation to other elements of the system for gas and the drive unit in engine compartment. In the designing process, the rules for installing the pressure reducer must be obeyed (regarding height, width ...). The position of one pressure reducer is shown in Fig. 11. On account of need for quick diagnostics and monitoring of gas behaviour in pressure reducer, i.e. monitoring of gas temperature change, the paper presents one of the possibilities in application of thermal camera in diagnostics, see Fig. 12.

Fig. 12 Application of thermal camera in diagnostics of system for gas

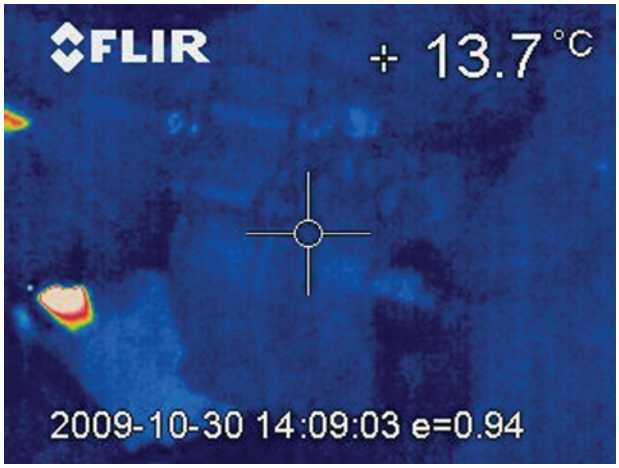


Fig. 13 Cold start

In case that the thermal camera is applied in diagnostics of pressure reducer functioning, the engine compartment must be open, similarly to the conditions of adjusting the system for gas at a standstill. Thermal camera can only be used when the vehicle is stationary, in conditions of idle stroke or at 3000o/min. The application can be customized to the system adjustment conditions, as it is the case in most private service-stations which install system for gas. In addition to that, thermal camera can be used for analyzing heat load of each element in engine compartment.

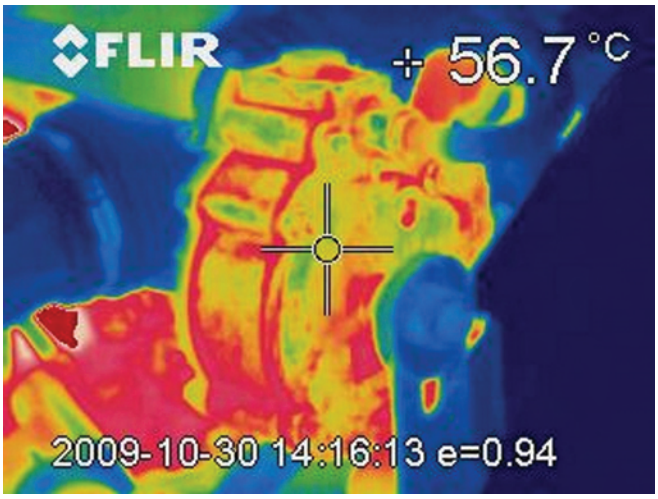


Fig. 14 Fifth measuring

In addition to pressure reducer heating, we can monitor the heating of other elements in the surroundings comparatively. After completion of this measuring, we can check the heat load of all other vehicle elements.

V. CONCLUSION

The results of investigating the effect of system for gas injection programming are presented in this paper. Additional adjustments of the system require a lot of time and adapting of gas ECU to petrol ECU. In the course of that attention must be paid to the speed of electronic control units adaptation to functioning on petrol and gas. Thermal camera can be used for fast analyses of system elements behaviour under heat. Data obtained in this way are useful and specific.

REFERENCES

[1] M. Milovanović, S. Petrović, N. Vitošević and S. Spasojević, ″APPLICATION OF GAS AS ALTERNATIVE FUEL″″, Journal, ECO VARNA 2006, pgs. 9-15, Varna, 2006.

[2] M. Milovanović, A. Dimitrov, N. Vitošević and Š. Uremović ″EXPERIENCE IN APPLICATION OF LIQUID PETROLEUM GAS ON VEHICLES″, Journal ECO VARNA 2007, pgs. 9-15, Varna, 2007.

[3] Producer's manual ″MANUAL ON SEQUENT FAST″.

[4] M. Milovanović, D. Begović and S. Spasojević "INFLUENCE OF PROGRAMMING OF SYSTEM FOR GAS INJECTION ON ENGINE RUNNING", YU INFO 2008, CD, Kopaonik, March 2008.

Analysis of LCC resonant DC-DC converter with capacity filter in the load circuit

Aleksandar S. Vuchev[1], Nikolay D. Bankov[2] and Georgi P. Terziyski[3]

***Abstract* – A study of LCC resonant DC-DC converter with capacity filter in the load circuit is accomplished. To this end, a harmonic analysis is used. Expressions for determination of main quantities of the converter are obtained. The areas of commutation of the power switches are determined at both zero current and zero voltage.**

***Keywords* – LCC resonant DC-DC converter.**

I. INTRODUCTION

Resonant converters are widely used in the building of power supply devices. It is largely due to their peculiar capacity for soft commutation of the power devices. Converters with second order resonant circuits are the most widespread among them and their resonant circuit comprises only an inductor and a capacitor. When their operating frequency is lower than the resonant one, the controllable switches commutate at zero current, i.e., Zero Current Switching (*ZCS*) occurs. Conversely, when the operating frequency is higher than the resonant one, this results in Zero Voltage Switching (*ZVS*). For obtaining better power features it is preferred to have the operating frequency higher than the resonant one.

Despite the great number of advantages, the converters with second order resonant circuits also reveal a considerable drawback – their disability to work in the whole range from a no-load to a short circuit state retaining at the same time the conditions of soft commutation of the controllable switches. This limitation can be overcome by the use of a higher order resonant circuit. Very good results are obtained if just one more reactive element – an inductor or a capacitor – is added to the circuit [1].

LCC converters have been lately used in building arc welders [2], power supplies of fluorescent lamps [3], lasers, etc. In this case, the load circuit behaves like a rectifier with a capacitive filter with respect to the resonant inverter.

Very often in theoretical studies of resonant converters harmonic analysis is applied. Besides, to obtain results with reasonable accuracy, the influence of the first harmonics of the currents and voltages are only taken into account [4] – i.e. the "first harmonic analysis" method is typically used.

The purpose of this paper is studying an LCC resonant DC - DC converter with a capacitive filter in the load circuit by the first harmonic analysis. As a result of the analysis dependencies of basic quantities are to be obtained by which the limits of the converter efficiency while retaining soft commutation conditions of the controllable switches are to be defined.

II. ANALYSIS OF THE CONVERTER

A. General Assumptions

The circuit diagram of the converter under consideration is shown in Figure 1. It comprises an inverter (controllable switches $S_1 \div S_4$ with reverse diodes $D_1 \div D_4$), a resonant circuit

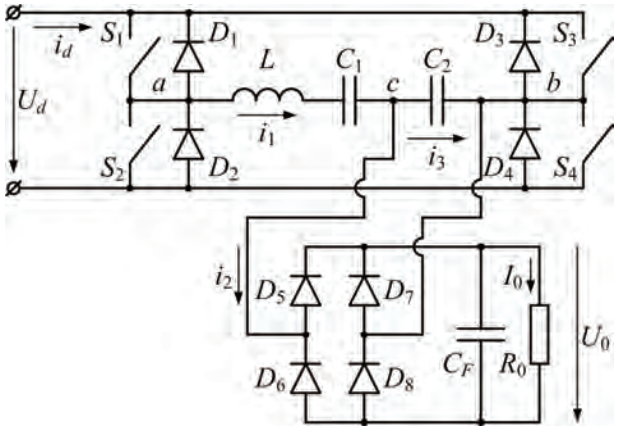


Fig.1. LCC resonant DC-DC converter with capacity load filter

(L, C_1 and C_2), an uncontrollable *DC-DC* converter ($D_5 \div D_8$), a capacitive filter (C_F), and a load resistor (R_0).

For the purposes of the analysis it is assumed that all elements in the circuit diagram are ideal (no losses in them), the power devices switch instantly and the pulsations of the power supply voltage U_d and the output voltage U_0 are negligible. It is also assumed that, according to the chosen analytical method, the first harmonics of the currents and voltages only act in the circuit under consideration.

In converters with third order resonant circuits two resonant frequencies are observed [5]. For the considered converter, they are defined as follows:

$$\omega_0 = 1/\sqrt{LC_1} \qquad \omega_{01} = 1/\sqrt{LC_1C_2/(C_1+C_2)}\,. \qquad (1)$$

[1]Alexandar S. Vuchev is with the Faculty of Electrical Engineering and Electronic, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: avuchev@yahoo.com

[2]Nikolay D. Bankov is with the Faculty of Electrical Engineering and Electronic, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: nikolay_bankov@yahoo.com

[3]Georgi P. Terziyski is with the Faculty of Electrical Engineering and Electronic, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: g_terziyski@yahoo.com

The lower resonant frequency (ω_0 here) is chosen to be the basic one.

In this study it is assumed that the operating frequency of the converter is higher than the basic resonant frequency ω_0.

In accordance with the above, by analogy with the analyses of converters wit second ordre resonant circuit, wave impedance and frequency distraction are determined:

$$\rho_0 = \sqrt{\frac{L}{C_1}} \qquad \nu = \frac{\omega_S}{\omega_0}, \tag{2}$$

where the operating frequency is denoted by ω_S.

Using (2), for the impedances of the inductor and the capacitors of the resonant circuit it is obtained:

$$Z_L = j\nu\rho_0 \qquad Z_{C1} = -j\frac{\rho_0}{\nu} \qquad Z_{C2} = \frac{Z_{C1}}{a}, \tag{3}$$

where $a = C_2 / C_1$ is the ratio between the capacitances of the two capacitors.

B. Output Characteristic

In accordance with the assumptions made, the operation of the converter in the steady-state could be illustrated by the vector diagrams shown in Fig.2. The voltage vectors are shown by a continuous thin line while the current vectors are presented by a continuous thick line. By $\dot{U}_{ab}$ the vector of the first harmonic of the inverter output voltage u_{ab} (between points *a* and *b*) is denoted, by $\dot{U}_{C2}$ - the voltage u_{C2} over the capacitor C_2, and by $\dot{U}_{LC1}$ - the voltage u_{LC1} over the inductor *L* and capacitor C_1 (between points *a* and *c*). By $\dot{I}_1$, $\dot{I}_2$ and $\dot{I}_3$ the vectors of the currents i_1, i_2, and i_3 are denoted, as shown in Figure 1. Angle φ indicates the phase shift of the current i_1 with respect to the voltage u_{ab}, and angle α – the phase shift between the voltages u_{C2} and u_{ab}.

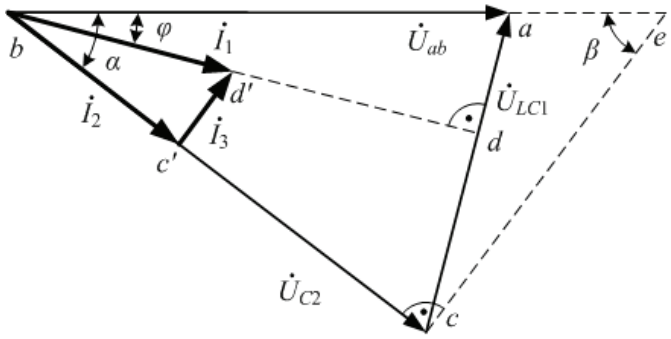


Fig.2. Vector diagrams of the voltages and currents

To better visualize the analysis results in Fig.2 additional constructions are made (shown by means of a dotted line).

Using vector diagrams a system of equations describing the processes in the resonant circuit can be worked out:

$$\begin{cases} \dot{I}_1 = \dot{I}_2 + \dot{I}_3 \\ \dot{U}_{ab} = \dot{U}_{C2} + (Z_L + Z_{C1})\dot{I}_1 \\ \dot{U}_2 = Z_{C2}\dot{I}_3 \end{cases}. \tag{4}$$

After solution of the system an equation is obtained for the output circuit of the resonant inverter:

$$\dot{U}_{ab} = \frac{\dot{U}_{C2}(Z_L + Z_{C1} + Z_{C2}) + (Z_L Z_{C2} + Z_{C1} Z_{C2})\dot{I}_2}{Z_{C2}}. \tag{5}$$

The operation of the power switches ($S_1 \div S_4$) and their antiparallel diodes ($D_1 \div D_4$) results in the rectangular shape of the voltage u_{ab}, its amplitude being equal to the supply voltage U_d. When the rectifier operates in a continuous current mode, the voltage u_{C2} also has a rectangular shape and its amplitude is equal to the output voltage U_0. Then the modules of the vectors of the first harmonics of u_{ab}, u_{C2} and i_2 are defined as:

$$U_{ab} = \frac{\sqrt{8}}{\pi}U_d \qquad U_{C2} = \frac{\sqrt{8}}{\pi}U_0 \qquad I_2 = \frac{\pi}{\sqrt{8}}I_0. \tag{6}$$

Substitution of (3) and (6) in (5) and certain formal transformations lead to:

$$\frac{8}{\pi^2}U_d^2 = \left[\frac{\sqrt{8}(1+a-a\nu^2)}{\pi}\right]^2 U_0^2 + \left[\frac{\pi(1-\nu^2)}{\sqrt{8}\nu}\rho_0\right]^2 I_0^2. \tag{7}$$

For obtaining summarized results all quantities are normalized as follows: voltages with respect to U_d; and currents with respect to U_d/ρ_0. Thus, from equation (7) the expression of the normalized output characteristic of the converter is obtained:

$$1 = (1+a-a\nu^2)^2 U_0'^2 + \frac{\pi^4(1-\nu^2)^2}{64\nu^2} I_0'^2. \tag{8}$$

The above expression is an equation of an ellipse with semi-axes defined by the voltage of a no-load and the current of a short circuit:

$$U'_{0\max} = \frac{1}{1+a-a\nu^2} \qquad I'_{0\max} = \frac{8\nu}{\pi^2(\nu^2-1)}. \tag{9}$$

Equation (8) shows that under certain conditions it is possible for the converter to behave like an ideal source of current. For this purpose it is necessary for the frequency distraction to obtain a limiting value:

$$\nu_{bound} = \sqrt{\frac{a+1}{a}}. \tag{10}$$

Then the output current is determined by the expression:

$$I'_{0bound} = \frac{8}{\pi^2}\sqrt{a(a+1)}. \tag{11}$$

The above equation shows that this value of the current depends only on the parameter *a*.

From (8), the dependence of the output voltage on the output current in relative units is easily obtained:

$$U_0' = \sqrt{\frac{\nu^2 - \left[\left(\pi^2/8\right)\left(\nu^2-1\right)\right]^2 I_0'^2}{\left(\nu + a\nu - a\nu^3\right)^2}} \,. \qquad (12)$$

In accordance with the above equation, a family of output characteristics obtained at $a = 1$ and at different values of the frequency detuning $\nu \geq 1$ is presented in Fig.3. With the exception of the boundary characteristic obtained at $\nu = \nu_{bound}$, all the other characteristics are arcs from ellipses. It can be seen that the characteristics obtained at $\nu > \nu_{bound}$ are concentric to each other and for them the no-load voltage and the short circuit current decrease with the increase of the operating frequency. The characteristics obtained at $\nu < \nu_{bound}$ intersect. In this case with the increase of the operating frequency the short circuit current decreases, but the no-load voltage increases.

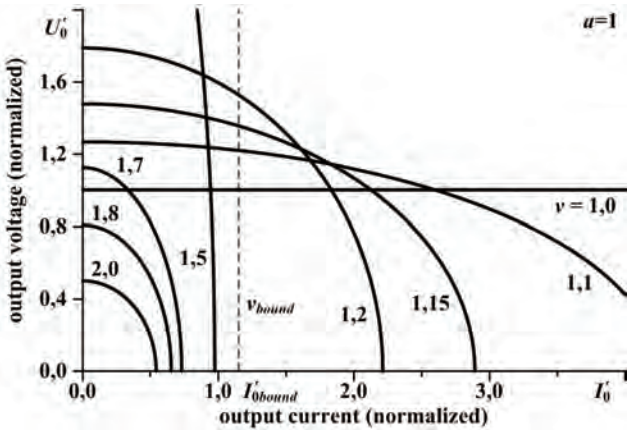


Fig.3. Output characteristics of the converter

Fig.3 shows that in the plan of the output characteristics an operating area in which $U'_0 > 1$ exist. Therefore, under certain conditions the output voltage is higher than the supplied one. This effect is observed for all characteristics obtained at $1 < \nu \leq \nu_{bound}$. The output voltage can be higher than the supplied one at $\nu > \nu_{bound}$.

The normalized dependence of the output power alteration on the output current is obtained from equation (12):

$$P_0' = I_0' U_0' = I_0' \sqrt{\frac{\nu^2 - \left[\left(\pi^2/8\right)\left(\nu^2-1\right)\right]^2 I_0'^2}{(\nu + a\nu - a\nu^3)^2}} \,. \qquad (13)$$

The study of the above expression shows the maximum value of the power:

$$P'_{0\max} = \sqrt{\frac{16\nu^2}{\pi^4\left(\nu^2-1\right)^2\left(1+a-a\nu^2\right)^2}} \,, \qquad (14)$$

which could be obtained when:

$$\begin{aligned} U'_{0(P\max)} &= 1/\left[\sqrt{2}(1+a-a\nu^2)\right] \\ I'_{0(P\max)} &= \sqrt{32}\nu/\left[\pi^2(\nu^2-1)\right] \end{aligned} \,. \qquad (15)$$

A family of dependencies of the output power on the output current obtained when a=1 and with different values of the frequency distraction $\nu \geq 1$ is presented on figure 4.

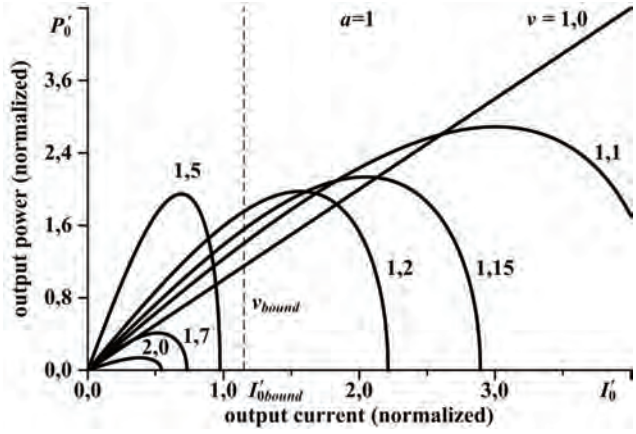


Fig.4. Dependencies of output power P_0 on the output current I_0

It is obvious that the characteristics have a clearly expressed maximum, the position of which shifts towards the beginning of the coordinates with the increase of the frequency distraction.

C. Conditions of soft commutation

As mentioned earlier, an important feature of the resonant converter operation is the mechanism of commutation of the inverter power switches. This is related, on the one hand, to the way of controlling the power switches, and on the other hand, to the type of auxiliary commutation circuits. The required commutation mechanism can be determined based on the phase shift φ between the inverter voltage u_{ab} and the inverter current (in this case i_1). When the voltage is ahead of the current the controllable switches commutate at zero voltage, i.e., zero voltage switching (*ZVS*) occurs. Otherwise, zero current switching takes place (*ZCS*). The angle φ can be determined based on the vector diagrams in Fig.2. Since the current i_1 and the voltage u_{LC1} are shifted with respect to each other at an angle of 90°, the segment *bd* is a height in the triangle *abc*. Its length is easily expressed by the lengths of the sides of Δabc. Then the same angle φ can be expressed by the lengths of the sides in the rectangular triangle *abd*. Before that, however, the length of the side ac has to be found (voltage u_{LC1}). This is possible by applying the cosine theorem. To find the phase shift α between the vectors of the voltages u_{ab} and u_{C2} equation (5) is used. Substitution of (3) and (6) leads to:

$$\alpha = arctg\left[\frac{\pi^2}{8}\frac{I_0'(1-\nu^2)}{U_0'(-a\nu^3+a\nu+\nu)}\right]. \qquad (16)$$

Then the final expression for the angle φ has the form:

$$\varphi = arctg\left(\frac{8\nu^2 a + \pi^2 \nu I_0' \sin\alpha}{\pi^2 \nu I_0' \cos\alpha}\right). \qquad (17)$$

Fig.5 presents dependencies of the angle φ on the output current I_0, obtained when a=1 for different values of the frequency distraction $\nu \geq 1$. It can be seen that when $\nu > \nu_{bound}$ the angle φ always has a positive value. Consequently, the inverter voltage u_{ab} is ahead of the inverter current i_1 and the controllable switches commute at zero voltage. When $\nu < \nu_{bound}$, with the increase in the output current I_0 from 0 to I_{0max} the angle φ changes from –90° до +90°. It means that whereas at heavy loads the controllable switches commute at zero voltage, at small loads switching is accomplished at zero current.

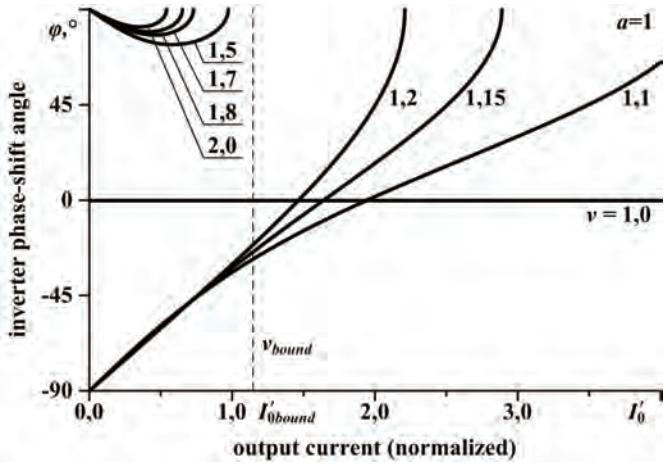


Fig.5. Dependencies of phase-shift angle φ on the output current I_0

During the process of operation while retaining the conditions for soft commutation, the commutation mechanisms of the controllable switches cannot be easily changed. Therefore, if it is necessary to use ZCS, the range of load should be limited to the modes, close to a no-load, and the frequency distraction should not exceed ν_{bound}. When the control devices have been chosen to switch at zero voltage the converter can operate in the whole range from a no-load to a short circuit. To maintain the conditions for soft commutation, however, the frequency distraction in the area of small loads should be higher than ν_{bound}.

III. Conclusion

A study of LCC resonant DC-DC converter with capacitive filter in the load circuit is accomplished by means of harmonic analysis, taking into account the influence of the firs harmonic only. It is assumed in the process of consideration that the converter operates with a higher frequency than the basic resonant frequency of the resonant circuit.

As a result of the analysis an expression for the output characteristic of the converter is obtained. Dependencies of characteristic angles of the phase shift between main quantities for the converter circuit under consideration are determined.

Based on the obtained results the type of the output characteristics of the LCC converters is defined. It is shown that under certain conditions these converters behave like an ideal current source without using any additional resources (e.g., as with converters with second order resonant circuits, where current feedback is used). This mode is possible at a precisely defined frequency distraction that depends only on the values of the capacitors in the resonant circuit. It is established that the output voltage of the converter may have a higher value than the supply voltage. Some of the conditions under which this effect is observed are discussed.

It is established from the dependencies of the angle of phase shift between the voltage and the current of the inverter what the commutation mechanisms of the controllable converter device should be. Although the operating frequency is higher than the basic resonant frequency of the resonant circuit, from small loads the switches are supposed to commutate at zero current, i.e., zero current switching occurs. This effect is observed at frequency distraction lower than that at which the converter behaves like a current source. In the rest of the cases, the controllable devices should switch at zero voltage. The transition from ZVS to ZCS is mainly due to the change of the load and can be avoided by changing the operating frequency.

The study results can be used in designing LCC converters used as power supplies of electric arc welders, fluorescent lamps, lasers, etc.

References

[1] Batarseh I., “Resonant Converter Topologies with Three and Four Energy Storage Elements”, IEEE Transactions on Power Electronics, Vol. 9, No. 1, January 1994 pp. 64-73.

[2] Malesani, L., L. Mattavelli, L. Rossetto, R. Tenti, W. Marin, A. Pollmann, “Electronic Welder With High-Frequency Resonant Inverter”, Industry Applications, IEEE Transactions, Vol. 31, Iss. 2, March/April 1995, pp. 273 – 279.

[3] Alexandrov F.I., “Resonant Inverter Including Feed Back Circuit Having Phase Compensator and Controller”, United States Patent, P.N. 7,045,966 B2, 16 May 2006.

[4] De Simone S., C. Adragna, C. Spini, G. Gatavari, “Design-Oriented Steady State Analysis of LLC Resonant Converters Based on FHA”, IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion, Speedam 2006 pp. S41-16 – S41-23.

[5] Haung G., A.J. Zhang, Y. Gu, “LLC Series Resonant DC-to-DC Converter”, United States Patent, P.N. 6,344,979 B1, 5 February 2002.

Load Characteristics of LCC Resonant DC-DC Converter with Capacity Filter in the Load Circuit

Aleksandar S. Vuchev[1], Nikolay D. Bankov[2] and Georgi P. Terziyski[3]

***Abstract* – A study of LCC resonant DC-DC converter with capacity filter in the load circuit is accomplished. On the basis of an existing harmonic analysis, expressions for determination of the currents through the power devices and the resonant circuit components are obtained. Main load characteristics of the converter are built. Methods for design of the converter are proposed.**

***Keywords* – LCC resonant DC-DC converter.**

I. INTRODUCTION

LCC converters have been used recently for realization of arc welders, for power supply of luminescent lamps, lasers etc. In this case the load circuit behaves similarly to a rectifier with a capacitive filter with respect to the resonant inverter. Analysis of an LCC converter with a capacity filter in the load circuit, carried out following the first harmonic method, is presented in paper [1]. As a result of the theoretical study presented there, an expression of the output characteristic of the converter has been obtained. The dependences of both the output power and the phase shift between the inverter voltage and inverter current on the output current have been determined as well. The required commutation mechanisms of the controllable switches have been defined.

The present work is dedicated to defining the load of the components, constituting the converter circuit. Based on the analysis presented in [1] as well as on the current investigation, a methodology for designing LCC converters with capacitive filter in the load circuit.

II. LOAD CHARACTERISTICS

A. General Assumptions

The circuit diagram of the converter under investigation is shown in Fig. 1. It is made up of: an inverter (controllable switches $S_1 \div S_4$ with reverse diodes $D_1 \div D_4$), a resonant circuit (L, C_1 and C_2), an uncontrollable rectifier ($D_5 \div D_8$), a capacitive filter (C_F), and a load resistor (R_0).

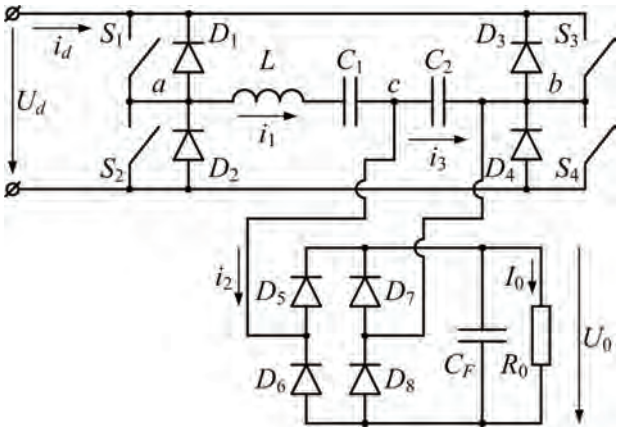


Fig.1. LCC resonant DC-DC converter with capacity load filter

For the sake of the analysis it is assumed in [1] that all the components in the circuit are ideal (there are no losses in them), the power devices switch on instantly, and the ripples of the supplying voltage *Ud* and the output voltage U_0 are negligible. It is also assumed that, according to the chosen method of analysis, only the first harmonics of the currents and voltages are active in the circuit.

For facility's sake the ratio $a = C_2 / C_1$ between the capacitances of the two capacitors in the resonant circuit has been introduced.

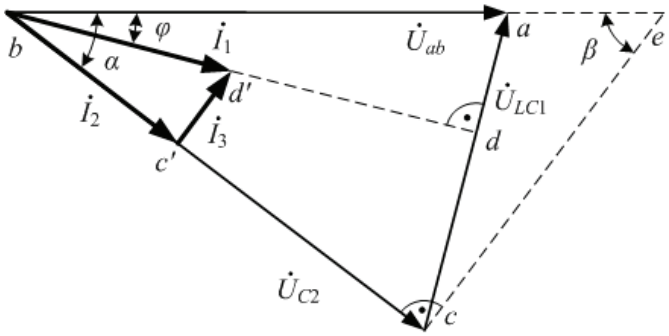


Fig.2. Vector diagrams of the voltages and currents

In correspondence with the assumptions mentioned above the converter operation in a steady-state condition can be illustrated by the vector diagrams shown in Fig.2. The voltage vectors are shown by a continuous thin line and the current vectors are presented by a thick one. The vector of the first harmonic of the voltage u_{ab} at the output of the inverter

[1]Alexandar S. Vuchev is with the Faculty of Electrical Engineering and Electronics, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: avuchev@yahoo.com

[2]Nikolay D. Bankov is with the Faculty of Electrical Engineering and Electronics, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: nikolay_bankov@yahoo.com

[3]Georgi P. Terziyski is with the Faculty of Electrical Engineering and Electronics, 26 Maritza st., 4002 Plovdiv, Bulgaria, E-mail: g_terziyski@yahoo.com

(between the points *a* and *b*) is denoted by $\underline{U}_{ab}$, the voltage u_{C2} on the capacitor C_2 is marked by $\underline{U}_{C2}$, and the voltage u_{LC1} in the inductor *L* and the capacitor C_1 (between the points *a* and *c*) is correspondingly denoted by $\underline{U}_{LC1}$. By $\underline{I}_1$, $\underline{I}_1$ и $\underline{I}_1$ the vectors of the currents i_1, i_2, и i_3, shown in Fig.1 are denoted. Angle φ shows the phase shift of the current i_1 with respect to the voltage u_{ab}, and angle α illustrates the phase shift between the voltages u_{C2} and u_{ab}.

For a better illustration of the results additional lines have been drawn (the dotted lines) in Fig.2.

B. Basic Dependencies

In order to obtain generalized results in [1] all quantities have been normalized as it follows: the voltages in relation to U_d; the currents – with respect to U_d/ρ_0. By ρ_0 the wave impedance of the resonant circuit is denoted and it is defined for the main resonance frequency ω_0:

$$\rho_0 = \sqrt{\frac{L}{C_1}} \qquad \omega_0 = \frac{1}{\sqrt{LC_1}}. \tag{1}$$

The theoretical investigation in [1] is carried out under the condition of frequency distraction $\nu = \omega_S/\omega_0$ being higher than a unity. The operating frequency is denoted by ω_S here. Thus for the normalized value of the output voltage U_0' as a function of the normalized output current I_0', the following expression is obtained:

$$U_0' = \sqrt{\frac{\nu^2 - \frac{\pi^4}{64}\left(\nu^2-1\right)^2 I_0'^2}{\left(\nu + a\nu - a\nu^3\right)^2}}. \tag{2}$$

The equation given above shows that for a certain value of v_{bound}, the converter behaves like an ideal source of current.

In result from the analysis, presented in [1], dependence between the output power of the converter and the output current in relative units has been obtained. It has its maximum, when the output voltage and the output current obtain the following normalized values:

$$U'_{0(P\max)} = \sqrt{\frac{1}{2\left(1+a-a\nu^2\right)^2}}. \tag{3}$$

$$I'_{0(P\max)} = \sqrt{\frac{32\nu^2}{\pi^4\left(\nu^2-1\right)^2}}. \tag{4}$$

The angles α and φ are also presented as functions of the output current:

$$\alpha = arctg\left[\frac{\pi^2}{8}\frac{I_0'\left(1-\nu^2\right)}{U_0'\left(\nu+a\nu-a\nu^3\right)}\right]. \tag{5}$$

$$\varphi = arctg\left(\frac{8\nu^2 a + \pi^2 \nu I_0' \sin\alpha}{\pi^2 \nu I_0' \cos\alpha}\right). \tag{6}$$

B. Load Dependencies

For the purposes of the designing process, it is necessary to define the values of the currents through the components of the converter circuit as well as of the voltages on the capacitors in the resonant circuit. This could be done by means of the vector diagram in Fig.2.

From the analysis presented in [1] it is known that:

$$U'_{C2} = \frac{\sqrt{8}}{\pi}U_0' \qquad I_2' = \frac{\pi}{\sqrt{8}}I_0'. \tag{7}$$

Then, taking into account equations (5) and (6), from the triangle of currents ($\Delta bc'd'$) for the root mean square value of the current in the inductor of the resonant circuit the following expression is obtained:

$$I_1' = \sqrt{\frac{\pi^4 I_0'^2 + 64a^2\nu^2 U_0'^2 + 16\pi^2 a\nu U_0' I_0'}{8\pi^2}}. \tag{8}$$

The average values of the current flowing through the diodes of the converter are defined based on the output current and having in mind that each of the diodes conducts just during a half-cycle:

$$I'_{D\_RECT\ AV} = \frac{I_0'}{2}. \tag{9}$$

Since it is assumed that the operating frequency ω_S is higher than the main resonant frequency ω_0, the action of the controllable switches and their anti-parallel diodes could be illustrated by the time diagrams, shown in Fig.3. They correspond to the vector diagram in Fig. 2.

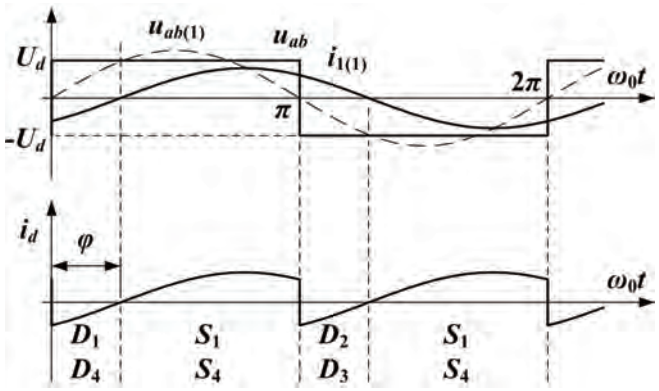


Fig.3. Invertor voltage and current waveforms

By $u_{ab(1)}$ and $i_{1(1)}$ the first harmonics of the inverter voltage u_{ab} and inverter current i_1 are denoted. From the figure it can be seen that for the interval, corresponding to angle φ the inverter voltage and the inverter current have opposite signs. Consequently, during these intervals the reverse diodes of the converter conduct electricity.

During the rest of the time, corresponding to angle π-φ, the controllable switches conduct electricity. Fig. 3 shows the sequences of conduction for all power devices of the inverter.

From what has been exposed above it follows that the normalized average values of the currents through the controllable switches are defined as follows:

$$I'_{S\,AV} = \frac{1}{2\pi}\int_{0}^{\pi-\varphi}\sqrt{2}I_1' \sin\omega t d\omega t = \frac{\sqrt{2}}{2\pi} I_1'(1+\cos\varphi)\,. \quad (10)$$

Likewise, for the currents through the anti-parallel diodes it is obtained:

$$I'_{D\,AV} = \frac{1}{2\pi}\int_{\pi-\varphi}^{\pi}\sqrt{2}I_1' \sin\omega t d\omega t = \frac{\sqrt{2}}{2\pi} I_1'(1-\cos\varphi)\,. \quad (11)$$

Then the normalized average value of the current, consumed by the power supply, is:

$$I'_{d\,AV} = \frac{1}{\pi}\int_{0}^{\pi}\sqrt{2}I_1' \sin(\omega t-\varphi) d\omega t = \frac{2\sqrt{2}}{\pi} I_1' \cos\varphi\,. \quad (12)$$

The maximum value of the voltage across the capacitor C_1 can be defined based on the effective value I_1 of the current through the inductor:

$$U'_{C1\max} = \frac{\sqrt{2}I_1}{\omega_S C_1}\cdot\frac{1}{U_d} = \frac{\sqrt{2}}{\nu}\cdot\frac{I_1}{U_d/\rho_0} = \frac{\sqrt{2}I_1'}{\nu}\,. \quad (13)$$

The maximum value of the voltage on the capacitor C_2 is actually equal to the output voltage U_0.

In accordance with equation (8) Fig.4 presents normalized dependencies of the current through the inductor as a function of the output current, obtained at a=1 and for different values of the frequency distraction $\nu \geq 1$.

Whatever the value of ν is, the characteristics look alike. They have a maximum, whose position moves toward the origin of the coordinate system with the increase in the frequency distraction. The dependencies show that for the whole range of load change the current through the inductor always has a non-zero value. Moreover, at frequency distraction close to the limiting one ν_{bound}, the current I_1 can have a significant value even at low output current. Consequently, considerable losses in the inductor of the resonant circuit will be observed under these modes as well.

On the basis of equation (10) normalized dependencies for the average values of the currents flowing through the controllable inverter switches as a function of the output current have been constructed. They have been obtained at a=1 and at different values of frequency distraction $\nu \geq 1$. The dependencies are shown in Fig.5.

The characteristics show that for the whole range of loads the currents flowing through the controllable switches have non-zero values. This guarantees the conditions of soft commutation. On the other hand, at frequency distraction close to the limiting value ν_{bound}, the currents through the controllable switches could be considerable, thus increasing the losses as well.

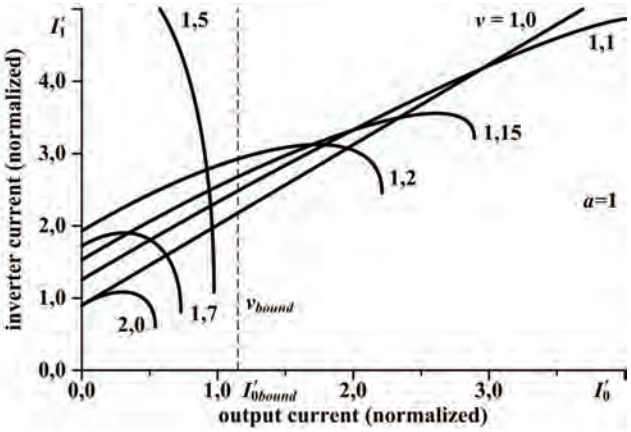


Fig.4. Dependencies of the converter currents on the output current

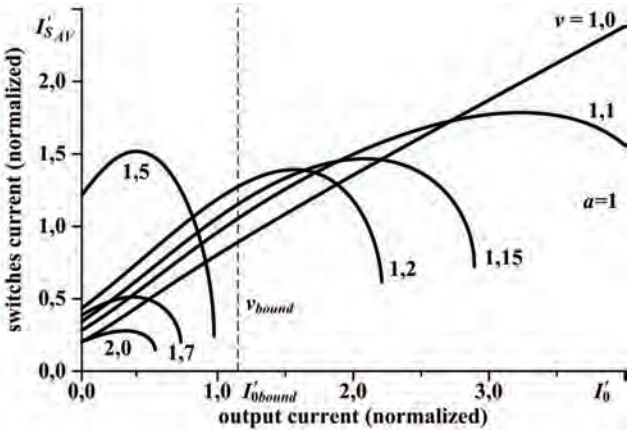


Fig.5. Dependencies of the average values of currents, running through the inverters switches on the output current

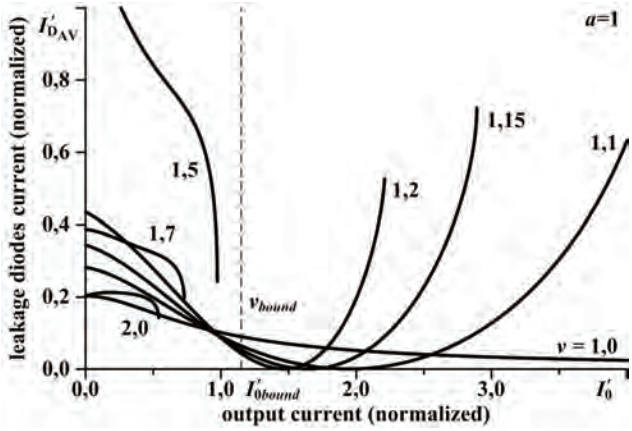


Fig.6. Dependencies of the average values of leakage currents, running through the inverter diodes on the output current

Fig. 6 presents normalized dependencies of the average current values through the reverse inverter diodes as a function of the output current. They have been obtained on the basis of equation (11) at a=1 and at different values of the

frequency distraction $\nu \geq 1$. The characteristics show that the currents through the reverse diodes could also have considerable values at frequency distraction close to ν_{bound}. From the figure it can be seen as well, that for certain values of I_0 и $\nu < \nu_{bound}$ the current flowing through the reverse diodes has a zero value. Comparing these points with the results from [1] it becomes clear that they correspond to the zero values of the phase shift angle φ between the inverter current i_1 and the inverter voltage u_{ab}.

III. Design methods

Converter design is usually accomplished based on predetermined values for the output power P_0, the output voltage $U_{0,}$ and the operating frequency f_S. Both the ratio between the capacitances a and the frequency distraction ν have to be chosen in advance as well. This is achieved by following the recommendations given in [1].

Defining the parameters of the power supply is carried out based on the assumption that there are no losses in the circuit and the converter operates at the point of maximum power. Then,

$$U_d = \frac{U_0}{U'_{0(P\max)}} \qquad I_d = \frac{P_0}{U_d}. \tag{14}$$

From the expressions (3) and (4) for the output voltage and current at the maximum of output power the normalized value of the load resistor can also be defined:

$$R'_{0(P\max)} = \sqrt{\frac{\pi^4 \left(\nu^2 - 1\right)^2}{64\left(\nu + a\nu - a\nu^3\right)^2}}. \tag{15}$$

As it is assumed that the converter operates at the point of maximum output power, then, in accordance with the chosen principle of normalization, it is obtained:

$$R'_{0(P\max)} = \frac{R_0}{\rho_0} = \frac{U_0^2}{P_0\sqrt{L/C_1}}. \tag{16}$$

On the other hand, the frequency distraction can be presented as:

$$\nu = 2\pi f_S \sqrt{LC_1}. \tag{17}$$

Then the values of the components in the resonant circuit L, C_1 and C_2 are defined by the expressions for the normalized value of the load resistor at maximum output power and frequency distraction as it follows:

$$L = \frac{4\nu^2 U_0^2}{\pi^3 f_S P_0}\sqrt{\frac{\left(1 + a - a\nu^2\right)^2}{\left(\nu^2 - 1\right)^2}}, \tag{18}$$

$$C_1 = \frac{\pi P_0}{16 f_S U_0^2}\sqrt{\frac{\left(\nu^2 - 1\right)^2}{\left(1 + a - a\nu^2\right)^2}}, \tag{19}$$

$$C_2 = aC_1. \tag{20}$$

Finally, by means of the expressions (8) ÷ (13) both the currents through the inductor and the power devices in the converter circuit and the voltages across the capacitors in the resonant circuit are defined. They will serve for the proper selection of components.

IV. Conclusion

A theoretical study of an LCC resonant DC-DC converter by means of harmonic analysis has been accomplished. The consideration has been carried out with the assumption that the operating frequency is higher than the basic resonant frequency of the converter.

As a result from the analysis expressions defining the values of the current flowing through the components constituting the converter circuit have been obtained. Their graphical dependencies on the output current have been drawn. Their analysis proves the opportunity to operate the converter in soft commutation within the whole range of loads from a no-load to a short circuit state. Apart from that it can be seen that the currents through the components could have considerable values even at small loads and at frequency distraction close to a certain limiting value. It has been established that at this value of frequency distraction, the converter behaves like an ideal source of current.

Expressions defining the voltages on the capacitors in the resonant circuit have been obtained as well.

Based on the results from the analysis a simple methodology for engineering design of LLC resonant DC-DC converters with capacity filters in the load circuit has been proposed.

References

[1] Bankov N., G. Terziyski, A. Vuchev. “Analysis of LCC Resonant DC-DC Converter with Capacity Filter in the Load Circuit”. XLV International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST’2010, Ohrid, Macedonia.

Investigation of Power MOSFET and IGBT Gate Drivers using SPICE

Georgi Tzv. Kunov[1] and Elissaveta D. Gadjeva[2]

Abstract **–The driver circuits for control of the power MOSFET and IGBT transistors are of significant importance in the field of converter device design. A number of companies offer specialized integrated circuits for this purpose. They are successfully used in the cases when the transistors operate in the mode of soft switching (ZVS and ZCS). In the present paper, the influence of the shape of control pulses on the commutation processes of MOSFET and IGBT commutation is considered, in the mode of hard switching. A special attention is paid on a driver circuit limiting the speed of variation of the voltage and the current (du/dt and di/dt), which leads to reducing of the overvoltages on the transistors. The influence on the commutation losses is investigated when applying the proposed approach using the PSpice simulator.**

Keywords **– MOSFET/IGBT Gate Driver, Computer simulation, MOSFET/IGBT Gate Driver Spice model**

I. INTRODUCTION

In the field of power converters realized by MOSFET/IGBT transistors, the commutation processes in transistors are significantly influenced on the type of the used Gate Driver circuits [1,2,3]. A number of companies offer specialized integrated Gate Driver circuits. Among the most widespread are the integrated drivers of the I21xx series of the company International Rectifiers and their analogs MPIC21 of the company Motorola, the series TC4xxx of MICROCHIP/TELCOM, IXDDxxx of IXYS, etc. The listed specialized Gate Driver circuits are intended for commutation of power transistors with differing powers. They do not take into account the influence of the parasitic capacitances (the Miller capacitance and the output capacitance) on the commutation processes. The time interval for charge and discharge of the parasitic capacitances significantly influences on the:

- switching speed (*dv/dt* and *di/dt*) and hence on the EMI/RFI emissions [4,5];
- dynamic power losses;
- overvoltages on the transistors;
- reliable turn-on and turn-off of the transistors during the switching.

Based on *dv/dt* and *di/dt* tracking, a reliable (but too complex) driver circuit is proposed in [6] for power IGBT transistors. It satisfies the listed above conditions. A simple and cheap solution to suppress the Miller capacitance current represents the specialized integrated circuit ACDL331J [7]. The function *two-level turn-off* is added in [8] to the abilities of [7], which allows to decrease *dv/dt* and hence the overvoltages on the transistors.

In the present paper, MOSFET/IGBT driver circuits are investigated using *Cadence PSpice*, which realize the functions *one-level turn-on/turn-off* and *two-level turn-on/ one-level turn-off*. The investigated circuit using two drivers IXDD414 corresponds to the circuit proposed in [5].

[1]Georgi Kunov is with the Faculty of Electronic Engineering and Technologies, Technical University of Sofia, Klimenrt Ohridski Blvd. 8, 1000 Sofia, Bulgaria, E-mail: gkunov@tu-sofia.bg

[2]Elissaveta Gadjeva is with the Faculty of Electronic Engineering and Technologies, Technical University of Sofia, Klimenrt Ohridski Blvd. 8, 1000 Sofia, Bulgaria, E-mail: egadjeva@tu-sofia.bg

II. INVESTIGATION OF POWER MOSFET AND IGBT GATE DRIVERS

A. One-level Turn-on/Turn-off Gate Drivers

The circuit of one-level turn-on/turn-off Gate Drivers is shown in Fig. 1. The presented connection of the power transistors allows to investigate the influence of the Miller capacitance and the output capacitance on the following characteristics of power transistors: peak current value, Gate-Source voltage and peak value of the commutation power. The simulation results are shown in Fig. 2 for V12= V22=0V, i.e. for the unipolar supply voltage of the integrated drivers IXDD414. The simulation results for V12=V22=5V are shown in Fig. 3. It is seen from the comparison of the results in Fig. 2 and Fig. 3 that for the case of bipolar supply voltage the peak value of the capacitance current through the transistors decreases three times and the peak value of the switching power decreases four times. The additional sources V12 and V22 increase the threshold voltage V_{th} of the transistors, thus avoiding the negative influence of the Miller capacitance current on the gate voltage.

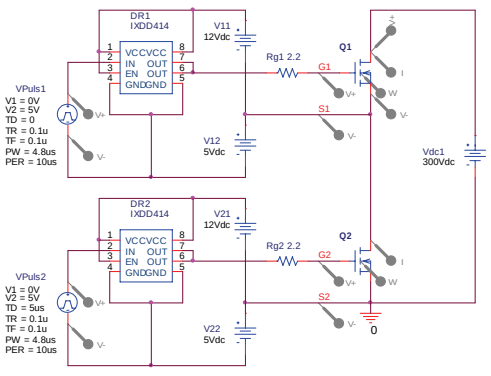


Fig. 1. The circuit of one-level turn-on/turn-off Gate Driver

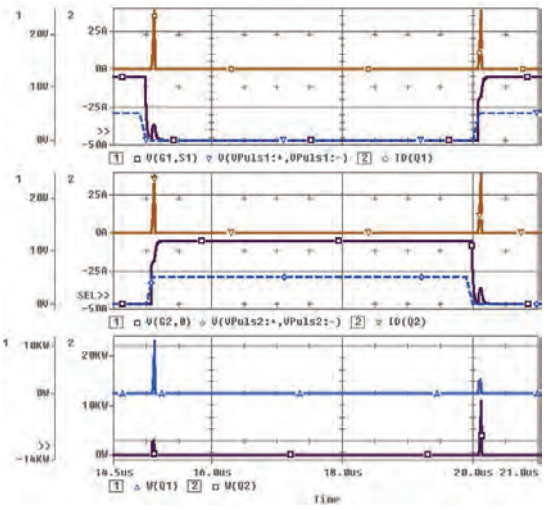


Fig. 2. Simulation results of the circuit in Fig. 1 for V12=V22=0V

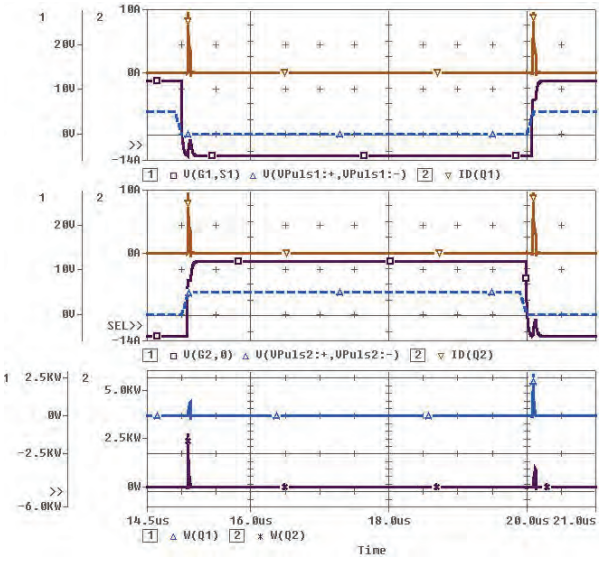


Fig. 3. Simulation results of the circuit in Fig. 1 for V12=V22=5V

B. Two-level Turn-on/One-level Turn-off Gate Driver

The circuit of the two-level turn-on/one-level turn-off Gate Driver is shown in Fig. 4. Each transistor is controlled by two drivers IXDD414.

The switching from "0" in "1" of DR12 and DR22 is delayed with respect to DR11 and DR12. This time delay is defined by the turn-on time t_{don} of the transistors. The moment of turn-on is tracked by the diodes D1, D2 and comparators V1A and V2A. During the time interval t_{don} the resistor dividers R1,R2 and R3,R4 limit the gate voltage at the level below the nominal level. It is seen from the simulation results shown in Fig. 5.

From the comparison of the obtained results in Fig. 5 and the results for one-level turn-on (Fig. 3) it can be seen that the peak value of the current through the transistors and the switching power decrease two times.

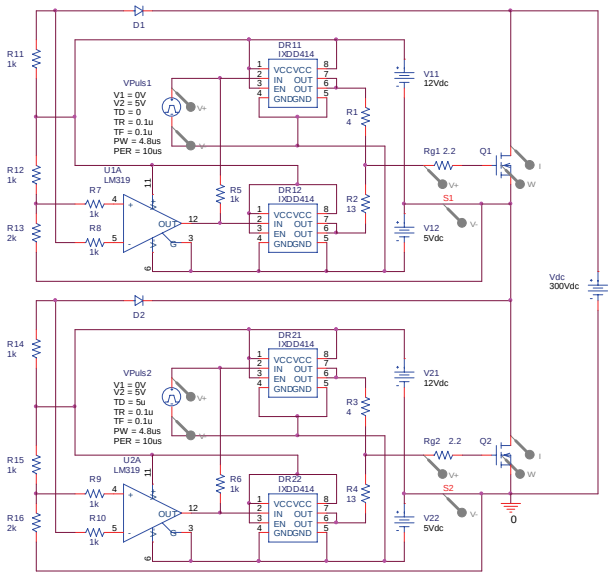


Fig. 4. The circuit of the two-level turn-on/one-level turn-off Gate Driver

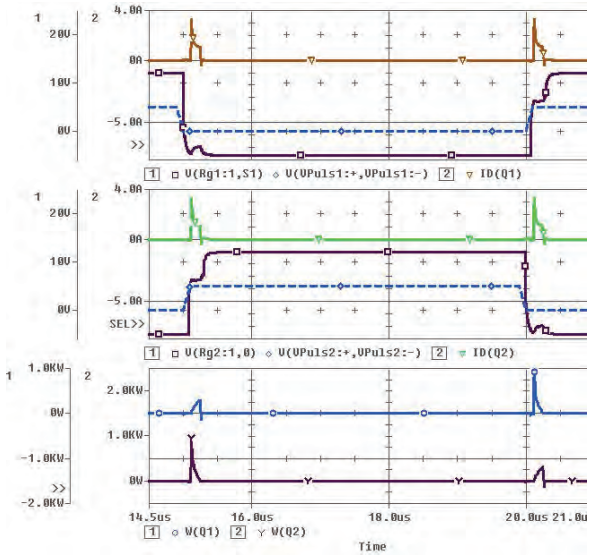


Fig. 5. Simulation results of the circuit in Fig. 4

C. Two-level Turn-on/One-level Turn-off Gate Driver of BUCK Converter

The simulation circuit is shown in Fig. 6. In this case the voltage value of the first turn-on level is defined parametrically. In order to avoid the transients of the load current, the load is represented by a current source I_{load}. The following dependencies are investigated:

- *di/dt* as a function of the threshold voltage V_{th} for a constant current I_{load} (Fig. 7);
- *dv/dt* as a function of the current I_{load} for a constant threshold voltage V_{th} (Fig. 8).

The simulation results for the variation of the collector-emitter voltage are shown in Fig. 9 for different values of the threshold voltage V_{th}.

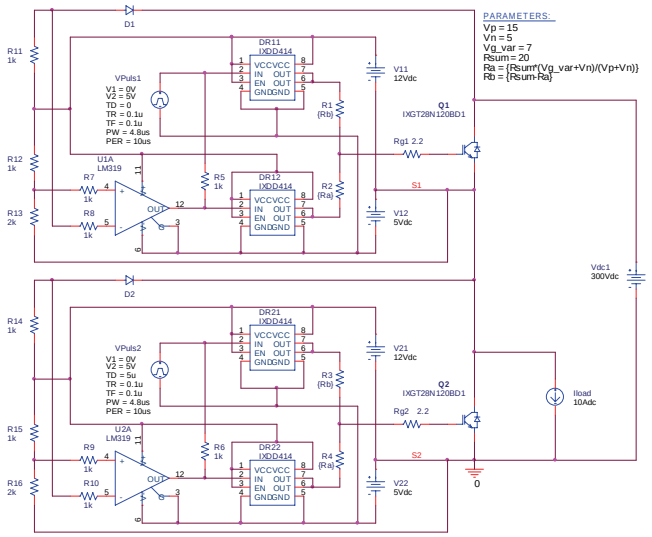


Fig. 6. The circuit of the two-level turn-on/one-level turn-off Gate Driver of BUCK converter

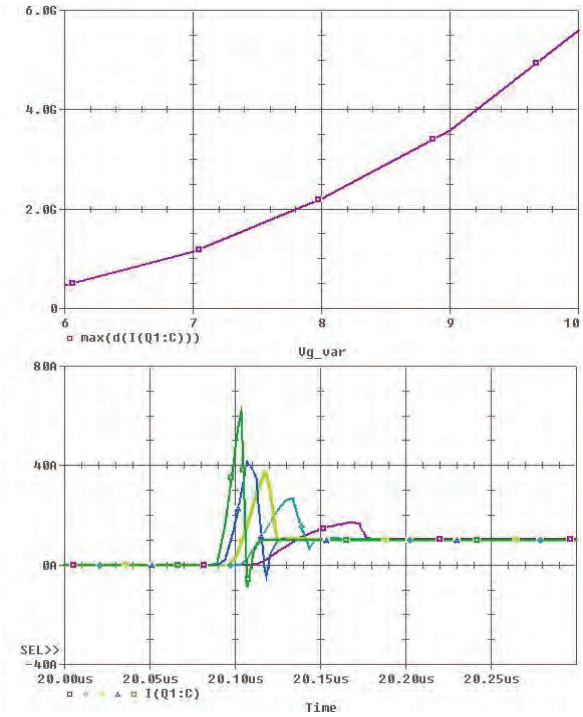


Fig. 7. The dependence of *di/dt* as a function of the threshold voltage for a constant current I_{load}

III. PSPICE MODEL

The *PSpice* model of the MOSFET Gate Driver IXDD414 [9] can be applied in the cases when the ground node DGND is connected to the ground 0 of the circuit. When the node DGND is not connected to the ground 0, this *PSpice* model cannot be used. It is necessary to modify the *PSpice* subcircuit description of the driver IXDD414 in order to avoid this restriction. For this purpose, the node 0, participating in the original model description [9], is replaced by the node GND, corresponding to the pin DGND.

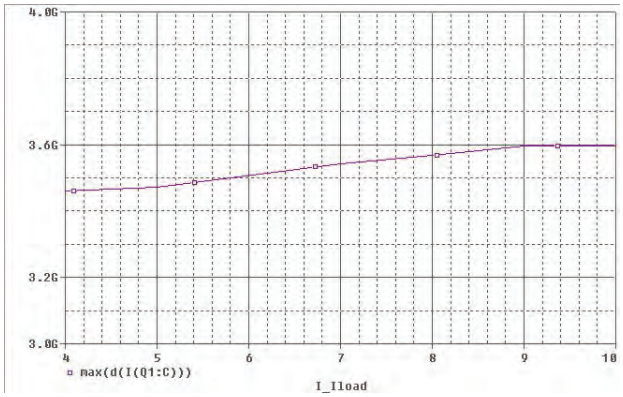


Fig. 8. The dependence of *dv/dt* as a function of the current I_{load} for a constant threshold voltage

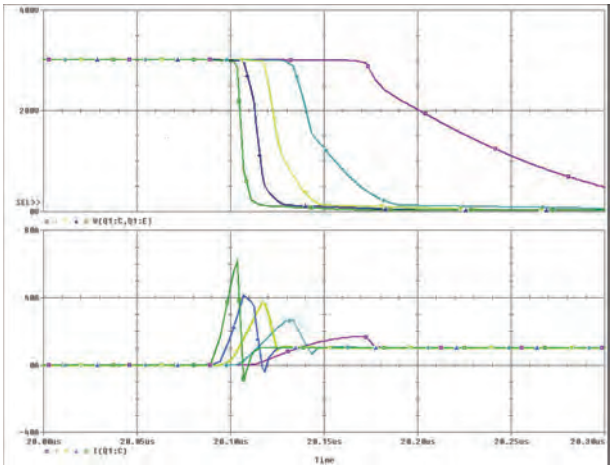


Fig. 9. Simulation results for the variation of the collector-emitter voltage for different values of V_{th}.

In addition, the elements R33 and R34, connecting the node GND to the ground 0 [9], are omitted (represented as comments by the symbol * in first column). The resulting modified *PSpice* model of IXDD414 is represented in Table I. The modified element descriptions are given in italic.

The obtained generalized *PSpice* model in Table I allows to connect the node DGND of the driver IXDD414 to arbitrary circuit node. In this way, circuits with two and more IXDD414 elements can be analyzed, as shown in Fig. 6.

IV. CONCLUSION

MOSFET/IGBT driver circuits have been investigated based on computer simulation. The influence of unipolar and bipolar supply voltage of the Gate Drivers is investigated on the peak value of the capacitance currents through the transistors and the switching power. A driver circuit is simulated that realizes the principle of two-level turn-on/one-level turn-off. The obtained results show the decreasing of the peak values of the current and the power during the commutation. The influence of the turn-on threshold voltage on *di/dt* and *dv/dt*. The PSpice model of IXDD414 is improved, which allows its usage as a high-side driver.

TABLE I: PSPICE MODEL OF IXDD414

```
.SUBCKT IXDD414 VCC IN EN GND  VCC OUT OUT GND
Q_Q5   N209276 N03676 GND BC107A
C_C5   N02728 GND  1uF
R_R9   N02728 GND  10k
C_C3   VEE GND  100uF
R_R24   GND TP  10K
R_R26   GND N242307  724
M_Q2 N04428 N209276 GND GND  MNMOS
R_R18   N01509 GND  10k
X_U4   VCC VEE GND LM7805C
R_R12    OUT GND  10k
R_R17    N82692 GND  10k
*R_R34  GND 0  0.0001
M_Q1  N04213 N209203 GND GND  MNMOS
C_C6   VSS GND  47uf
R_R70   GND TEST  1k
R_R15    N161018 GND  10k
C_C2  GND VCC  0.1uF
D_D6  GND N02728 BZ-056
C_C4   VCC GND  100uF
R_R39  GND N260557  2k
R_R35  GND VT  10k
D_D5  GND VSS BZ-056
R_R19   GND N295262  10k
R_R6   N03676 GND  10k
R_R7  N03600 GND  10k
R_R10  N02728 GND  10k
*R_R33   GND 0  0.0001
R_R27  GND N336660  10k
R_R37   GND TP  10k
Q_Q6  N209203 N03030 GND BC107A
R_R4   N03030 GND  10k
R_R36  GND N259753  10k
D_D2  GND N82834 D1N4376
M_Q4  N104168 N04213 GND GND MNFET
R_R28  GND N295639  10k
D_D4 GND N02182 D1N4376
R_R8   N02959 GND  10k
C_C1  VCC GND  220uF
R_R30   GND N242974  100k
R_R16   IN N82692  100
R_R25   N242307 EN  10K
M_Q3  N104168 N04428 VCC VCC MPFET
R_R1   VCC N04213  10K
R_R38  N260557 VCC  10k
R_R2   VCC N04428  10K
D_D6   GND N02728 BZ-056
R_R3    N02959 N03030  5
C_C7    N02959 N03030  1nf
D_D1    N82834 VCC D1N4376
R_R22   VCC N209276  5.0K
R_R5    N03600 N03676  5
C_C8    N03600 N03676 1nf
R_R13   HENABLE N02728  100
D_D3    N02182 VCC D1N4376
R_R23   VCC N209203  5.0K
R_R21   TEST N01509  0.0001
R_R31   VCC VSS  100
R_R11   N104168 OUT  0.01
R_R29   HENABLE N242974  10
X_U9A  N02728 N01509 N03600 VEE GND 74ACT00  PARAMS:
+  IO_LEVEL=0  MNTYMXDLY=0
X_U2A N260557 N260557 N259753 VSS  GND 74ACT08
+ PARAMS:: IO_LEVEL=0  MNTYMXDLY=0
X_U6A N336660 N336660 N295639 VEE  GND 74ACT08
+ PARAMS:  IO_LEVEL=0  MNTYMXDLY=0
X_U22A N295262 N295262 TEST VEE GND  74ACT00 PARAMS
+  IO_LEVEL=0  MNTYMXDLY=0
X_U10A  N01509 N161018 N02959 VEE  GND 74ACT02
+ PARAMS  IO_LEVEL=0 MNTYMXDLY=0
X_U5A TP N336660 VEE GND 74ACT04  PARAMS:
+IO_LEVEL=0   MNTYMXDLY=0
X_U1A  N242307 VT VSS GND 74ACT14 PARAMS:
+ IO_LEVEL=0    MNTYMXDLY=0
X_U7A N82692 TP VEE GND 74ACT04 PARAMS: IO_LEVEL=0
+  MNTYMXDLY=0
X_U3A VT N259753 HENABLE VSS GND  74ACT08 PARAMS
+   IO_LEVEL=0   MNTYMXDLY=0
X_U8A N02728 N161018 VEE GND 74ACT04  PARAMS:
+ IO_LEVEL=0   MNTYMXDLY=0
X_U23A N295639 N295639 N295262 VEE  GND 74ACT00
+  PARAMS: IO_LEVEL=0  MNTYMXDLY=0
.MODEL MNMOS NMOS LEVEL=3 W=4E-6  L=2E-6
+ NSS = 0.00000E+00 VTO = 6.30E-01 TOX = 1.90000E-08
+  XJ = 1.66632E-07  LD = 0.04E-06 RSH = 6.40359E+02
+  NSUB = 1.82429E+16 NFS = 1.26963E+11
+  UO = 5.95254E+02 VMAX = 1.86907E+05
+  DELTA = 8.22502E-01  THETA = 5.77858E-02
+ ETA = 2.1931E-02  KAPPA = 2.60564E-1  CGSO = 7.263E-11
+ CGDO = 7.263E-11 CGBO = 1.1189E-10  CJ = 2.59E-4
+CJSW= 2E-10 PB = 0.811 MJ = 0.36  MJSW = 0.31 TPG = 1
+ DW =-0.18E-6  DL =-0.72500E-06  XQC = 1
.MODEL MPFET PMOS(LEVEL=3 W=4e-3 L=0.01e-006)
.MODEL MNFET NMOS(LEVEL=3 W=4E-3 L=0.01e-006)
.MODEL BC107A NPN
.MODEL BZ-056 D RS=1.E-3  CJO=1.E-12  M=.3333 VJ=.75
+ ISR=7.539E-6  BV=5.009 IBV=43.11E-3 TT=144.3E-9
.MODEL D1N4376 D
.ENDS
```

REFERENCES

[1] Power MOSFET Transistor Data, Motorola Inc., 1992.

[2] IGBT Device Data, Motorola Inc., 1998.

[3] A.D. Pathak, "MOSFET/IGBT Drivers. Theory and Applications", Application Note IXAN0010, www.ixys.com

[4] Ch. Gerster and P. Hofer-Noser "Gate Controlled dv/dt and di/dt - Limitation in High Power IGBT Converters", EPE Journal, vol. 5, no. 3/4, pp. 11-16, Jan. 1996.

[5] N. Idir, J. Fraunchaud and R. Bausiere, " How to Reduce EMI Generated by IGBTs and MOSFETs", New Control Technique Achieves Low dv/dt and di/dt, PCIM EUROPE – Power Electronic, n.2, pp. 28-30, February 2000.

[6] H. Rüedi and P. Köhli, „Dynamic Gate Controller (DGC) – A new IGBT Gate Unit for High Current /High Voltage IGBT Modules", PCIM 1995, Nürnberg, Germany, pp.241-249,June 1995.

[7] Miller Clamping IGBT Gate Driver ACPL-331J, www.avagotech.com

[8] IGBT/MOSFET driver TD350, www.st.com

[9] MOSFET Gate Driver IXDD414, PSpice Model, www.ixys.com

CCTV Industry

Elena N. Koleva[1] and Ivan S. Kolev[2]

***Abstract* – Classification was carried out surveillance equipment in the industry and the optical elements used. Classification was carried out the matrix semiconductor photoconverter. Two systems have been developed to monitor certain processes in the industry in black and white cameras, color camera or thermal imaging cameras. Used cameras are connected in a closed circuit television (CCTV) camera, system and an internet connection (IP camera, Internet Protocol camera, Web camera) - IP system. The used monitors are picture in picture (PIP, Picture In Pucture). Observations can be carried out during the day and night.**

***Keywords* – *CCD* matrix semiconductor photoconverters, Infrared security CCD cameras, Night vision devices**

I. INTRODUCTION

Monitoring industry using black and white or color security cameras, solid state video camera (Camcorder, Handycam), digital cameras (PDA, Photographic Digital Apparatus), thermo vision cameras, equipment and night vision goggles, endoscopes, borockopes, fiberoscopes, periscopes, telescopes, visual tubes, microscopes and more.

II. CLASSIFICATION OF INFRARED (IR) SECURITY CAMERAS

There are two types of infrared security cameras: black and white and color.

Minimum illumination in black and white camera is 0,01 lx; 0,05 lx; 0,1 lx.

Minimum illumination for color cameras – 3 lx.

A resolution of the IR cameras, 360, 380, 420, 480, 600 Horizontal TV lines (TVL, TeleVision Lines).

The cameras can be installed in PID (PIR) (Passive Infrared Detector) detector, smoke detector, in a clock, a wall sconce, in chandelier, in glasses, rings, Spies, jewelry, earrings, necklaces, bracelets and more.

The cameras can be night or day surveillance and nighttime observation.

Security cameras work in the near infrared range (NIR) or in the far infrared range (FIR).

Infrared cameras can be integrated with the IR source (IR LEDs, Infrared Light Emitting Diode (s)), without built – IR source or halogen projector for visible light (Floodlight camera).

Voltage of security cameras battery is +12 V.

The cameras can be a Wireless security camera (radio) and wire connection with headquarters.

Infrared cameras are divided in cupola, cylindrical, parallelepiped, board, cameras on a single chip.

Depending on the infrared camera assembly are divided into external, internal and secret cameras. Mounting has waterproof outside cameras.

Cameras are CCD (Charge Coupled Devices) camera or CMOS (Complementary Metal – Oxide – Semiconductor) camera.

The cameras can be connected to the Internet (IP camera and Web camera).

Infrared cameras can be connected in closed circuit television sistem (CCTV camera system).

The cameras can be a zoom lens with an option for rotation and tilt PTZ camera (Pan Tilt Zoom camera).

Information from infrared cameras unless it can be observed directly on a monitor and can be recorded continuously or for a specified period of time.

Video cameras may be digital and analog. Image standards PAL (Phase Alternation Line (s)), CCIR (Consultative Committee for International Radio), NTSC (National Television System Committee).

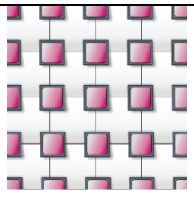
The observed object or process can be displayed on a cathode – ray tube (CRT, Cathode - Ray Tube), TFT LCD monitor (Thin - Film Technology Liquid Crystal Display screen).

Audiocompresion format Dolby Digital.

Record information on a magnetic medium (Mini DV (Digital Video) cassette, Compact Flash (CF); MultiMedia Card (MMC); Smart Media (SM); Secure Digital (SD); Memory Stick, Stick card (SC); USB flash drive or optical media; DVD (Digital Versatile Disk) drive and Blu – ray disk)

Video compresion formats MPEG2 (Motion Picture Expert's Group) 2 and JPEG (Joint Photographic Experts Groups).

III. CLASSIFICATION OF THE MATRIX SEMICONDUCTOR PHOTOCONVERTERS – FIG. 1 AND FIG. 2



Normal CCD matrix

Super CCD matrix (Fuji)

Super CCD SR matrix

[1]Elena N. Koleva is Ph. D., Department of Electronics, Technical University – Gabrovo, Street "Hadji Dimiter" No. 4, 5300 Gabrovo, Bulgaria, phone: +359 898 226 464,
e-mail: elena_ndpt@yahoo.com

[2]Ivan S. Kolev is Prof., Dr. Sci., Department of Electronics, Technical University – Gabrovo, Street "Hadji Dimiter" No. 4, 5300 Gabrovo, Bulgaria, phone: +359 898 634 633,
e-mail: ipk_kolev@yahoo.com

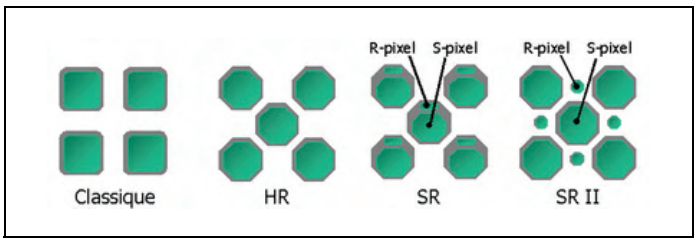


Fig. 2. Types of CCD matrixes

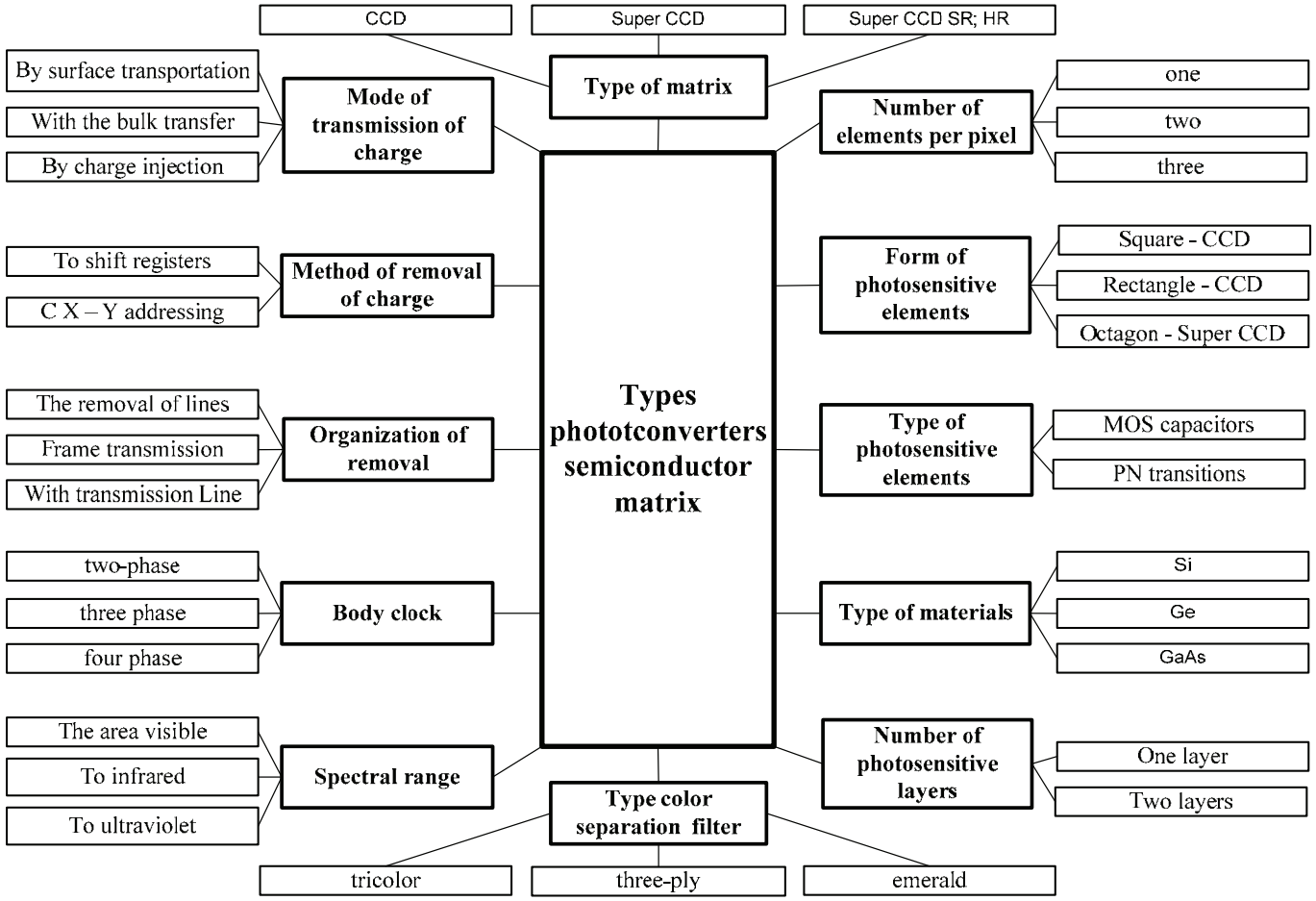


Fig. 1. Matrix semiconductor photoconverters

A. Super CCD matrixes semiconductor photoconverters

Under the new Super CCD matrix elements are angled 45° and have a shape of an octagon. This will reduce the area occupied by circuit buses. New principle available to the photosensitive elements resembles honeycomb. Thus increasing the effective resolution of 60% compared to traditional matrixes. These matrixes with higher resolution CCD and Super CCD matrixes.

B. Super CCD HR (High Resolution) matrixes

These are matrixes with very high dynamic range – fourth generation dies. The structure of the Super CCD SR lies in the combination of large photodiodes (S – pixel), having greater sensitivity to small photodiodes (R – pixels), having less sensitivity.

Monitoring with very high resolution bright and dark sectors are used to Super CCD SR matrixes.

C. Other types of CCD arrays and CCD devices

Super Hole – Accumulation Diode CCD (Super HADCCD); Electron Multiplying Charge Coupled Device (EMCCD); Intensified Charge Coupled Device (ICCD).

IV. BLACK AND WHITE OR COLOR SECURITY CAMERAS FOR MONITORING IN INDUSTRY

A. Black and white security cameras

Application for continuous monitoring of industrial processes in normal and low light.
Parameters of IR security CCD cameras used to work near infrared range (NIR).

Parameters of the IR camera type KP – F2A company Hitachi:

- Maximum spectral sensitivity of - 760 nm (700 to 1000 nm);
- Signal format output - 30 fps;
- Dimensions of the sensitive element - 1 / 3″ (8,47 mm) CCD;
- Number of pixels horizontally and vertically (H x V) - 658 x 496;
- Pixel size (H x V) - (7,4 x 7,4) mμ;
- Sensitive area (H x V) - (4,87 x 3,67) mm;
- Horizontal resolution - 500 TV lines (TVL);
- Video Output - 1 Vp-p / 75 Ω BNC;
- Installation of the lens - C-mount;
- Distance (min) - 17,5 mm;
- Synchronization - Internal/ external;
- Minimum illumination of light - 0,3 lx;
- Signal/ noise ratio - 50 dB;
- Electronic Shutter speed - or off 1/ 60, 1/ 125, 1/ 250, 1/ 500, 1/ 1000, 1/ 2000, 1/ 4000 and 1/ 8000 s (second);
- Gamma - 0,45/ 1,0 elected;
- Automatic gain control - On/ off;.
- Working temperature - (-10 ÷ +50) °C
- Supply voltage - 12 V DC;
- Power supply - 220 V/ 50 Hz;
- Dimensions - (44 x 44 x 87) mm;
- Weight - 170 g.

(Fps – frames per second (frame (s) per second), H/ V (Horizontal/ Vertical), pp (peak - to - peak), BNC (Bayonet Nut Connector), DC (Direct Current)).

In Fig. 3 is a spectral characteristic of infrared CCD camera type KP – F2A company Hitachi.

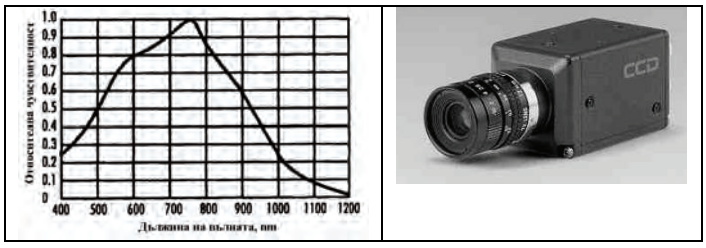


Fig. 3. Spectral characteristic and CCD camera

CMOS black – white camera company Conrad (2010):

- Supply voltage - (9 to 12) V DC;
- Login - audio/ video (A/ V);
- Antibloming – yes;
- Consumed current - 18 mA;
- Video Output - 1 V_{p-p} / 75 Ω;
- Standard - CCIR, 628 pixels (H) x 582 (V);
- Angle of acceptance - 70° x 63°;
- Signal/ noise ratio > 48 dB;
- Sensitivity - 0,5 lx;
- Dynamic Range - 78 dB;
- Photodetector sensor - OV7410;
- Aperture - 2,0;
- Focal length - f = 3,7;
- Number of pixels - 365 496;
- Length of the camera - 44 mm;
- Chamber diameter - 24 mm.

A/ V (Audio Video), OV (Operational View)

B.Colour security cameras

Color video camera for surveillance PIH70XX/ H

- Lighting - 0,01 lx;
- Digital signal processing (DSP);
- Horizontal resolution - 480 TV lines;

C. PDA

PDA are classified in different signs: the number of sensor arrays: one CCD, three CCD matrixes, the number of photodetectors (PD) in a pixel: PD one, two PD, the number of color filters: with three filters, four filters.

V. DEVICES AND NIGHT VISION GOGGLES

Application: monitoring of processes and equipment in the industry at very low light.

A. Night vision devices (NVD)

A major element in them is electro converter EOP (EIT, Electron Image Tube).

NVD four generations. When I - st generation NVD currents are sensitive photocatode (PC) 60 μA/ lm. In III – Generation NVD photocathode is gallium arsenide, a microchannel plate amplified coated with ion barrier. Spectral sensitivity of the PC is 450 to 950 nm. In Fourth Generation NVD currents sensitivity of PC is 1800 μA/ lm, Watt sensitivity (λ = 830 nm) - 190 mA/ W, the resolution - 64 lp/ mm, halo - 0,75 mm, signal/ noise ratio – 25, reliability - 10 000 pm, detected distance - 480 m. Tube type used is "Gated filmless tube". Microchannel plate (MCP) consists of a bundle of optical fibers to 10^6, which increase the brightness.

B. Night Vision Goggles

Reference data for night vision goggles with a double pipe type "Diana TT", production company OPTIX – Panagyurishte.

1. Magnification (1x).
2. Field of view (FOV) - 50,9 degrees.
3. Resolution - 45 ÷ 57 lp/ mm.
4. Finding rangem - ~ 270 m.
5. Orientation range - ~ 225 m.
6. Detection range (Recognition range) ~ 105 m.
7. Range of Recognition (Identification range) ~ 55 m.
8. Focus range - 0,2 m ÷ ∞.

9. Operating wavelength diode to illuminate - 880 nm.
10. Exit pupil relief - 25 mm.
11. Exit pupil diameter - 18 mm.
12. Eye basis - fixed - 66 mm.
13. Inter - pupil (ary) distance (IPD) 6 mm - 54 ÷ 98 mm.
14. Diopter adjustment - + 4 ÷ - 6 diopters (dpt).
15. Power supply alkaline battery – 1,5 V.
16. Power supply lithium alkaline accumulator - 1,2 V.
17. Life of the battery - 30 hours.
18. Built in automatic disconnection at illumination higher that 5 ÷ 10 lx.
19. Weight - 630 grams.
20. Overall dimensions (Length x width x height) - 103 x 106 x 64 mm.
21. Operating temperature – 50 ^{o}C ÷ + 50 ^{o}C.
Couple lp/mm- lines pairs per millimeter.

VI. THERMOVISION SYSTEMS OR THERMAL IMAGING CAMERAS

Sensitive elements in thermo vision systems for the far infrared part of the spectrum are the photodetectors (after 6 μm). On the thermal imaging system, except transmission by object falling and disturbing radiation emitted by the environment and reflected radiation from surrounding objects.

Application of IR systems: monitoring of temperature change, sites of chemical industry (process of formation of parts, cooling rates of the polymer melt, temperature distribution of injection molds, drawing optical fiber) oil industry (temperature control in tanks); Construction (analysis of thermal insulation breakdown) electronics (production of semiconductors, thermotransfer, especially in powerful components and appliances), medicine (distribution of skin temperature, diagnostics of the thyroid gland, subcutaneous location of disease outbreaks) measuring temperature of the sheet; detection of flying objects (airplanes) to measure the temperature of the Sun, Moon and planets, border guards and others.

VII. ENDOSCOPES, AND BOROSKOPES FIBEROSKOPES MONITORING IN INDUSTRY

Application: For monitoring and non – destructive control and diagnostics of internal surfaces of machinery, equipment, tools, assemblies and components which are not accessible to the naked human eye.

		
Video image endoscope of VID EXPERT+ Series	318 Series fiber optic borescopes	Fiberscope

A. Video image endoscope (VID) series:

- Diameter - 6/0mm - 8/0mm;
- Operating length - 1.0 m ÷ 6.0 m;
- Articulation - 2 or 4 ways (120° left/ right and (or) over 160° up, 130° down, 100° left/ right);
- Image Sensor - 1/ 4"color super - HAD PZS (Hole - Accumulation Diode Positioning Zone Sensor);
- Resolution - 330 TV lines;
- Video Output - NTSC/ PAL;
- Camera/ Image control - on the handhold;
- Handhold - aluminium with a black powder treatment;
- Survey direction - 0° forward;
- Objectives - are provided with the objective of 58° angular field where the color code black and the objective of 90° angular field where the color code red is;
- Covering of tube - Tungsten armor;
- Bending radius of the tube - 50mm;
- Operating temperatures - Operating end: - 10° ÷ 80° C;
- Body (case) and cable of the lighter: - 10 °C ÷ 50 °C;
- Bars.

VIII. CONCLUSION

Classification is done by the authors of the matrix semiconductor photoreceivers, classification of the cameras and classification of devices for monitoring processes in industry. Proposed are two systems for monitoring CCTV and IP. Elected representatives are typical of instruments for monitoring during both day and night. Here are the basic parameters and characteristics of the equipment.

REFERENCES

[1] Колев, И. С. и Е. Н. Колева. Оптоелектронни сензори и оптоелектронни охранителни системи. Габрово, ТУ, Университетско издателство „Васил Априлов”, 2009.
[2] Opticoelectron group. Panagyurishte, Bulgaria, 2008.
[3] Press, 1994. Optix company. Panagyurishte, Bulgaria, 2008.

PSPICE Simulations of Pulse Distributor in Industry

Elena N. Koleva[1] and Tsanko V. Karadzhov[2]

***Abstract* – The pulse distributor is realized on three integral timers. It generates square pulses shifted in time. By using PSPICE software, a simulation of the operation of the distributor has been done, as well as an adjustment of the time-setting elements and the parameters of the pulse packet. The pulse distributor has been designed to be used in industry.**

***Keywords* – PSPICE Simulations, Pulse Distributor.**

I. INTRODUCTION

Are proposed scheme of distribution of pulses developed on the basis of the timer 555 and converter angle into digital code and pulse duration on the basis of a timer 555, which is managed by a single – chip microcontroller PIC18F252.

Simulations are made of schemes of work with the software PSPICE.

II. ALLOCATOR PULSE TIMER 555

Scheme distributor pulses is shown in Fig. 1. It's based on three timer which are connected in series. The first timer is started using a brief pulse low, which made its finding 2. He made pulse duration 105 μs. The end of this second pulse starts a timer which produce pulse duration 105 μs. The end of the second pulse timer starts timer third, who also produced pulse of 105 μs. The duration of each pulse timer is determined by time constant of the chains time determine timers R1-C1; R3-C4; R5-C7. Thus the outcome of each timer can be obtained dephasig sequences of pulses with different duty cycle.

In Fig. 2 and Fig. 3 shows the timing diagram of operation of the scheme in the period of the clock pulses 750 μs and 1500 μs.

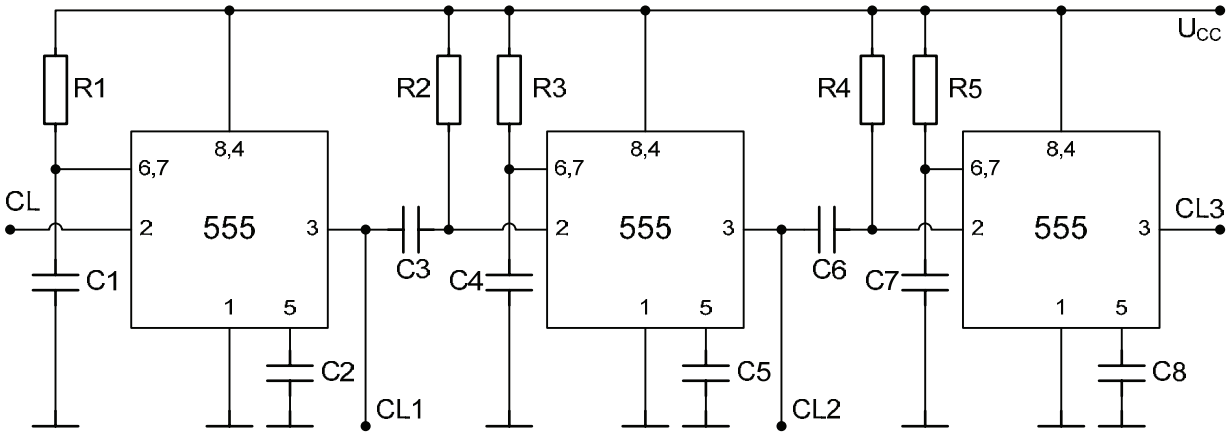


Fig. 1

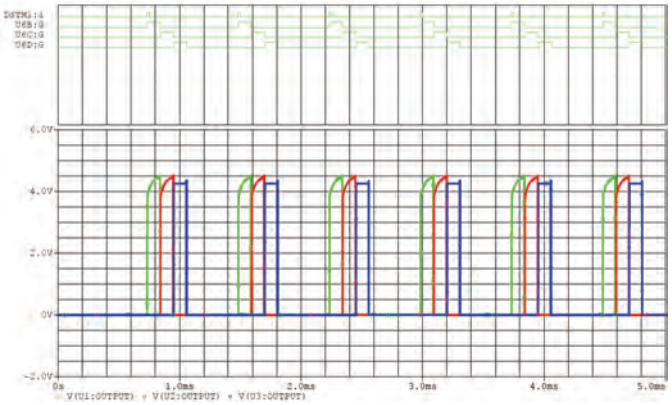


Fig. 2

[1]Elena N. Koleva Technical University of Gabrovo, 4 H. Dimitar, 5300 Gabrovo, Bulgaria, E-mail: elena_tugabrovo@abv.bg

[2]Tsanko V. Karadzhov Technical University of Gabrovo, 4 H. Dimitar, 5300 Gabrovo, Bulgaria, E-mail: karadjov_st@abv.bg

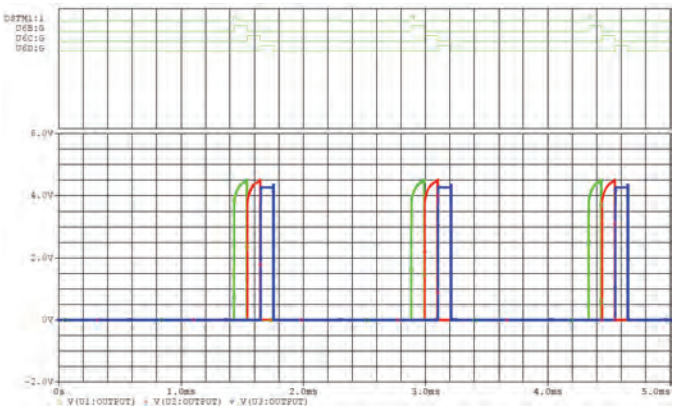


Fig. 3

III. CONVERTER ANGLE OF ROTATION IN A NUMERIC CODE AND PULSE DURATION

The scheme of the converter is shown in Fig. 4. Converter is developed based on single – chip microcontroller PIC18F252 and timer 555. For visualization of the information used two line alphanumeric LCD displays. Table I gives the characteristics of the chosen microcontroller.

TABLE I

Program memory capacity	32 K
Available I / O ports	PORT A, B, C
RAM memory capacity	1536
Number of sources of interruption	17
Number of channels of analog-digital converter	5
Body/Casing/	28 pin DIP
EEPROM memory capacity	256
Timers	4
Maximum clock frequency	40 MHz
CCP modules	2
Modules in series communication	MSSP, USART
Number of instructions	75

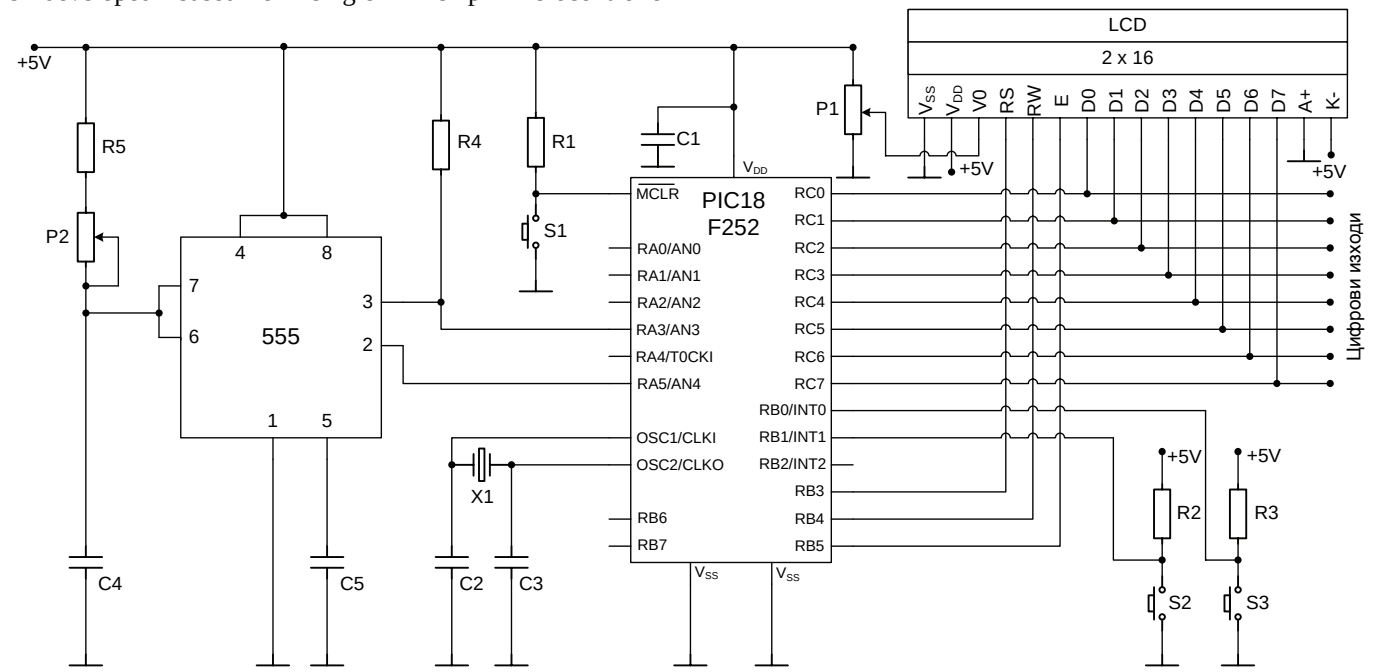


Fig. 4

Timer 555 operates as a waiting multivibrator. Allowing low-level pulse is applied to the conclusion of the microcontroller 2 of the timer. Authorizing pulse is less than pulse duration, the timer generates. Therefore, each authorization timer generates a single pulse output 3. At the same time and start timer 1, which is one of four built-in timer of the microcontroller. The duration of the output pulse depends on time constant of the time determine chain which is made up of elements R5, P2 and C4. Time constant altered by changing the resistance in time determine circuit by the potentiometer P2.

This regulates the duration and rate of filling

the output pulses. The period of these pulses is fixed by the microcontroller and 10 ms (100 Hz). The duration of each output pulse is measured by timer 1. As a clock stroke signal for timer 1 is used $f_{OSC}/ 4$, which is f_{OSC} clock stroke frequency of the quartz oscillator of the microcontroller - 10MHz. Timer 1 counting pulses during the duration of the output pulse of timer 555. Thus, the Timer 1 registers will have a numeric value that is proportional to the duration of the pulse output of timer 555. This value can be derived from the digital outputs of the microcontroller in binary form or can be displayed on the LCD alphanumeric display.

Thus, microprocessor system can convert the angle of rotation of the potentiometer P2 in numeric.

In Fig. 5 and Fig. 6 are given the signal timing diagram of output 3 timers 555 at two different positions of rotation of the potentiometer P2.

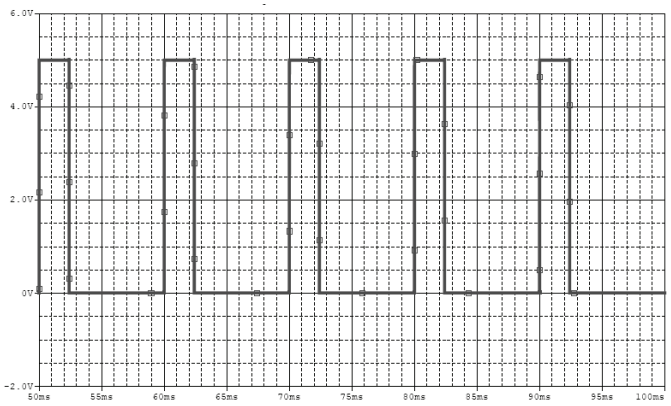


Fig. 5

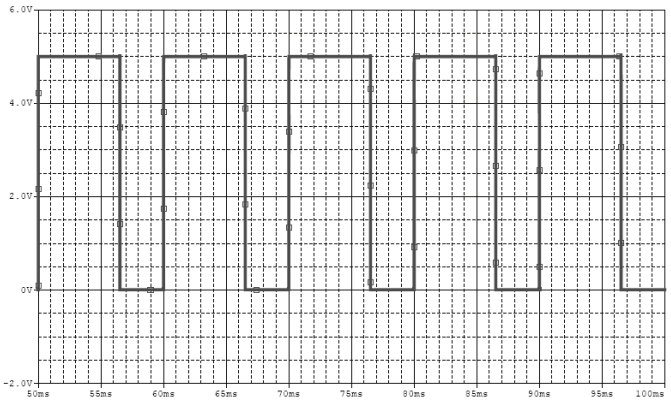


Fig. 6

In the Table II are given a pulse duration of the source and the coefficient of filling for various values of the resistance of the potentiometer P2.

TABLE II

Resistance, kΩ	t, ms	t/T, %
0	0.755	7.55
5	1.130	11.30
10	1.505	15.05
15	1.880	18.80
20	2.254	22.54
25	2.628	26.28
30	3.002	30.02
35	3.374	33.74
40	3.750	37.50
45	4.125	41.25
50	4.496	44.96
55	4.871	48.71
60	5.245	52.45
65	5.618	56.18
70	5.990	59.90
75	6.364	63.64
80	6.737	67.37
85	7.111	71.11
90	7.489	74.89
95	7.862	78.62
100	8.231	82.31
105	8.611	86.11
110	8.983	89.83

A. Depending on the duration of the pulse function in the resistance of potentiometer

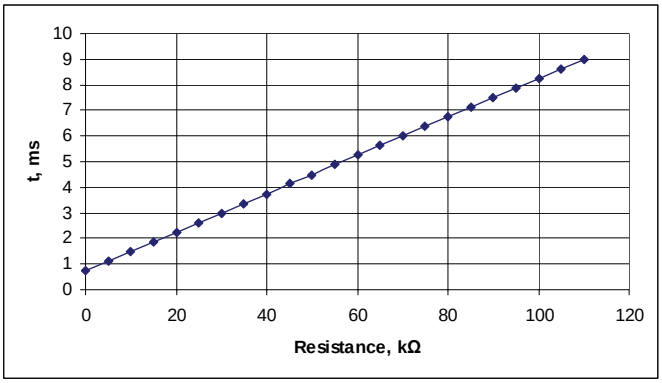


Fig. 7

B. Depending on pulse duty factor in the function of the potentiometer resistance

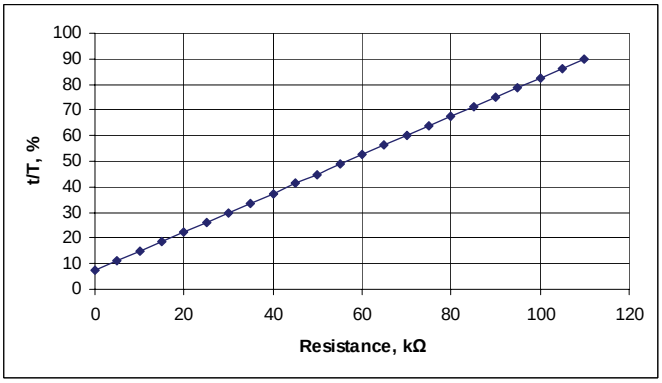


Fig. 8

Advantages of the converter of Fig. 4:

1. Converts an angle in the pulse duration of high linearity;
2. Pulse can be measured with very high accuracy, because clock frequency of Timer 1 (2.5 MHz) is much greater than the frequency of pulses generated by timer 555 (100 Hz);
3. Information from the angle of rotation can be displayed on LCD display or transmit parallel code as a digital device to another.

IV. CONCLUSION

The proposed schemes are developed and tested with the software PSPICE. The developed converter angle into digital code and pulse duration can be used in many areas of pulse and digital equipment such as a generator of pulses with constant period and variable duty cycle or as a digital potentiometer. Developed a program in assemble language by which the microprocessor system can convert the angle into a digital code or ASCII code which is suitable for displaying information on alphanumeric displays.

REFERENCES

[1] Н. Кенаров, PIC Микроконтролери,. част 2, Млад конструктор, Варна, 2006.
[2] Р. Трейстер, Радиолюбительские схемы на ИС типа 555, Мир, Москва, 1988.

Generator of Synchronizing Digital Signals with Microcontroller PIC18F252

Tsanko V. Karadzhov[1] and Ivelina S. Balabanova[2]

***Abstract* – Synchronizing signal generators are an integral part of the synchronous digital devices. This makes it necessary to develop new methods and principles in the design of such generators. Deficiencies in the classical schemes are synchronizing signal delays in logic elements that can lead to incorrect output responses. This paper presents the design of a generator for digital synchronizing signal microcontroller PIC18F252, in which the parameters of the output signals have been improved.**

***Keywords* – Microcontroller, Digital signals, Generator.**

I. INTRODUCTION

This paper presents a circuit for producing dephased pulses with CMOS logical elements, a rectangular pulse generator voltage and a frequency divider, a circuit control by TTL logical elements. Circuit operation with PSPICE software has been simulated. A microprocessor system with PIC18F252 microcontroller has been developed which can perform the function of the three circuits.

II. DEFASED PULSE GENERATOR

Fig. 1 shows a circuit for obtaining dephased pulses. D trigger is used to switch on the rising edge.

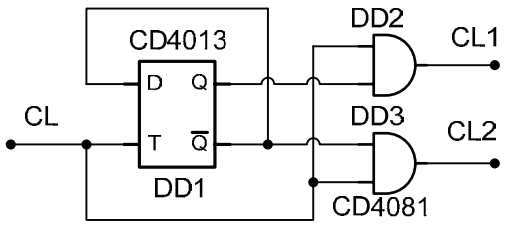


Fig. 1

Fig. 2 shows the time charts of the simulated operation of the circuit. A clock signal, two trigger output and the output signal CL2 in larger scale are shown. From the time chart it is seen that, due to delays in switching the logical elements incorrect output effects are obtained that may impair the operation of digital devices using such a generator.

[1]Tsanko V. Karadzhov Technical University of Gabrovo, 4 H. Dimitar, 5300 Gabrovo, Bulgaria, E-mail: karadjov_st@abv.bg

[2]Ivelina S. Balabanova Technical University of Gabrovo, 4 H. Dimitar, 5300 Gabrovo, Bulgaria, E-mail: ivstoeva@yahoo.com

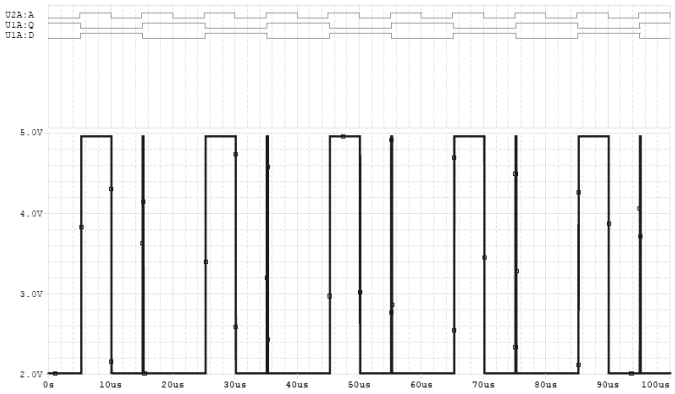


Fig. 2

III. RECTANGULAR PULSE GENERATOR WITH VOLTAGE CONTROL

Fig. 3 shows the circuit of the generator.

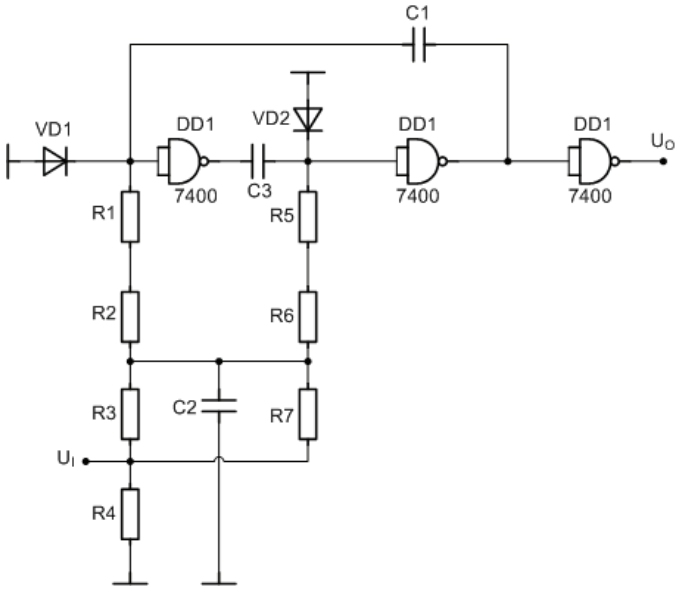


Fig. 3

Frequency variation range depends on the values of the elements C1, C3, R1, R2, R3, R5, R6, R7 the control input voltage is applied from 0 to 1 V. Fig. 4 and 5 show the time chart of the system at two different voltages 0 V - Fig. 4 and 0.8 V - Fig. 5.

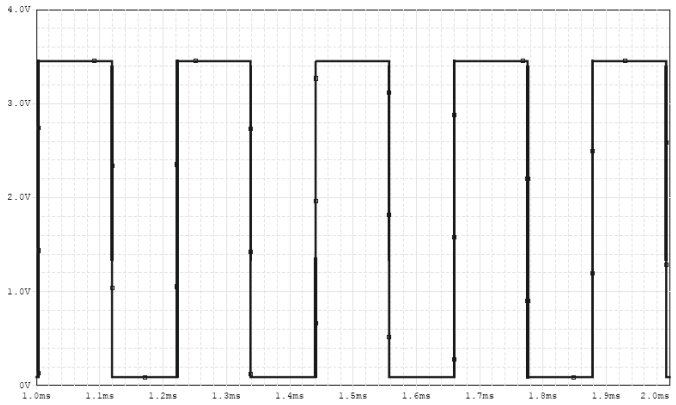


Fig. 4

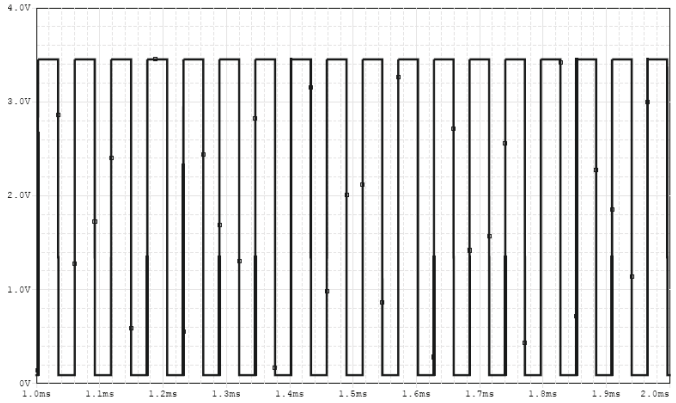


Fig. 5

The frequency can vary within wide limits. Table I gives the generator frequency at different input voltages.

TABLE I

voltage	time	frequency
0 V	218 µs	4.578 kHz
0.1 V	207 µs	4.831 kHz
0.2 V	193 µs	5.181 kHz
0.3 V	180 µs	5.556 kHz
0.4 V	158 µs	6.329 kHz
0.5 V	132 µs	7.576 kHz
0.6 V	107 µs	9.346 kHz
0.7 V	82 µs	12.20 kHz
0.8 V	56 µs	17.86 kHz
0.9 V	31 µs	32.26 kHz

Fig. 6 gives a diagram of generator frequency variation as a function of input control voltage. The disadvantage of this generator is that the dependence is nonlinear, but exponential. The input voltages frequency near 1 V is unstable. It changes at a very little change in the control voltage.

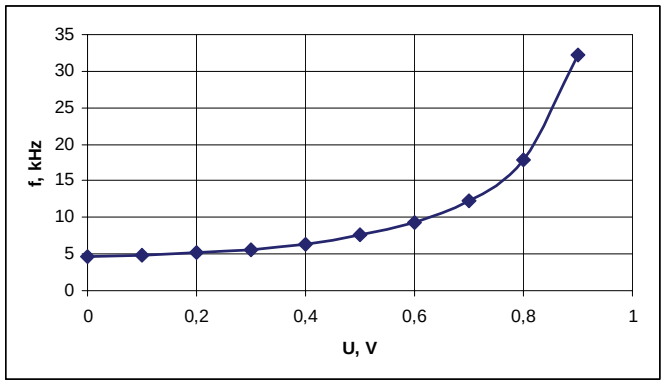


Fig. 6

IV. FREQUENCY DIVIDER

The frequency divider is developed with an adjustable division coefficient in Fig. 7.

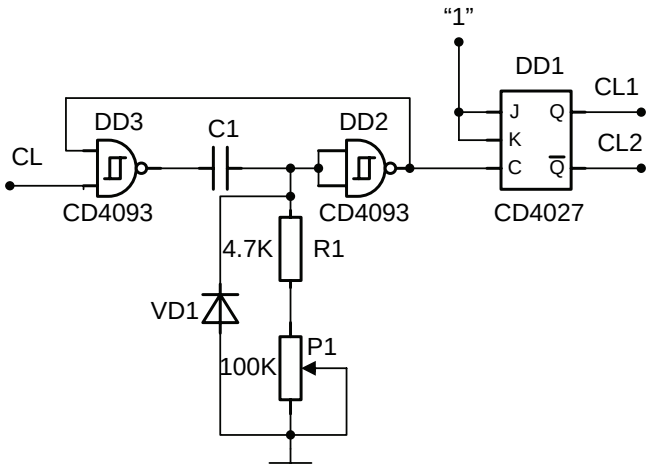


Fig. 7

Division coefficient may be modified a very wide range by adjusting the resistance of potentiometer P1. The divider is made on the bases of two logical elements AND-NOT. JK bistable is used to form the output pulses. The two elements form a monostable multivibrator. In the initial state of circuit operation condenser C1 isn't charged and the input of logic element DD2 is going to have a logical "1" as well as at the output of DD3. At that time, the output of DD2, which is connected to one input of DD3 has a logical "0" and the input clock pulses aren't get pass a pulse. The condenser is charged by the output of logic element DD3 through the resistor R1 and potentiometer P1. Once it is loaded DD2 switches and its output has high level. This enables the get pass a pulse of input pulses. When a pulse input condenser is quickly discharged in diode VD1 and the process is repeated. The time in which input pulses get pass a pulse is determined by constant of C1, R1 and P1, and this determines the coefficient of division. Fig. 8 and 9 show the time charts of operation of the circuit for division ratios of 2 and 6. The input pulses, the output pulses and the voltage condenser are given. Signal generator with PIC18F252.

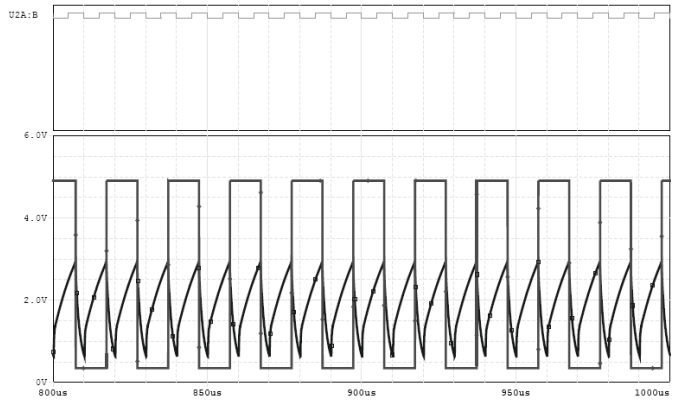


Fig. 8

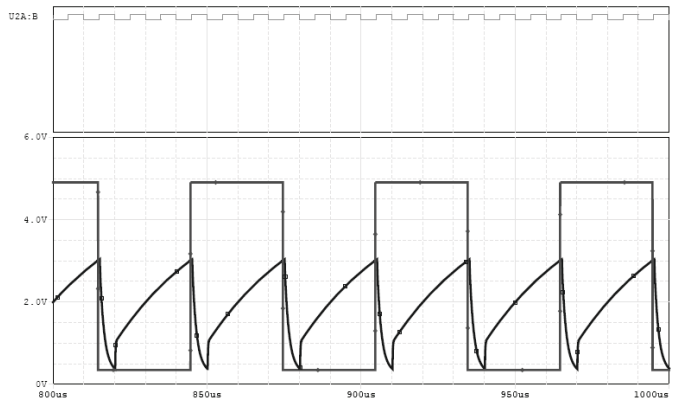


Fig. 9

V. SIGNALS GENERATOR WITH PIC18F252

Generator circuit is shown in Fig. 10

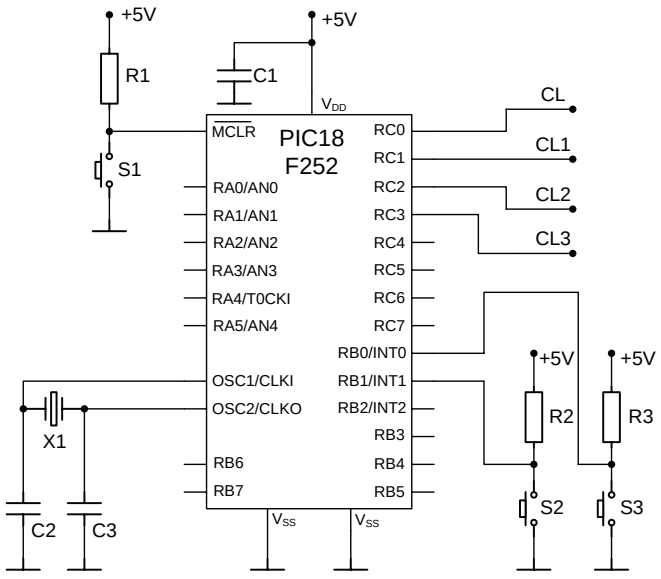


Fig.10

Generator is developed on the bases of PIC18F252 single-chip microcontroller. It can generate four output signals and has two buttons to control the operation modes. An advantage of the circuit is that the external elements necessary for the operation of the microcontroller are a very small number.

The characteristics of the chosen microcontroller are given in Table II.

TABLE II

Program memory capacity	32 K
Available I / O ports	PORT A, B, C
RAM memory capacity	1536
Number of sources of interruption	17
Number of channels of analog-digital converter	5
Body/Casing/	28 pin DIP
EEPROM memory capacity	256
Timers	4
Maximum clock frequency	40 MHz
CCP modules	2
Modules in series communication	MSSP, USART
Number of instructions	75

A program has been developed in assembly language by which the microprocessor system can cover the functional capacity of the generators discussed above. Control button S2 can switch different generators, and control button S3 sets the parameters of the generated signals.

Some of the shortcomings of the circuits discussed above have been avoided in this generator.

The time charts of the operation generator signals PIC18F252 are given in Fig.11, 12 and 13.

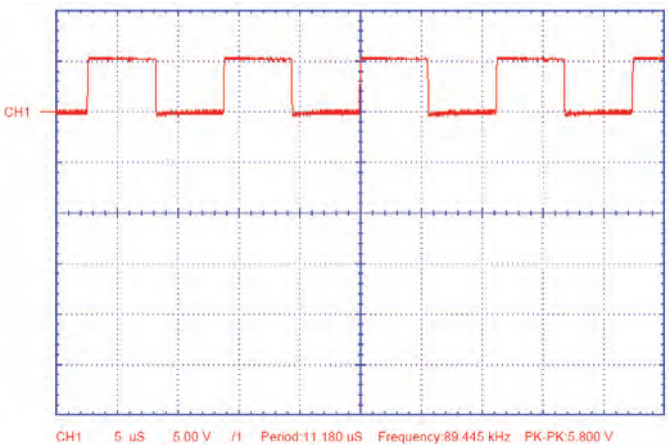


Fig. 11

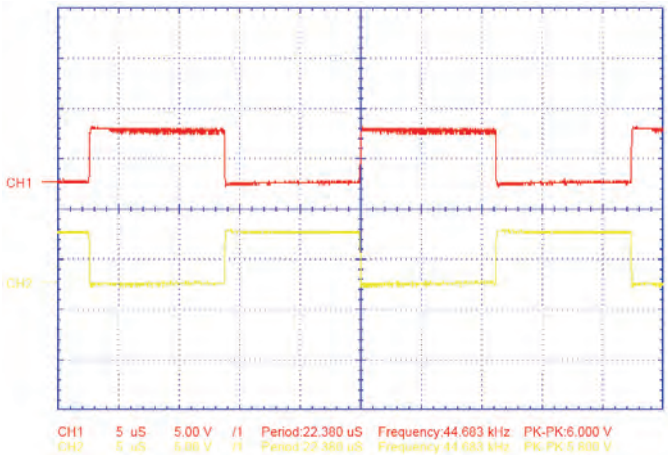


Fig. 12

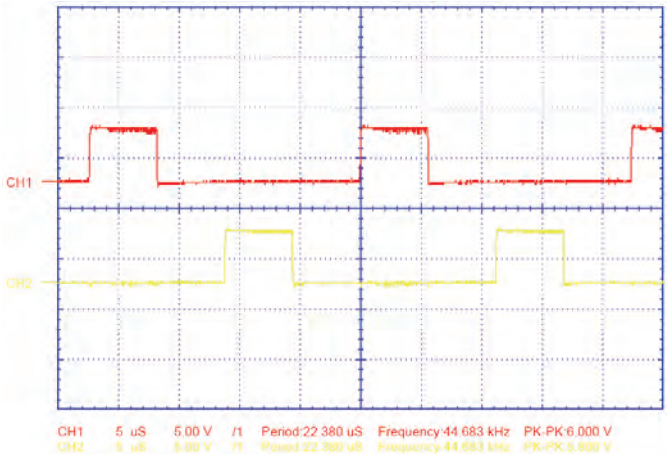


Fig. 13

VI. CONCLUSION

The proposed circuits have been developed and tested using PSPICE software. The developed of digital synchronizing signals with PIC18F252 microcontroller can find application in synchronous digital devices that require clock control signals in telecommunication.

REFERENCES

[1] Кенаров Н., PIC Микроконтролери,. част 2, Млад конструктор, Варна, 2006.

[2] Iliev, A. K., Research of Digital Sinchronizers, Unitech'09, Gabrovo, 20-21 November, 2009, volume I, p.124.

Improved Grid Connected Power Supply for Drivers in Power Electronic Converters

Vencislav C. Valchev[1], Dimitar Bozalakov[1] and Angel St. Marinov[1]

***Abstract* – The paper presents a specialized power supply for drivers in power electronic converters. The proposed topology can be directly interfaced to the electric grid without the need of a rectifier or high voltage filter capacitor. Advantages of the presented circuit are: reduced complexity, size, weight and price. The suggested device supplies power and galvanic separation of the driver stages of high power electronic converters. The scheme was tested and validated experimentally.**

***Keywords* – Switch mode power supply, Power electronics, MOSFET/IGBT drivers, Insulated Power Supply**

I. INTRODUCTION

The paper presents a specialized power supply for drivers in power electronic converters. The suggested device is meant to provide power supply and galvanic separation of the driver stages of high power electronic converters. Figure 1 shows a classical driver topology, presented for one power leg. Utilizing it in a 3-phase system requires high number of power supplies with galvanic separation: one power supply for each high side transistor plus one for all the low side transistors. Using this type of driving topologies however improves the safety and reliability, as well as the EMC characteristics of the power electronic converter. Anyhow using it leads to some certain disadvantages, due to the influence of the overall complexity of the whole scheme, which increases the gross price, weigh and size.

The paper introduces a topology that can be directly interfaced to the electric grid without the need of a rectifier or high voltage filter capacitor. This in terms will reduce the circuits complexity, improving the parameters stated in the previews paragraph. The introduced circuit in its specifics can be described as a switch-mode AC/AC converter that uses high frequency transformer to obtain the required voltage transfer ratio and number of insolated outputs. Only a small capacitor in the output is required in order to filter the output voltage.

Using a separate connection to the AC grid makes the power supply independent from the power converter. This provides a stable power source for converter, which DC line cannot be used for driver supply, for example converters controlling energy generation.

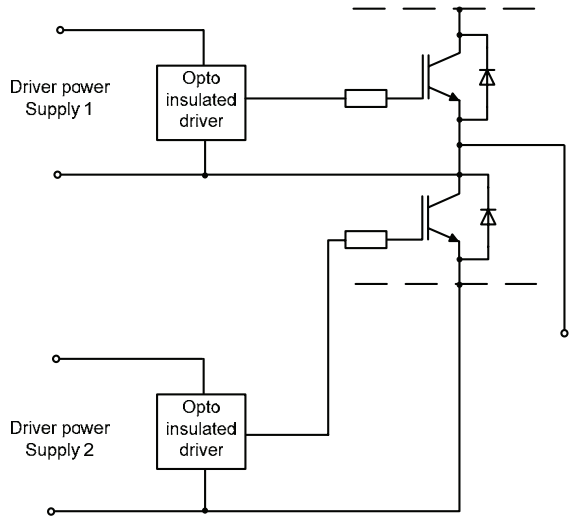


Fig. 1. Drivers with galvanic separated power supply

II. PRESENTED TOPOLOGY OPERATION

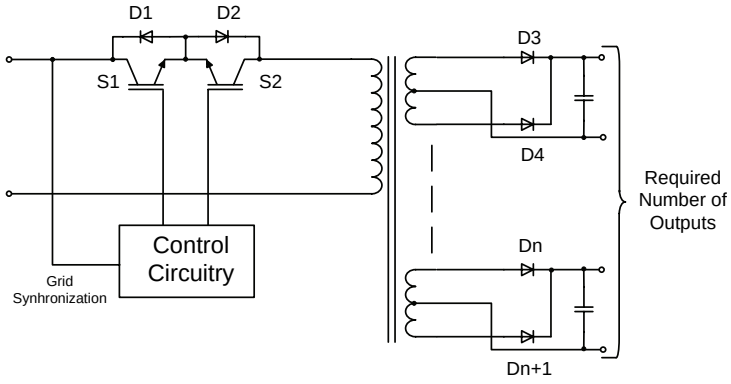


Fig. 2. Proposed Power Supply topology using AC voltage

A. Description of the presented circuit

The suggested topology is given at figure 2. In its essentials the circuit is a modified AC/AC converter formed by transistors S1, S2 and their corresponding anti-parallel diodes D1, D2. In the developed modification the transistors are

[1]Vencislav C. Valchev, Dimitar Bozalakov and Angel St. Marinov are with the Faculty of Electronics, Technical University of Varna, 9010 Varna, Bulgaria, E-mail: vencivalchev@hotmail.com, igdrazil@abv.bg

directly interfaced to the AC grid. The input voltage is transferred to the primary winding of a pulse transformer, as S1, D2 and S2, D1 respectively conduct the positive and negative half-wave. During each half-wave only one transistor is controlled, the other is off. The corresponding wave is being chopped by the corresponding transistor - creating a high frequency (30kHz to 50kHz) square pulse voltage applyied to the primary winding of the transformer. The amplitude of each pulse is determined by the instant value of the input sine wave voltage. This voltage is then transferred to the required number of outputs with the necessary voltage transfer ratio. The voltage from the secondary windings is rectified and filtered by the output circuitry. Thus, for each half-wave the circuit behaves as a flyback converter. For example during the positive half-wave S1, D2 and D3 form the 'first' flyback converter. During the negative half-wave S2, D1 and D4 form the 'second' flyback converter. Thus the low frequency sine wave voltage can be transferred to the secondary windings and then rectified and filtered. Due the nature of the chopped sin wave input voltage, the output voltage will have higher level of pulsations then when using a regular flyback working with rectified DC voltage, this however will not affect neither the supplied drivers – considering a sufficient output voltage, nor the other components of the circuit – the pulse transformer or the power transistors.

B. DC and AC input interfacing

As described, the circuit can be directly interfaced either to an AC or to DC line, depending on the available voltage source. When connected to a DC line the circuit can operate as normal flyback converter utilizing only one of the transistors –S1 and one half of the output winding. This gives the circuit significant flexibility as a pre-prepared circuit can be used for wide range of applications.

C. Output interfacing

The circuit can posses a high number of output windings, thus providing a sufficient number of output voltages with galvanic separation necessary for supling the required number of drivers. For example for a three phase inverter, the circuit can provide four power supplies with galvanic separation needed to supply drivers and drive the transistors (three power supplies for the high side transistors and one power supply for the low side transistors). Since drivers can sometimes act as open circuit (when the controlled transistors is been kept off for prolonged durations of time) a special protection circuit – zener diode D1 and resistor R1 is added. The circuit provides alternative path for the transferred energy during open circuit mode and protects the capacitor against overvoltages. (figure 3) The zener diode should be selected with a stabilization voltage higher than the required output.

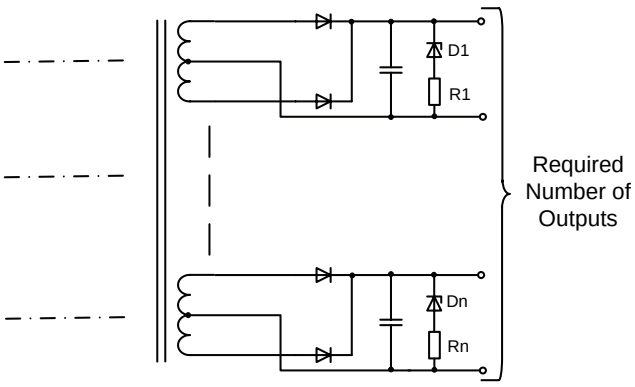


Fig. 3. Output Interface

D. Advantages and Disadvantages of the presented topology

The suggested topology has several main advantages because of the reduced number of components compared to the classic flyback circuits. The presented circuit avoids a separate rectifying and filter circuit which reduces the overall complexity of the circuit and allows the fallowing advantages:

1. Reduced size and weight since no high voltage input capacitor is used
2. Reduced gross price, determined by the difference of the value of one extra IGBT compared to the high voltage capacitor and rectifier.
3. Flexibility – the circuit can be interfaced to both DC and AC lines with different voltage. Thus it can be used as a laboratory supply for drivers or pre-prepared circuits can be connected directly to the available source.
4. Spin-off technologies – the circuit can be introduced in different fields where power supplies with large number of outputs with galvanic separation are necessary.

III. Experimental Results

The suggested circuit was tested for functionality and it proved to be operational, working without notable drawback. Due to safety issues the circuit was tested using reduced galvanic separated AC voltage, where only the positive half-wave was utilize. Complete tests utilizing the whole AC voltage will be shown at the paper's presentation.

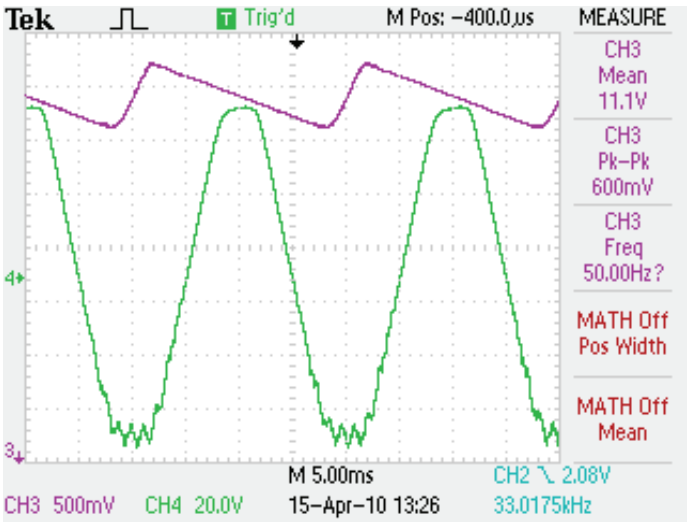


Fig. 4. Waveform of the input and output voltage

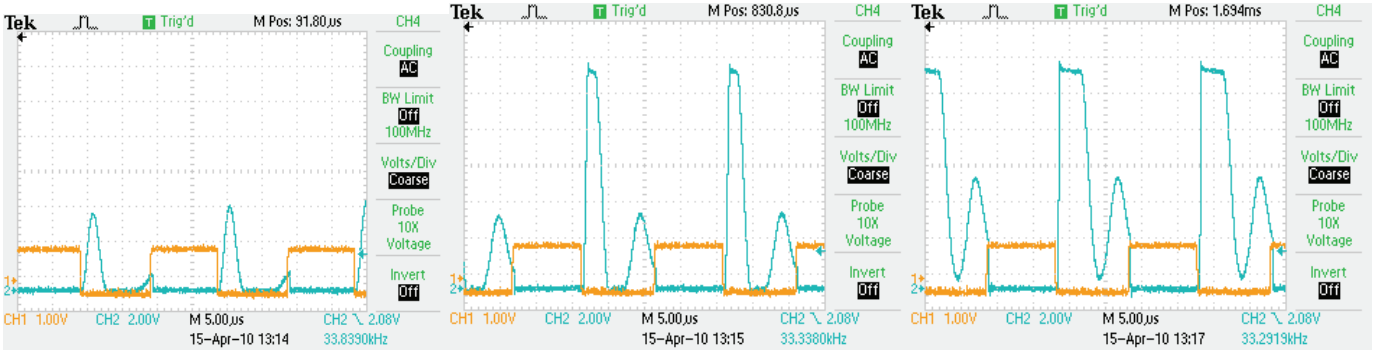


Fig. 5. Waveforms of the collector emitter voltage, and control voltage (gate-emitter) at the moments of 4ms and 5ms after zero crossing

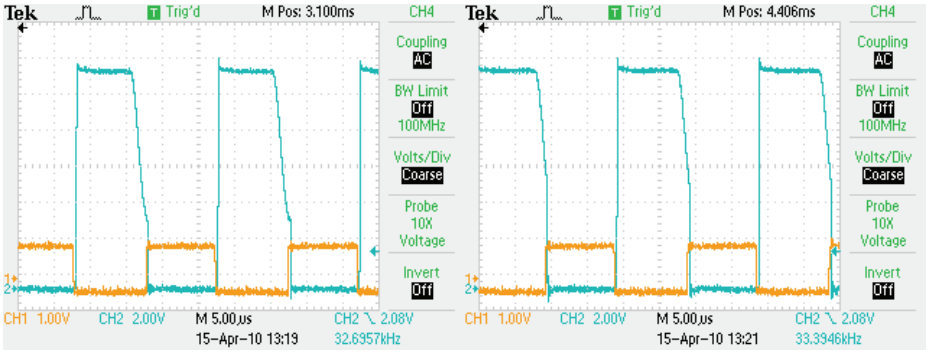


Fig. 6 Waveforms of the collector emitter voltage, and control voltage (gate-emitter) at the moments of 4ms and 5ms after zero crossing

Figure 4 shows the waveforms of the output versus the input voltage, where:

Channel 4 (green) is the input voltage

Channel 3 (purple) - output voltage

It can be seen that the output voltage has small ripple with gird frequency due to the sinewave component. This ripple however will not affect the proper operation of supplied driver circuits.

Figure 5 presents the Collector Emitter (CE) voltage on the transistor and its control pulses. The different waveforms are taken in different moments of the input sine-wave, as it follows 0,5ms, 2ms and 3ms after zero crossing of the input voltage. It can be seen that there is a certain transition period in the CE voltage of the transistor due to the discontinuous mode of operation of the transformer. This transition process can be additionally used to improve the circuit efficiency, obtaining soft switching of the transistor.

The waveforms of the Figure are:

Channel 2 (blue) is the CE voltage of the transistor

Channel 1 (yellow) - control pulses on the transistor

Figure 6 has the same depictions as figure 5 but for time periods 4ms and 6ms after the zero crossing of the input sinewave. It can be seen that for high voltages the transformer begins operating in continuous mode as the transition process are diminishing. Wave placement on the diagrams repeats that of figure 5.

Figure 7 presents the waveforms of the control pulses, the output voltage and the voltage on the primary winding of the transformer. The figure gives a visual explanation of the principle of operation of the suggested circuit.

The waveforms of the Figure are:

Channel 3 (purple) is the output voltage

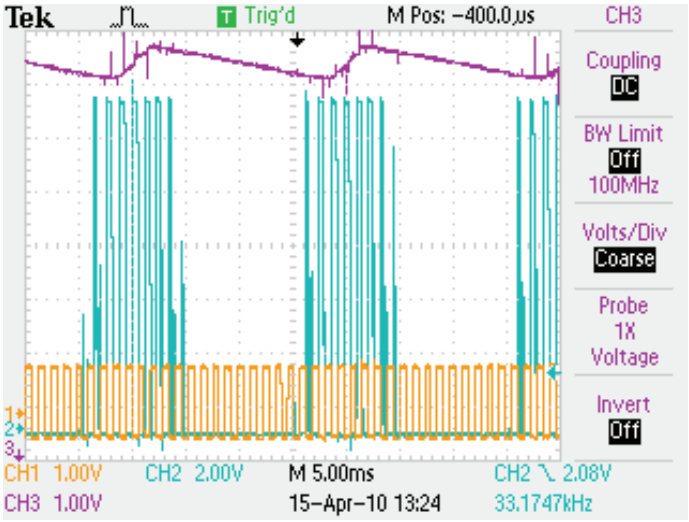


Fig. 7. Waveform of the positive half-wave on the primary winding of the transformer, the control pulses and the output voltage

Channel 2 (blue) – Voltage on the primary winding of the transformer

Channel 1 (yellow) - control pulses on the transistor

IV. CONCLUSION

The presented circuit provides notable advantages over the standard driver power supplies, as it offers reduced complexity which benefits size weight, and prize. The circuit was experimentally tested and validated. The results show possibilities that the circuit can be utilized in different applications and that its efficiency can be further improved.

ACKNOWLEDGEMENT

The paper is developed in the frames of the project 'Development of specialized scientific infrastructure for investigation of wind and solar energy potential', № DO02-48/10.12.2008, Ministry of Education, Youth and Science, Fund 'Scientific Research', Bulgaria.

REFERENCES

[1] Smith K. et. al., 'A Comparison of Voltage-Mode Soft-Switching Methods for PWM Converters', IEEE Transactions on Power Electronics, vol.12, No2, March 1997, pp.376-386.

[2] [Zhang J. M., et. Al., 'Comparison Study of Phase-Shifted Full Bridge ZVS Converters', 2004, 35th Annual IEEE PESC, Aachen, Germany, 2004, cd-rom.

[3] Vencislav Valchev et al, 'Design Considerations and Loss Anylasis of ZVS Boost Converter,' IEE Electric Power Applications, vol.148, No1, January, 2001, pp.29-33.

[4] Chang-Sun Kim, et. al., 'Alternately Zero Voltage Switched Forward, Flyback Multi Resonant Converter Topology', IEEE IES, Sevilla, November 5-8, 2002, cd-rom.

[5] Farrington R, F. Lee, M. Jovanovic, 'Constant frequency zero-voltage multi-resonant converters: topologies, analysis and experiments', IEEE PESC, 1991, pp. 33-40.

[6] Gradinarov N., N. Hinov, D. Arnaudov, "Analisys and Design of Resonant Inverters with Improved Output Characteristics, Zero-Current Switching', PCIM'03, Germany, 05.2003, pp. 423-427.

POSTER SESSION PO XIII

PO XIII - Renewable Energies

Preferential Tariffs and Administrative Procedures for Obtaining Licenses for Use of Electric Power From Photovoltaics

Miroslav V. Labudovic[1], Vladimir M. Labudovic[2] and Hristina M. Spasevska[3]

***Abstract* – Like any other investment, for installing a solar power plan first we should do a techno-economic analysis. By definition, techno-economic analysis compares all benefits and costs associated with some project during its working life. Costs include initial and annual recurring costs, while benefits include incomes from sale of energy. Depending on the results of the economic analysis of projects for solar power plants, the value of the economic indicators, the final decision for the investment should be done. The process of building of a solar power plant continues with obtaining all necessary permits: building permit and permission from the municipality where we are going to build a solar power plant; a license for performing energy activities, the final license for performing the activity and energy solution for preferential tariff; enrolment in a register, solution for preferential production; agreement with MEPSO for the purchase of produced electricity and power line from EVN. All these licenses, decisions and agreements are obtained in accordance with regulations prescribed by the entities which are authorized for this activity.**

***Keywords* – Photovoltaics, Tariff Book, Photovoltaic power plant.**

I. INTRODUCTION

By the end of 20th century, the United Nations Organization has started to pay greater attention to reduction of the emission of greenhouse gases in the atmosphere which are responsible for the greenhouse effect. Regarding the reduction of the greenhouse effect, in the last several years recommendations for increased use of renewable sources in the field of electricity production have been intensified. Solar energy as an inexhaustible source plays an important role in the production of electric power in many counties worldwide [1]. The need of electric power being increased, many new generations of photovoltaic panels started to appear, each with greater energy efficiency in order to gain the most of the space in which they would be installed.

Republic of Macedonia as an area with extremely favourable conditions [2,3] for the production of electric power from photovoltaics has adopted regulations for building a photovoltaic power plant, but unfortunately, an important capacity for production of electric power from photovoltaics is not installed yet.

In the last twenty years the production capacities in the developed countries show a constant growth regarding the production of electric power from photovoltaic systems. For development and realization of a photovoltaic power plant, a sound technical and economic analysis has to be made in order to check if it fulfils all the requirements. The technical and economic analysis, by definition, compares all the gains and costs related to a certain project while it last [4].

Costs include initial and annual recurring costs, while benefits include income from the sale of energy. The economic analysis of projects for solar power plants, usually provides the basis for a final decision whether to continue with such an investment - in the case that it is positive, i.e. gives positive economic results, or, if economic indicators are negative, to stop it [5].

Besides all the processes to build a solar power plant, it is necessary to obtain all necessary permits for its implementation by several entities, such as site conditions, building permit and use permit is obtained from the municipality of residence[1] (in accordance with the Law on Construction/ Law on Energy) where the solar power plant would be built. The licenses for carrying out energy activities, for performing energy activities and solving the preferential tariff are obtained from the Regulatory Commission of the Republic of Macedonia[6]. Enrolment in the Register and Approval/Decision for preferential manufacturer is obtained from the Energy Agency of Macedonia[7]. A contract with MEPSO is needed for purchase of the electric power produced, whereas EVN should approve a gain power line connection to the power network. All of these licenses, approvals and agreements are pursuant to the Law on Energy and all the Regulations and Rulebooks passed by the bodies in charge of this particular field.

II. PREFERENTIAL TARIFFS

Today in the world the current trend is the tendency for continuous energy saving and a particular attention is paid to environmental pollution. With it, the conventional fuel savings and maximum use of renewable sources of energy are imposed. In such a situation, intensifying of the development

[1]Miroslav V. Labudovic is with the Energoplan inzenering, Vasil Gjorgov, 1000 Skopje, Macedonia, E-mail: Labudovic_m@yahoo.com

[2]Vladimir M. Labudovic is with the Energo Art - Labudovik, Koco Racin br. 14-6/2, 1000 Skopje, Macedonia, E-mail: Labudovik@gmail.com

[3]Hristina M. Spasevska is with the Faculty of Electrical Engineering and Information Technologies, P.O. Box 574, 1000 Skopje, Macedonia, E-mail: hristina@feit.ukim.edu.mk

[1] According to the installed power capacity

of all types of energy with special place for photovoltaic plants is inevitable.
Given that the solar potential is inexhaustible source of energy and from a view of environmental protection it is the purest form, it is by chance subject to major investigations and scientific studies.
The most appropriate way to absorb the solar potential is the construction of photovoltaic plants. Their construction is of great importance for the economic exploitation of so far unexploited solar potential.
Tests are performed on various models of financing and construction of photovoltaic power plants in order to obtain better economic indicators.
Brought in 1997, The Kyoto Protocol recommended that each State Party, to advocate for reducing emission of carbon dioxide (CO_2) into the atmosphere, thus to slow down the process of global warming. Construction of plants of renewable sources for electric energy production is part of the entities that are prescribed with the protocol.
Construction of a plant for production of electric power from renewable energy sources is an expensive investment. Because of the relatively new technology and its rapid technological development that strives for greater efficiency, this investment is very expensive.

TABLE I
PRICES OF THE FIRST TARIFF BOOK

Group	Installed power (kW)	Preferential tariffs (euro cents/kWh)
I	to 50	45,00
II	51 - 1000	41,00

In order to attract investors in this field and stimulate their quick return on investment, the state provides price list for the sale of electricity. In Republic of Macedonia, the Regulatory Commission had issued a price list for purchase of electricity produced by a photovoltaic plant, but after a certain period of time, a new price list was issued by the same body.
The prices of the first Tariff book are given in the Table1.
In order to stimulate building of smaller solar power plants, the price of ransom kWh is greater. The price for a plant up to 50 kW differs from the price for photovoltaic plants in a way that the price is lower for the photovoltaic plants.

TABLE II
PRICES OF THE NEW TARIFF BOOK

Group	Installed power (kW)	Preferential tariffs (euro cents/kWh)
I	to 50	38,00
II	51 - 1000	34,00

Regardless of the location of the photovoltaic plant, on the ground or on a roof, the cost of the purchased energy is the same. The second Tariff book is with lower prices of approximately 17% for a sold watt, which may lead to investors hesitating because the period of return on investment will be 15% longer than the previous one.
The new Tariff book is given in Table 2.

III. ADMINISTRATIVE PROCEDURES FOR OBTAINING LICENSES FOR USE ELECTRIC POWER FROM PHOTOVOLTAICS

Construction of a photovoltaic power plant is far from easy and simple operation. After obtaining an approval for a construction of a photovoltaic power plant, the process needs to be stated by applying for licences, approvals, and permits in accordance with the laws and regulations in the country.
When selecting the most suitable location, i.e. finding the land where the plant would be built, it is needed to be determined in the Cadastre whether land is in private possession or it is owned by the state. Accordingly, in the case that the land is owned by the state, it is necessary to sign a concession agreement to use the land for some number of years. After the land is chosen, it is required urban plan to be made, to transform the land into a building land and an urbanistic decision to be adopted by the Municipality Council. After the completion of these processes and creation of a draft of the photovoltaic plant, a request for location conditions is filed in the municipality. The next step is application for building license with completed documentation for the plant. The Municipality Council is obliged to respond to the request in due period prescribed by the laws and regulations.
Application for issuing energy agreement is to be made to the company for power distribution EVN Macedonia. Energy approval is issued pursuant to statutory limit. Documents required for completing the documentation are in accordance with the conditions of EVN.
Licensing requirements for performing energy activities shall be submitted to the Regulatory Commission of the Republic of Macedonia. The required documents to be submitted are given in a Rulebook on the conditions, manner and procedure for issuing, changing, renewal and revocation of licenses for conducting energy activities, adopted at a meeting of the Regulatory Commission. The documentation should contain the energy approval provided by EVN and the draft of the photovoltaic plant. The Commission is obliged to provide the license in a period of time prescribed by the law.
The Energy Agency must be filed for registration in the registry under the policy for renewable electricity production and require proper documentation. The Agency is bound to issue the decision within the period of time prescribed by the Agency and in accordance with the regulations.
The Energy Agency is also responsible for issuing a decision on preferential producer of electric power, provided by the Rulebook on a preferential producer. The Agency is bound to issue this decision in a certain period of time, prescribed by the legal regulations.
The Energy Regulatory Commission of the Republic of Macedonia is the body where it should be applied for a decision on preferential tariff, prescribed according to the Rulebook for preferential tariff for photovoltaics.

MEPSO is a legal entity with which ends the procedure for deploying a production capacity of photovoltaics in the power system with signing an agreement for purchase of the produced electric power.

The period needed for obtaining the complete required documentation is 8 to 10 months, in case that all the due dates are taken into consideration.

The initiative for building photovoltaic power plants in Republic of Macedonia is in initial stage. The interest of investors to invest in energy in our country is growing in time. It is necessary to establish long-term strategy and program for intensifying the construction of photovoltaic power plants and, very importantly, creating conditions for attracting more investors.

In this regard, complete eradication of the factors that negatively affect the investment and construction of photovoltaic power plants is needed, including the following:

- Weak technological and economic documentation
- Lack of positive examples and lack of information
- Weak social policy for photovoltaic plant construction
- Weak legislation and passivity of the institutions in the issuance of the necessary documentation
- Bad signal sent to investors for reduction of the tariff rate.

IV. CONCLUSION

The exploitation of solar energy by building photovoltaic power plants is a topic that is very relevant in the Republic of Macedonia. During this period, applications for obtaining the necessary licenses and permits for the construction of photovoltaic plants are submitted. According to the data from relevant laboratories for solar irradiation, the regions around Ohrid and Bitola are most suitable for the construction of photovoltaic plants.

According to the analysis and studies carried out in the field of solar energy which concern investing in energy, the utilization of solar potential, and evolving process of construction of photovoltaic power plants in countries from Europe, including Republic of Macedonia.

Regarding the fact about the expected application of the Republic of Macedonia for membership into the European Union, in the near future it can be expected:

- Essential use and adaptation of the overall legislation, including the processes in the energy sector;
- Defining the conditions, as well as unknown elements associated with investing in photovoltaic plants, and for which important and necessary is the role and the assistance from the World Bank for Reconstruction and Development.

REFERENCES

[1] Solar Energy International, "Photovoltaics: Design and Installation Manual, , Ch.3, ISBN-13: 9780865715202, 2004.

[2] Real energy security is starting us in the face: renewable energy case studies South East Europe, http://bankwatch.org/documents/real_energy_security.pdf, p.p. 37, 2007.

[3] Report Linker, Photovoltaic profile of Macedonia, 2009.

[4] Bozin Donevski, Sustainable Energy Technologies: Energy Sector in Macedonia: Current Status and Plans, ISBN978-1-4020-6723-5, pp. 303-320

[5] USAID, Macedonia energy efficiency and renewable energy assessment, Final report, 2009,

[6] http://www.erc.org.mk

[7] http://www.ea.gov.mk

This page intentionally left blank.

Renewable Energy Technologies and Challenges for Their Applications

Aleksandar Lj. Malecic

***Abstract* – In this paper renewable energy technologies and challenges for their applications and popularization are considered. These challenges are observed from different aspects, both technological (design and control, lifecycle assessment and demand response and smart grids) and social (economy and popularization and social networks).**

***Keywords* – Renewable energy, wind turbines, economy, lifecycle assessment, social networks**

I. Introduction

Renewable energy sources (such as wind, solar and geothermal) have a role not just to supply with electrical energy, but also to replace existing non-renewable energy sources. There is a strong urge to reduce CO2 emissions and to prepare us for peak oil (some sources claim that we are about to face it [1]). Applications (or a lack of applications) of renewable energy technologies can and will reflect on climate change and society. We need the best solutions as soon as possible.

II. Design, control, and location

We need the best solutions out of different technologies and designs.

Wind as an energy source is highly unpredictable over time. There are different control strategies of wind turbines, such as with variable or fixed pitch and/or speed and collective or individual control of blades. Physical model should be accurate enough (no more or less) for design of good controllers. Control systems are "expected not merely to keep the turbine within its safe operating region but also to improve efficiency and quality of power conversion" [2].

Besides of different control strategies, to increase efficiency of wind turbines, one must consider different designs of blades and turbines [3], materials used [4], and location (onshore or offshore [5], close or remote from forests [6] etc.).

Solar energy [7] can be used for thermal conversion (collectors), photovoltaic conversion (solar cells) and combined (hybrid collectors). Different designs and materials have been examined, such as monocrystalline, policrystalline and amorphous materials for solar cells. The exposition to solar energy of a certain region (and energy capture) changes over a day and a year. Some solar cells (more in experiments than in a practical usage) have a built-in control mechanism that allows them to follow Sun and to adjust the angle of exposition.

Geothermal energy [8] comes from the natural generation of heat under the Earth's surface. The locations for geothermal energy plants are on the places with higher gradients of temperature in which drilling as shallower and less costly. The resources of geothermal energy are from hot water and molten rocks. It is primarily used for heating and cooling, but also for electric power generation. Plants are big and require high degrees of automation for extruding (production well) and returning (injection well) of material.

Aleksandar Lj. Malecic is with the Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mail: manoma@bisinter.net

III. Lifecycle assessment

While discussing about lifecycle assessment of renewable energy technologies, we must consider three kinds of factors: environmental, economic and social [9]. People obviously do not focus on renewable energy just to make profit. The problem is that we can't evaluate the true value of energy systems because we can't evaluate their effects on quality of life and long-term economic effects. Lifecycle assessment takes into account [9]:

- Product lifecycles and their total system-wide impacts
- "Cradle to Grave" (lifecycles of materials, how they fit into a bigger whole)
- Quantity
- It is data-intensive
- Standardization (ISO)
- Becoming global (we need the best solutions on the global level during lifecycles of wind turbines)

When compared with coal power plants [9], wind power plants are much better for environment. Their main shortcomings are materials and their lifecycles: eco toxicity (comparable with coal power plants) and a big demand for minerals.

When for example new wind turbines are being designed, their components should be traceable [10]. Product Lifecycle Management (PLM) is a strategic approach to the management of information relating to a product, from its definition, into manufacturing, and through to include maintenance. Through "a redefinition of the various manufacturing processes and better communication and integration between the related heterogeneous systems", a change in one part of the system can reflect on the system as a whole. A requirements change can change the process, the products and/or the enabling products.

IV. Demand response and smart grids

While a penetration of renewable power generation increases, one must consider both supply and demand. The demand response [11] deals with many uncertainties, but it could (with the improvement of wind power prognosis tools) enhance costs and availability of energy generated by wind turbines.

Cyber-controlled smart grids [12], if widely implemented, could help in the usage of renewable energy technologies. The goals of their implementation are:

- To develop technology for integration, control of renewable energy sources, control of energy consumption and load management.
- To empower energy user for a sustainable living and to develop Distributed Generation system where energy user is also an energy producer.

Also, engineers and researches must be aware of the challenges for implementation of the smart grids:

- Highly variable supply patterns… stochastic problem – particularly wind turbines and solar energy
- Efficient distributed algorithms required to process massive amounts of data for real time control
- Adaptive and self-healing control algorithms required for attaining high efficiency, reliability, and security of the large-scale distributed system
- Secure protocols, firewall mechanisms, intrusion prevention

V. Non-renewable vs. renewable energy – economic aspects

Generally, wind turbines are considered as the most promising source of renewable energy. Our planet receives enormous amounts of energy from Sun, but it seems that it will not be the leader, at least not in the near future. Different sources anticipate different penetrations, such as e.g. 29% globally by 2030 [13], 15% globally by 2030 [14] or 44% by 2020 in the United States of America [15] (Earth Policy Institute, mentioned in the next section).

If renewable energy succeeds to produce more electrical energy than nonrenewable resources, it will turn the global economy upside-down. It's obviously one of reasons why people hesitated to financially and politically support it in the past. New and emerging technologies and their innovative implementations can be an engine for economies in certain regions, especially in this interconnected and globalized world and its unpredictable dynamics with possibilities for rich to become richer and poor poorer very quickly [16].

On one hand, renewable energy resources still fight the competitive battle with short-term investments. They are more perspective in the long run and they can in not so distant future cause tectonic changes in the global economy. There could appear new "wind turbine sheiks" with economically strong (and probably politically influential) regions around them. Wind power plants are owned and by businesses and not by public utility companies [17].

Besides of suggestion to regulate nonrenewable energy resources through taxes and prices that reflect real economic effects of their usage [15], it is worth mentioning here another innovative economic mechanism. Kiwah [18] (shortened from kilowatt/hour) is a suggested monetary system (currency) that encourages sustainability and reductions of CO2 emissions. It could be exchanged for usual currencies and used directly only for investments in renewable and sustainable technologies and activities. Its creators have a goal to reduce our CO2 emissions with 80% in a short period of time" (before the year 2020).

VI. Popularization and online social networks

There are some initiatives on the internet that try to gather people working on sustainable development. I shall mention here some that look interesting to me: Earth Policy Institute [19], Repower America [20], Four Years. Go. [21], WiserEarth [22] and 2020 Climate Solutions Meshwork [23].

Earth Policy Institute is relatively well known between people interested in sustainability. Its director is Lester R. Brown [24], an internationally recognized author of books on global environmental issues. They have the book Plan B (right now it's the fourth edition, Plan B 4.0). According to this institute, we must radically change the way we produce energy before 2020. It means that we should globally lower our energy production through nonrenewable resources by 80 percent before the year 2020. Their ambitions are bringing us to another optimistic initiative, Repower America.

America has been the strongest world's economy for very long time. Its economic strength relies heavily on consumerism and consumption of energy and resources. When Al Gore, a former American vice-president and Nobel Prize laureate stands behind Repower America, it must be considered seriously. According to their website [20], the goal is a transition to a new clean energy future that will revitalize American economy, strengthen their national security, and solve the climate crisis.

It seems that something is indeed happening. But, according to State of the World Forum (and current dynamics) it's not even remotely enough. Perhaps all initiatives concerned about sustainability should make their activities more visible and to mobilize society as a whole. It brings us to online networks that are obviously influenced by Facebook, MySpace and similar social networks.

WiserEarth gathers people interested in activism. As one can see on its website, it is "a free online community space connecting the people, nonprofits and businesses working toward a just and sustainable world". It looks very similar to groups in Facebook. It's probably good in developing a sense of community. Its success depends as much on design and functionality as on members' ability to create something new. I can't see how the activities there could be attractive for people not involved in the groups and for the media.

The Four Years. Go. campaign started in March this year. Its goal is to gather different organizations and companies and through online collaboration to move the global society to sustainable path. There is already a significant number of organizations, but they should still start seriously work together and make their collaboration and its results visible.

I am active in 2020 Climate Solutions Meshwork. Their goal is the same as Earth Policy Institute's: to reduce nonrenewable energy production by 80 percent before 2020. It's moving disappointingly slowly at the moment, but, with probably some improvements in design and functionality (I have suggested some changes that would in my opinion be improvements), it looks promising, especially because everyone can contribute and make his/her contributions, regardless of name and working/political position. Also, it supports different types of activities: groups, discussions, blogs (content), pages and documents. types of activities: groups, discussions, blogs (content), pages and documents.

In my opinion these (or very similar) online activities should be considered seriously if we wanted to shift to renewable energy sources (with wind energy as probably their leader). It seems that, with the growth of social networking, we are about to reach a tipping point in which the thin line between virtual and material reality will be less and less visible. It seems that it has been predicted by Turchin (metasystem transition) [25] and Engelbart (augmented mind and experiments with hypertext as a collaborative tool) [26] and by Friedman ("The World Is Flat" - already a witness of something new emerging) [27]. These initiatives and social networks need success stories attractive for the media.

VII. Conclusion

In this paper renewable energy technologies and challenges for their usage (technology and popularization) are analyzed.

References

[1] R. Munroe, "Oil Supply Crunch: 2011-2015", Energy Bulletin, Post Carbon Institute, 2010.

[2] F. Bianchi, H. De Battista and R. Mantz, "Wind Turbine Control Systems", Springer-Verlag, London, 2007.

[3] P. B. Andersen, M. Gaunaa, C. Bak and Thomas Buhl, "Load Alleviation on Wind Turbine Blades using Variable Airfoil Geometry", EWEC 2006, Scientific Proceedings, pp. 22-25, 2006.

[4] A. Antoniou, T. Philippidis, "Optimizing Material Use in Blade Design by Improving Failure Prediction Methodology and Introducing Damage Tolerant Concepts in FRP Composites", EWEC 2006, Scientific Proceedings, pp. 79-83, 2006.

[5] P. Pinson, L.E.A. Christensen and H. Madsen, "Fluctuations of offshore wind generation - Statistical modeling", EWEC 2007, Scientific Proceedings, pp. 87-92, 2007.

[6] B. Dalpe and C. Masson, "Recommended Practices When Analyzing Wind Flow Near a Forest Edge with WaSP", EWEC 2007, Scientific Proceedings, pp. 54-58, 2007.

[7] T. Pavlovic, Fizika i tehnika solarne energetike", Gradjevinska knjiga, Beograd, 2007.

[8] J. Lund, "Characteristics, Development and Utilization of Geothermal Resources", Geo-Heat Centre Quarterly Bulletin, 28 (2), pp. 1-9, Oregon Institute of Technology, Klamath Falls, 2007.

[9] G. Norris, "Life Cycle Assessments of Wind Energy and Other Renewables"…, 2006.

[10] N. Moubayed, M. Bernard and M. H. El-Jamal, "Requirements Change Using Product Lifecycle Management for Wind Turbine Manufacturing Processes, Journal of ARISE, 2009.

[11] M. Klobasa and M. Ragwitz, "Demand Response –A new Option for Wind Integration", EWEC 2006, Scientific Proceedings, pp. 85-89, 2006.

[12] F. Khorrami, "Challenges in Cber-Controlled Smart Grid", Mechatronics/Green Research Laboratory, Politechnic Institute of NYU, Brooklyn, 2010.

[13] I. Paraschivoiu, O. Trifu and F. Saeed, "H-Darrieus Wind Turbine with Blade Pitch Control", International Journal of Rotating Machinery, Volume 2009, Article ID 505343, 2009.

[14] D. Barrie and D. Kirk-Davidoff, "Weather response to management of a large wind turbine array", Atmos. Chem. Phys. Discuss., 9, pp. 2917–2931, 2009.

[15] L. Brown, "Plan B 4.0", W. W. Norton & Company, London, New York, 2009.

[16] E. Reinert, "Globalna ekonomija", Cigoja, Beograd, 2006.

[17] N. Moubayed, A. El-Ali and R. Outbib, "Control Methods of a Wind Turbine", Journal of ARISE, 2009.

[18] http://www.qoin.com/en/kiwah/kiliwatt-hour-money.html

[19] www.earth-policy.org

[20] http://repoweramerica.org/why-clean-energy/

[21] www.fouryearsgo.org

[22] www.wiserearth.org

[23] http://2020.global.gaiaspace.org

[24] http:///en.wikipedia.org/wiki/Lester._R._Brown

[25] V. Turchin, "The Phenomenon of Science", Columbia University Press, New York, 1977.

[26] D. Engelbart, "Augmenting Human Intellect: A Conceptual Framework", Air Force Office of Scientific Research, 1962.

[27] T. Fridman, "Svet je ravan", Dan graf, Beograd, 2007.

This page intentionally left blank.

Influence of the Monthly Solar Radiation Variations to the Performances of the Hybrid Photovoltaic-Diesel Systems

Dimitar Dimitrov[1] and Atanas Iliev[2]

***Abstract* – Due to the stochastic nature of the solar radiation, the performances of the hybrid photovoltaic diesel systems vary. In this paper the solar radiation occurrence probability has been analyzed and modelled. The model is used for sensitivity analysis of a hybrid system performances. The variations affect the fuel consumption of the tested system, but the continuity of electricity supply is kept stabile.**

***Keywords* – Solar Radiation, Occurrence Probability, Hybrid Photovoltaic Systems, Sensitivity Analysis.**

I. INTRODUCTION

Despite the intensive growth of the installed number of grid-connected photovoltaic (PV) systems, enabled by various incentives, in the following decades, the niche market of PV systems will remain in the field of stand-alone applications. In such applications, for enhancing economic and technical performances, usually are combined more electricity sources (incl. diesel gensets, wind generators, etc.) and energy storage components (batteries, fuel cells, etc.), creating a hybrid structure. However, such hybridization requires employing complex system component sizing along with selecting a proper dispatching strategy [1].

Hybrid PV systems are primarily driven by the incident solar energy on the PV panels, which has regular and stochastic variations. For quantifying the regular fluctuations, one needs to know the solar geometry rules [2]. However, many research efforts have been done to explore the stochastic (statistical and sequential) rules of the measured solar radiation datasets [3-6]. The findings are applied to create so called *Typical Meteorological Years* (TMYs), which are yearly synthetic hourly solar radiation datasets, consisting of data that recover the same statistical and sequential properties as the real (measured) solar radiation data [7]. TMYs are suggested and usually used for simulation purposes within the sizing procedures of hybrid PV systems [1,8,9]. Still, this approach does not give a complete reflection of the system's operation. Namely, as the weather changes from year to year, even for the same locations in the same months, the incident solar energy values fluctuate around the monthly means. Generally, this may lead to unexpected system behavior, such as variation of o&m costs, but in some cases may even lead to disrupting the continuous electricity supply.

Most of the regular variations of the solar radiation are removed when introducing the variable called clearness index, defined as a ratio of the solar radiation on a horizontal plane and extraterrestrial solar radiation [3-6]. In [3] is stated that the long-term probability distribution, on a monthly level, of the daily clearness index, depends only on their monthly mean value, but not of the location or of the month.

The analyses of yearly fluctuations on a monthly level basis are given in [8, 9]. As a result, the probability density distributions of mean monthly daily clearness index yearly fluctuations are obtained. These distributions are fitted with Gaussian probability density functions. Additionally, dependence between the total mean value of the daily clearness index and the standard deviation is revealed. This allows calculating of annual sets of mean monthly clearness indexes that occur with a specified probability. Some of this theory is presented here.

These analyses can be used for calculation of monthly mean daily clearness index values, which for a given location, occur with a specified probability. With these sets as inputs, synthetic hourly solar radiation data series can be generated. In this paper, these solar radiation data series are used to estimate the fuel consumption of the given hybrid PV system.

II. SOLAR RADIATION DATA ANALYSIS

In [8] was analyzed the database of the World Radiation Data Center (WRDC) [10], which is consisted of measured and calculated, monthly, daily and hourly values of the solar radiation components, from 1964 up to 1993 for locations all around the world. There was considered daily solar radiation data from 44 meteorological stations from the Balkans Region. For each year and location the monthly mean daily clearness index was calculated by:

$$\overline{K}_t(m,y,l) = \frac{1}{J}\sum_{j=1}^{J} K_t(j,m,y,l) \qquad (1)$$

where $K_t(j, m, y, l)$ is the daily clearness index, $j = 1, 2, \ldots J$ is the number of the day in the month m, J is the total number of days in the month, $y = 1963, \ldots 1993$ is the year and $l = 1, 2, \ldots 44$, denotes different locations.

[1] Dimitar Dimitrov is Assistant Professor at the Faculty of Electrical Engineering and Information Technologies, University "Sts. Cyril and Methodius", Karpos 2 bb, 1000 Skopje, Macedonia, e-mail: ddimitar@feit.ukim.edu.mk

[2]Atanas Iliev is Associate Professor at the Faculty of Electrical Engineering and Information Technologies, University "Sts. Cyril and Methodius", Karpos 2 bb, 1000 Skopje, Macedonia, e-mail: ailiev@feit.ukim.edu.

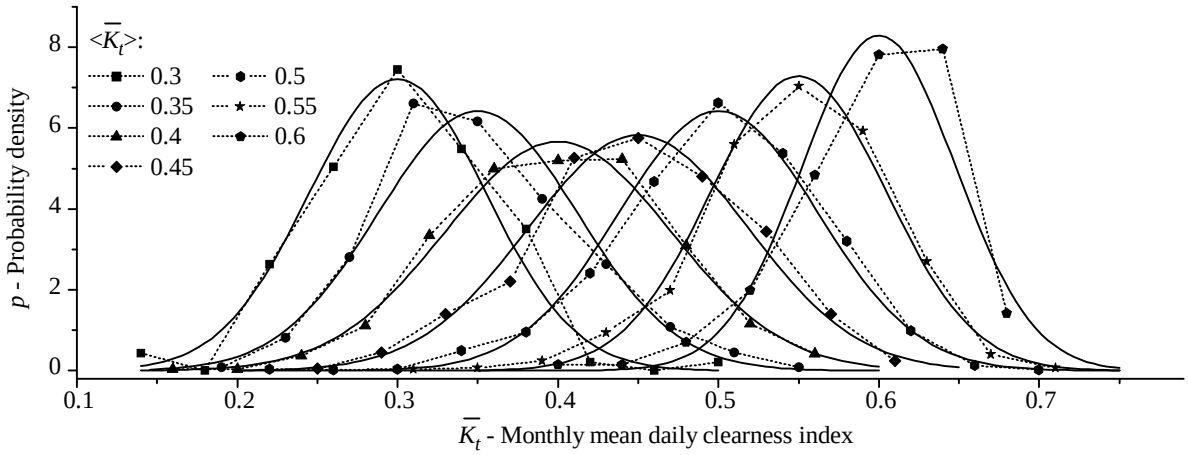


Fig. 1. Probability density distributions of K_t values and the fitting Gaussian functions

For each location *l*, the grand monthly mean daily clearness index, was calculated by:

$$<\overline{K}_t(m,l)>=\frac{1}{Y}\sum_{y=1964}^{1993}\overline{K}_t(m,y,l) \qquad (2)$$

where *Y* is the total number of years with having complete and regular $K_t(m, y, l)$. In the further text, for simplifying, $<K_t(m, l)>$ will be noted with $<K_t>$ and $K_t(m, y, l)$ with K_t.

Independent of location and month, sets of K_t values, having grand monthly mean daily clearness index $<K_t>$ that belongs to a narrow interval of ± 0.01 around the specified nominal clearness index, were grouped together. Because the intervals' length was 0.02, they overlap, thus one set of K_t values belongs to two neighbouring intervals.

The empirical standard deviation of all $K_t(i)$ belonging to the groups with a grand mean value $<K_t>$ was calculated by:

$$\sigma_{emp}=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[K_t(i)-<\overline{K}_t>\right]^2} \qquad (3)$$

where: $i = 1, ..., N$, *N* is the total number of $K_t(i)$ in a group.

For each group, the frequency distribution has been done. For all groups the bin length was $h = 0.04$. These distributions were then normalized, by dividing their values with the total number of the K_t-values in the groups. The normalized frequency distributions represent the occurrence probability $P(K_t, <K_t>)$ of K_t in a specified month for a given location, having grand monthly mean daily clearness index $<K_t>$.

In order to fit the empirical probability distributions with the theoretical probability distribution functions, it is more convenient to deal with the corresponding probability density distributions $p(K_t, <K_t>)$. Assuming that $p(K_t, <K_t>)$ is constant for the whole bin length *h*, it can be calculated with:

$$p\left(\overline{K}_t,<\overline{K}_t>\right)=\frac{P\left(\overline{K}_t,<\overline{K}_t>\right)}{h} \qquad (4)$$

In Fig. 1, with dotted lines, are shown 7 empirical probability density distributions with a range of $<K_t> = 0.30,.., 0.60$. As can be noted, the distributions are bell-shaped, which suggests a possibility to be fitted with the Gaussian probability density function. The theoretical probability density functions, when a random variable K_t, with mean $<K_t>$, has Gaussian distribution, are determined by:

$$p_{th}\left(\overline{K}_t,<\overline{K}_t>\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\overline{K}_t-<\overline{K}_t>}{\sigma}\right)^2} \qquad (5)$$

where σ is the standard deviation.

For fitting the empirical probability density distributions with a Gaussian probability density function, the widely used non-linear least square Levenberg-Marquardt method was employed. The standard deviation value σ_{LM} is chosen by finding a function given with Eq. (5), which optimally fits the empirical distributions. The fitting was done for the whole range of distributions, thus for each distribution, the parameter σ_{LM} was obtained. In Fig. 1, with solid lines, are also shown the Gaussian probability density functions that fit the corresponding empirical probability density distributions. Furthermore, it was approximated the dependence of the standard deviation to $<K_t>$, with the following expression:

$$\sigma_{fit}=-0.515<\overline{K}_t>^2+0.436<\overline{K}_t>-0.026 \qquad (6)$$

When the probability density function of a random variable is known, the probability can be obtained by integrating the function given in Eq. (6). In this case:

$$P\left(\overline{K}_t,<\overline{K}_t>\right)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\overline{K}_t}e^{-\frac{1}{2}\left(\frac{\overline{K}_t-<\overline{K}_t>}{\sigma}\right)^2}d(\overline{K}_t) \qquad (7)$$

Observing the measured data, all values of K_t belong to the range $0.05 < K_t < 0.8$. However, the significance of the above equation exceed this range from both sides. Still, the errors when using the Gaussian distribution in this case is the order of 10^{-7}, which is practically negligible.

The integral in Eq. (7) cannot be solved in exact form, and the solution is represented through error functions. On the other hand, in this paper, is required the inverse solution of the Eq. (7), i.e. the calculation of K_t for known $<K_t>$ and a specified probability $P(K_t, <K_t>)$. In order to easy the procedure of calculation of K_t, in this paper was not done

TABLE I

MONTHLY MEAN DAILY VALUES OF THE SOLAR RADIATION, CLEARNESS INDEX AND AIR TEMPERATURE

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
H [kWh/m^{2}day]	1.76	2.59	3.62	4.90	5.97	6.50	6.65	6.04	4.62	3.40	2.15	1.47
$<K_t>$	0.449	0.473	0.489	0.516	0.544	0.560	0.586	0.596	0.560	0.552	0.495	0.422
T_a [°C]	0.1	2.6	5.7	11.2	16.0	19.1	21.0	21.3	17.4	12.1	7.5	2.2

TABLE II

MONTHLY MEAN DAILY CLEARNESS INDEX WITH A SPECIFIED PROBABILITY

Month	$<K_t>$	σ_{fit}	Probability												
			0.001	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.999**
Jan	**0.449**	**0.066**	0.653	0.557	0.534	0.504	0.484	0.466	**0.449**	0.432	0.414	0.394	0.364	0.341	0.245
Feb	**0.473**	**0.065**	0.674	0.580	0.556	0.528	0.507	0.489	**0.473**	0.457	0.439	0.418	0.390	0.366	0.272
Mar	**0.489**	**0.064**	0.687	0.594	0.571	0.543	0.523	0.505	**0.489**	0.473	0.455	0.435	0.407	0.384	0.291
Apr	**0.516**	**0.062**	0.707	0.618	0.595	0.568	0.548	0.532	**0.516**	0.500	0.484	0.464	0.437	0.414	0.325
May	**0.544**	**0.059**	0.726	0.641	0.619	0.593	0.575	0.559	**0.544**	0.529	0.513	0.495	0.469	0.447	0.362
Jun	**0.560**	**0.057**	0.735	0.653	0.633	0.608	0.590	0.574	**0.560**	0.546	0.530	0.512	0.487	0.467	0.385
Jul	**0.586**	**0.053**	0.749	0.673	0.653	0.630	0.614	0.599	**0.586**	0.573	0.558	0.542	0.519	0.499	0.423
Aug	**0.596**	**0.051**	0.753	0.680	0.661	0.639	0.623	0.609	**0.596**	0.583	0.569	0.553	0.531	0.512	0.439
Sep	**0.560**	**0.057**	0.735	0.653	0.633	0.608	0.590	0.574	**0.560**	0.546	0.530	0.512	0.487	0.467	0.385
Oct	**0.552**	**0.058**	0.730	0.647	0.626	0.601	0.582	0.567	**0.552**	0.537	0.522	0.503	0.478	0.457	0.374
Nov	**0.495**	**0.064**	0.692	0.600	0.577	0.549	0.528	0.511	**0.495**	0.479	0.462	0.441	0.413	0.390	0.298
Dec	**0.422**	**0.066**	0.627	0.531	0.507	0.478	0.457	0.439	**0.422**	0.405	0.387	0.366	0.337	0.313	0.217

deeper analysis of the error functions. Instead, there was used the build-in function NORMINV from the Microsoft Excel®. It is a three-parameter (probability, $<K_t>$ and standard deviation) function, which, as a result, gives the inverse solution of Eq. (7), i.e. K_t. This function applies numerical solving and some errors in the results are possible. The methodology has been validated for several locations from the Balkan region.

III. INPUTS AND SENSITIVITY ANALYSIS

The shown methodology was used to estimate the sensitivity of a hybrid PV system which was planned to be installed in a small rural village Vrbica in the R. of Macedonia (42.13°N, 22.25°E, altitude 800 m). For this location monthly mean global solar radiation values on a horizontal plane and the ambient temperature were obtained from the nearest meteorological stations (Table I). These values were used to calculate the mean monthly values of the daily clearness index. Since they are a result of long-term observations, we considered them as grand monthly mean daily clearness indexes $<K_t>$. From those, we calculated the monthly mean of daily clearness index values K_t that appear with a specified occurrence probability, as previously shown (Table II). Then we calculated the mean monthly values for the global solar radiation on a horizontal plane. Further, using the software tool METEONORM [7], we generated the monthly sets of hourly solar radiation data, and they occur with a specified probability. On Fig. 1, for each month are given the solar radiation vs. probability diagrams. The diagrams can be read as follows: for example for September, with probability of 20 %, the global solar radiation on a horizontal plane is *equal or less than* 5.01 kWh/m^{2}day, or with 90 % probability – 4.02 kWh/m^{2}day. To note, that the values of the solar radiation given in Table I can be obtained for probability of 50 %.

The analyzed system is optimally sized to continuously supply several households in the mentioned rural area with the total average electricity consumption of 10 kWh/day. The optimization was done by using the techniques of genetic algorithms, and is elaborated in detail in [9, 11]. The system is consisted of PV generator – 1.68 kWp (fixed tilt angle of 40°), battery capacity – 13.2 kWh, diesel genset –2 kVA, inverter – 1.5 kVA. It uses a setpoint (SOC_{ON} = 0.398; SOC_{OFF} = 0.831) frugal (P_d = 817 VA) dispatching strategy [1, 9].

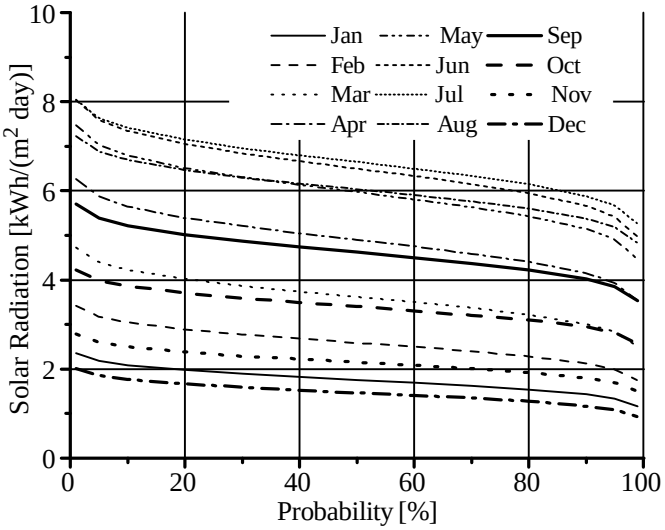


Fig. 1. Occurrence probability of the solar radiation

A year to year variations (for the same months) of the solar radiation, result with variations of the PV generator energy output. To ensure a continuous electricity supply, in case of shortage of PV energy, it is compensated by a prolonged diesel genset operation, i.e. with increased fuel consumption, and vice versa. The sensitivity of the fuel consumption in regard to different solar radiation input is obtained by a hybrid PV system simulation, using a software tool based on PVFORM [12], to which are implemented many modifications [9].

On Fig. 2, is given the PV generator output energy vs. its occurrence probability. Fig. 3 shows the dependence of the fuel consumption in regard to the occurrence probability of the solar radiation. The diagrams can be read on a following way: for example for December, when the mean daily solar radiation on a horizontal plane is 1.47 kWh/m^{2}day (for probability of 50 %), the total PV generator output energy equals to 124.6 kWh, while the fuel consumption is 85 liters. In 90 % of the time, the average value of the solar radiation is more or equal to 1.17 kWh/m^{2}day (20.4 % less than the average) and the PV output energy is more or equal to 90.2 kWh (27.6 % less) and the fuel consumption is at least 95 liters (11.8 % more than the average). In 10 % of the time, the solar radiation is more or equal to 1.77 kWh/m^{2}day (20.4 more than the average), the PV generator output is 158.4 kWh (27.12 % more) while the fuel consumption is at least 76 liters (10.6 less). In all considerations, no shortage of supply has been identified.

IV. Conclusion

This paper deals is presented the methodology for estimation of the solar radiation occurrence probability. Starting from the basic measurements, i.e. average monthly value of the solar radiation, the methodology enables to create hourly datasets of solar radiation that occur with a specified probability.

The variation of the solar energy has impact to the operation of the hybrid PV system. Lower values result with increased fuel consumption and vice versa. Hence, the methodology enables to quantify the expenditures and operation and maintenance costs and relate them to probability.

However, the proposed sensitivity analysis can give the answer if the designed hybrid PV system is able to supply the consumers without interruption.

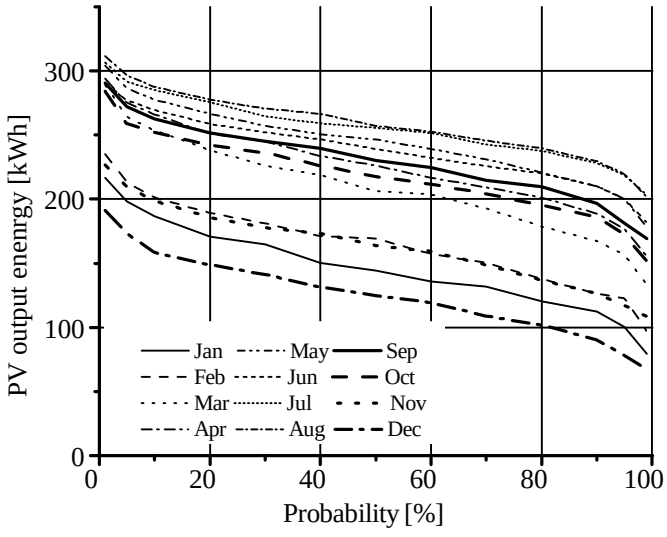


Fig. 2. Occurrence probability of the PV generator output energy

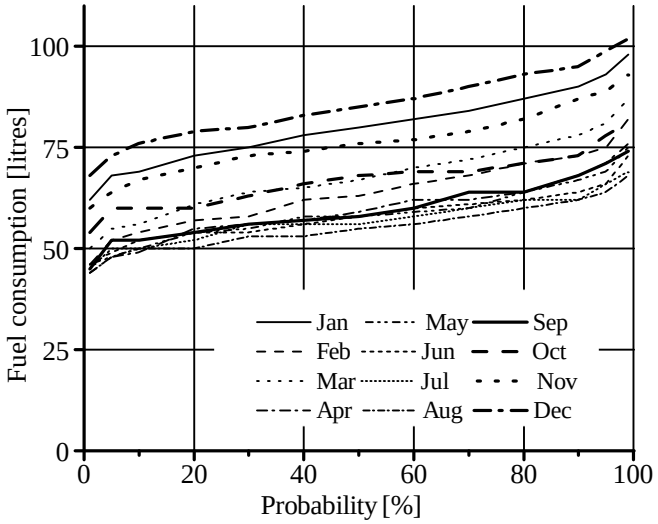


Fig. 3. Probability of the fuel consumption

References

[1] M. Ashari, C. V. Nayar, "An Optimum Dispatch Strategy Using Set Points for a Photovoltaic (PV)–Diesel–Battery Hybrid Power System", Solar Energy, Vol. 66, No. 1, pp. 1-9, 1999.

[2] C. Honsberg, S. Bowden, http://pvcdrom.pveducation.org/.

[3] B.Y.H. Liu, R.C. Jordan, "The interrelationship and characteristic distribution of direct, diffuse and total solar radiation", Solar Energy, Vol. 4, No. 1, pp. 1-19., 1960.

[4] P. Bendt, et. al., "The frequency distribution of daily insolation values", Solar Energy, Vol. 27, No. 1, pp. 1-5, 1981.

[5] R. Aguiar, et al., "Simple Procedure for Generating Sequences of Daily Radiation Values Using a Library of Markov Transition Matrices", Solar Energy, Vol. 40, No. 3, pp. 269-279, 1988.

[6] J.M. Gordon, T.A. Reddy, "Time Series Analysis of Daily Horizontal Solar Radiation", Solar Energy, Vol. 41, No. 3, pp. 215-226, 1988.

[7] METEONORM, http://www.meteonorm.com/.

[8] D. Dimitrov, "Occurrence Probability of Daily Solar Radiation Data Series, with Application to Hybrid Photovoltaic Systems", MEDPOWER'04, Conference Proceedings CD-ROM, ISBN 9963-8275-2-7, Lemesos, Cyprus, 2004.

[9] D. Dimitrov, "Contribution to the Optimization of Hybrid Photovoltaic Systems", PhD thesis, USCM-FEIT, Skopje, 2009.

[10] World Radiation Data Centre, http://wrdc-mgo.nrel.gov/.

[11] D. Dimitrov, Z. Andonov, "Determination of Genetic Parameters for Optimization of the Hybrid PV-Diesel Systems with Genetic Algorithms", MEDPOWER'08, Conference Proceedings CD-ROM, MED08-152, Solun, Greece, 2008.

[12] D.F Menicucci., J.P. Fernandez, "User's Manual for PVFORM", Report #SAND85-0376-UC-276, Sandia Natl. Labs, Albuquerque, NM, 1988.

Optimal Distribution of PV Modules According to Dynamic Optimization Methods

Borislav Hr. Dimitrov[1], Hristofor P. Tahrilov[2]

***Abstract* – The article examines the possibility of an optimal location of photovoltaic modules of a photovoltaic power plant depending on the characteristics of the terrain. It is accented the possibility for determination of the maximum filling of a given territory, using various modules and inverters, but with different strings and that the number of modules in them. It is proposed to use the algorithm for dynamic optimization, which resizes a photovoltaic area. The target function is the maximum filling of the terrain, and with it a maximum installed capacity.**

***Keywords* – PV system, dynamic optimization**

I. INTRODUCTION

One of the main stages in the design of photovoltaic plants is the placement of photovoltaic modules in the chosen field. Typically, the specialized software [5], [6] offers calculating the inverter, depending on the required electrical parameters of the construction system. Some of the application software offer desktop environments to determine the location of the photovoltaic modules, analysis of shading, etc. However, the designer is responsible for the desired location according to terrain's characteristics. Usual practice is to experiment with different possibilities. This approach is laborious and time-consuming and the decision may not be optimal.

The paper examines the possibility to solve above described task by using validated algorithms for dynamic optimization. The resolve of the problem would increase the effectiveness of the design work and to improve the optimal use of terrain.

Use of dynamic optimization according to [1] requires:

• Optimal substructures of the decision – the decision of optimization task can be found as a function of optimal solutions of the subtasks.

• Overlapping subtasks - subtasks is calculated only once, thereby limiting the actual number of subtasks.

Obviously the application of dynamic programming problem in question is subject to analysis.

II. ANALYSIS

The location of the modules over the terrain can be done differently. The decision of the task concerning the optimal placement of photovoltaic modules on the selected area must consider:

• *Determination of the terrain* - it may be with an irregular geometric shape and / or divided into individual parcels. This complicates finding the optimal placement of the strings in order to best fill. In addition the requirement for insulation distances, paths, cable paths and others.

• *Calculating the inverter* - the ability to use several inverters with different parameters, which defines a different configuration of the strings. Thus needed optimal configuration for optimal filling can be selected.

• *Defining the necessary photovoltaic modules* - they may have different electrical and geometrical characteristics and they are grouped to a separate inverter. Thus the length of the strings can be changed. The ability to work with different constructions is particularly important when the area is with irregularly shaped parcels.

• *Defining the parameters of the strings* – the inverter allows a certain tolerance of the input variables, as they can define strings with different parameters.

Those features require reporting of possible combinations when working with different inverters and permitted parameters of strings.

Fig.1 shows the design of a photovoltaic power plant, which uses the field of irregular shape. In order to optimal fill a territory it is parceled of the parts, in which the inverters and the strings with different parameters can be used. It is assumed that the parts are not predetermined and can be defined by the designer. In this sense, work environment can be considered as a variable.

Table 1 contains the parameters of used inverters. The number of PV in the string is determined by the tolerance in the U_{DC}, which allows working with different lengths. Thus a set of configurations applicable to different parts of the terrain is received.

An exemplary realization is made with Polycrystalline Framework modules. Sizing the model system was made according to validated literature sources [2], [3], [4] and aims to propose an experimental study of the applicability of the proposed calculation algorithms. Calculations are in the following sequence:

[1]Assist Prof. PhD. Borislav Dimitrov – Department of "Electrical Engineering and Electrical Technologies" , TU- Varna, e-mail: bdimitrov@processmodeling.org

[2]доAssoc. Prof. PhD Hristofor Tahrilov – Department of "Electrical Engineering and Electrical Technologies", TU- Varna, e-mail: h.tahrilov@gmail.com

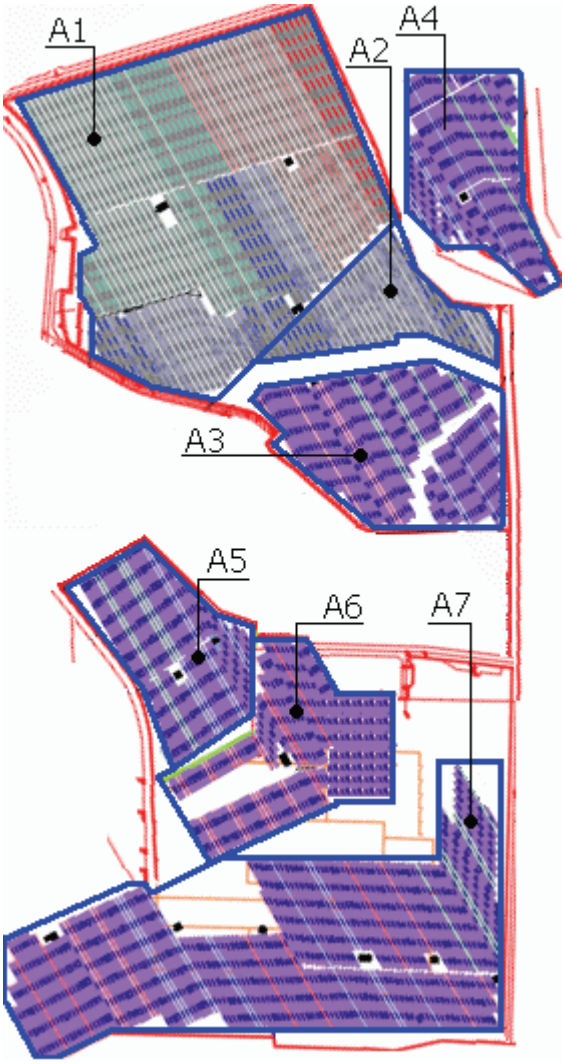


Fig. 1. PV plant's project
A1÷A7 – Area of the whole terrain, divided into separate parcels

Minimum nominal (working) voltage:

$$U_{min(\tau T)} = U_{mpp(stc)} + \left(T_u.\left(\tau_T - \tau_{STC}\right)\right) \qquad (1)$$

$\boldsymbol{\tau_T}$– temperature for calculations (module's temperature);
$\boldsymbol{\tau_{stc}}$ – temperature of the module under standard testing conditions (25°C)
$\boldsymbol{U_{mpp(stc)}}$ – voltage at maximum power for standard testing conditions;
$\boldsymbol{T_u}$– temperature coefficient of the voltage;

For chosen modules at $\boldsymbol{U_{mpp(stc)}}$ = 35,5V, $\boldsymbol{T_u}$= –0,1665 V/C° and maximum temperature $\boldsymbol{\tau_T}$=70°C, the minimum voltage is:

$$U_{min\,(70)} = 35{,}5 + (-0{,}1665.(70 - 25)) = 28{,}0075\text{V}$$

The maximum nominal voltage:

$$U_{max(\tau T)} = U_{mpp(stc)} + \left(T_u.\left(\tau_T - \tau_{STC}\right)\right) \qquad (2)$$

For chosen modules at minimum temperature $\boldsymbol{\tau_T}$ = -15°C:

$$U_{min\,(-15)} = 35{,}5 + (-0{,}1665.(-15 - 25)) = 42{,}16\text{V}$$

The voltage of the open circuit:

$$U_{0max(\tau T)} = U_{0(stc)} + \left(T_u.\left(\tau_T - \tau_{STC}\right)\right) \qquad (3)$$

TABLE I
PARAMETERS OF THE INVERTERS SOLARMAX

Инвер-тор	P_{DCmax} [KW]	U_{DC}	U_{DCmax} [V]	I_{DC} [A]	P_{ACnom} [KW]	P_{ACmax} [KW]	I_{AC} [A]	U_{AC}, f[Hz]
SolarMax 20C	24	430 ÷ 800;	900	0÷48	20	22	0÷31	3X 400 50Hz
SolarMax 35C	45			0÷78	35	38,5	0÷45	
SolarMax 50C	66			0÷120	50	55	0÷77	
SolarMax 80C	105			0÷180	80	88	0÷122	
SolarMax 20S	24	400 ÷ 800	900	0÷48	20	22	0÷31	3X 400
SolarMax 35S	45			0÷78	35	38,5	0÷54	

TABLE II
CALCULATING VOLTAGES OF POSSIBLE STRINGS.

	Limit according Table 1	**One module**	**In string**		
			15	**16**	**17**
$U_{min\ mpp}$ [V]	430	28	420,1	448,1	476,1
$U_{max\ mpp}$ [V]	800	42,2	632,4	674,6	716,7
$U_{max\ oc}$ [V]	900	51,7	774,9	826,6	878,2
Applicable inverters (Table 1)			SolarMax 20S SolarMax 35S	all	all

TABLE III

MPOWER AND PERMITED NOMBERS OF THE STRINGS OF A INVERTER

Number of modules	P_{DC} [KW] AT NUMBER OF STRINGS																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	8,4	12,6	16,8	21	25,2	29,4	33,6	37,8	42	46,2	50,4	54,6	58,8	63	67,2	71,4	75,6	79,8	84
16	8,96	13,44	17,92	22,4	26,9	31,4	38,8	40,3	44,8	49,3	53,8	58,2	62,7	67,2	71,7	76,2	80,6	85,1	89,6
17	9,52	14,28	19,04	23,8	28,5	33,3	38,1	42,8	47,6	52,4	57,1	61,9	66,6	71,4	76,2	80,9	85,7	90,4	95,2
Inverter	SolarMax20C				-														
	SolarMax35C								-										
	SolarMax50C													-					
	SolarMax80C																		
	SolarMax20S				-														
	SolarMax35S								-										

$U_{0(stc)}$ – idling voltage.

For chosen modules $U_{0(stc)}$ = 45V:

$$U_{\min(-15)} = 45 + (-0{,}1665.(-15 - 25)) = 51{,}66V$$

Maximum nominal current

$$I_{\max(\tau T)} = I_{\text{mpp(stc)}} + \left(T_I.\left(\tau_T - \tau_{STC}\right)\right) \qquad (4)$$

$\mathbf{I_{mpp(stc)}}$ – nominal current at standard testing conditions;

$\mathbf{T_I}$ – temperature coefficient of the current;

For chosen modules $\mathbf{I_{mpp(stc)}}$=7,89A, T_I=0,00501A/C°:

$$I_{\max(70)} = 7{,}89 + (0{,}00501.(70 - 25)) = 8{,}115A$$

Geometrical parameters of selected modules are: 1.9 x 1 m. Parameters of possible strings (with 15, 16 and 17 modules) are shown in Table 2. Whereas catalog PV power P (stc) = 0,28 kW and the resulting current of 8,1 A, the power of the three configurations of strings can be determined: P (15 modules) = 4,2 kW, P (16 modules) = 4, 48kW, P (17 modules) = 4,76 kW. The allowable number of strings to each inverter is shown in the table 3. It is obtained: SolarMax20S up to 5 strings, String SolarMax35S up to 9, etc.

The task can be reduced to the following:

A peak installed power of a photovoltaic power plant being built on a terrain with an irregular shape. In order to optimal sizing, the terrain is divided into individual parcels, which have capacity (M) of a number of modules (N), respectively certain maximum power. It is necessary to find the optimum amount of strings that can be placed on the terrain, taking into account the data from Table 1, 2,3. Put parameters are strings, numbers (m) and total power (r) (Table 3). Thus the optimal combination of inverters, working with allowable number of strings and which can fill most the terrain, need to be determined.

The task is reduced to finding the maximum amount [1]:

$$\sum_{i=1}^{N} x_i c_i \qquad (5)$$

At the limiting conditions:

$$\sum_{i=1}^{N} x_i m_i \le M$$
$$c_i > 0, m_i > 0, x_i \in \{0,1\}, i = 1,2...,n \qquad (6)$$

The limitations over $\boldsymbol{c_i}$ and $\boldsymbol{m_i}$ are connected to the condition of the task; $\boldsymbol{x_i}$ is the restriction relates to the sought decision.

It is necessary to define a recurrent target function, giving the decision to parcel with a capacity i.

$$F(i) = \begin{cases} 0 & i = 0 \\ \max[c_j + F(i - m_j)] & i > 0 \\ j = 1,2,...,N; m_j \le i & \end{cases} \qquad (7)$$

Iterative solution is obtained by successively calculating the target function. A realization is shown in listing 1. In the case of proposed PV system is sized using the "task of the backpack [1]:

Listing 1:

```
void opt_photovoltaic (void) {
for (i = 1; i <= M; i++) {
/* Find the maximum value of the target function M - Displacement
parcel  */
  maxV = maxI = 0;
/*maxV – maximum achieved value */
/*maxI – index at which it is achieved */
  for (j = 1; j <= N; j++) {       /*N – strings */
    if (m[j] <= i && !(set[i - m[j]][j >> 3] & (1 << (j & 7))))
/*m[]–array of number of modules in the string according to Table 3
*/
    if (c[j] + Ph[i - m[j]] > maxV) {
/*Ph – target function */
    maxV = c[j] + Ph[i - m[j]];
    maxI = j;
    }
  }
  if (maxI > 0) {/* Check for lower index m */
```

```
    Ph[i] = maxV;
    memcpy(set[i], set[i - m[maxI]], (N >> 3) + 1);
/*set[] – array with elements satisfying the target function */
    set[i][maxI >> 3] |= 1 << (maxI & 7);
  }
    if (Ph[i] < Ph[i - 1]) {
/* The plot fits all strings */
    Ph[i] = Ph[i - 1];
    memcpy(set[i], set[i - 1], (N >> 3) + 1);
  }
}
}
```

The use of recursive function is proposed in the listing 2 [1]:

Listing 2:

```
void Photovoltaic (unsigned k){  /* Recursive functions */
  unsigned i, bestI, fnBest, fnCur;
/* Initialization of variables * /
/ * Calculates the largest value */
  for (bestI = fnBest = 0, i = 1; i <= N; i++) {
    if (k >= m[i]) {
    if (if_not_calculated == Ph[k - m[i]]) Photovoltaic(k - m[i]);
/*if_not_calculated – value -1 if the target function Ph is not calculated*/
    if (!set[k - m[i]][i])
      fnCur = c[i] + Ph[k - m[i]];
    else
      fnCur = 0;
      if (fnCur > fnBest) {
      bestI = i;
      fnBest = fnCur;
     }
    }
  } /*registering the biggest value of the function*/
  Ph[k] = fnBest;
  if (bestI > 0) {
  memcpy(set[k], set[k - m[bestI]], N);
  set[k][bestI] = 1;
  }
}
void opt_photovoltaic (void) {
/*Calculating the value of the function */
  unsigned i, sumM;   /* Initialization of variables */
  memset(set, 0, sizeof(set));
/* Initialization of the multitudes of possible configurations */
  for (i=0; i<=M; i++)
/* Initialization of the values of the target function */
    Fn[i]= if_not_calculated;
  for(sumM=m[1], i=2; i<=N; i++)
/* Check for all configurations */
    sumM += m[i];
  if (M >= sumM) {
    printf("\n The field can accommodate all configurations ");
    return;
  }
  else {
    Photovoltaic (M);
/* computation of recursive function */
    printf("\n configuration number:\n");
    for (i = 1; i <= N; i++)
    if (set[M][i])
      printf("%5u", i);
    printf("\n Maximum value: %u", Ph[M]);
  }
}
```

Using the proposed algorithm gives the following results:

• parcel A1 (Fig. 1) - SolarMax80S inverter, PV system 20 string about 17 modules;
• A2 - SolarMax35S, 9 string of 16 modules;
• A3 - SolarMax35S, 9 string of 16 modules, SolarMax20, 5 string about 17 modules;
• A4 - SolarMax35S, 9 string about 17 modules;
• A5 - SolarMax35S, 9 string about 17 modules;
• A6 - SolarMax20S, 5 string about 17 modules, SolarMax20, 5 string of 16 modules;
• A7 - SolarMax50S, 14 string of 15 modules, SolarMax35, 8 string about 17 modules;

For given example at this distribution optimal fill the field in figure 1 is obtained. As seen from the decision in parcels A3, A6, A7 two different inverter are used with different numbers of strings and components in them. This was imposed by the typical geometric features of the parcel. Thus, the maximum duty cycle and the maximum installed capacity is determined by the necessary combination of strings. There may be other solutions that do not meet the target function:

- In A3 parcel instead of two inverter SolarMax35S (9 string of 16 modules) and SolarMax20 (5 string about 17 modules), using one SolarMax50;
- In A6 parcel instead of two inverter SolarMax20S (5 string about 17 modules) and SolarMax20 (5 string of 16 modules), using one SolarMax35;
- In the Land A7 - instead of two inverter SolarMax50S (14 string of 15 modules) and SolarMax35 (8 string about 17 modules) using one SolarMax80;

III. CONCLUSIONS

Dynamic optimization methods are applicable in the design work of photovoltaic plants. The algorithms in the proposed listings provide a decision on the location of the modules for optimal use of terrain. Thus the eligible configurations inverter-number of strings- number of modules that they are determined. The proposed approach allows, by decision of the task set out to determine the maximum possible installed capacity of a given territory. Multivariate decision, limited by the target function, creates options for the type of each element with additional requirements and patterns.

REFERENCES

[1] Nakov P., P. Dobrikov. “Programming =++ Algorithms” Sofia 2003.
[2] Christopher L, Photovoltaics. Design and installation manual. Solr Energy international 2004.
[3] Messenger R. Photovoltaic Systems Engineering. Washingon DC. 2003
[4] Brooks Engineering, Advanced PV System. Design and Installation. 2008
[5] http://www.pvsyst.com/
[6] http://solarmax.com/

Autonomous PV Heating System with Foil Heaters

Borislav Hr. Dimitrov[1], Hristofor P. Tahrilov[2]

Abstract – This paper analyzes the possibility of create an autonomous PV heating system from foil elements with wide surface. It is proposed that the implementation is with a direct connection between the source and the load. The paper includes analysis of the technical characteristics of the installation, designing foil heater according to the source parameters and experimental study. The last option is realized by selected characteristics of the heaters.

The proposed system photovoltaic source- foil heater does not use standard inverters, which enables the construction of low-cost solutions.

***Keywords* – Foil heaters, PV systems, Heating system**

I. INTRODUCTION

The proposed photovoltaic folic-heating system aims to realize the following advantages:

• *It is possible that the system be sized for direct connection of photovoltaic strings with foil heaters. This is an opportunity to not use an inverter, which is usually expensive. Possibility of obtaining low-cost solutions is essential for autonomous household photovoltaic systems in roof structures, small drying ovens, etc.*

• *The folic heaters are low heating elements that can be designed for any supply voltage range. This allows the implementation of various schemes with series and parallel connection of modules that meet specific performance requirements. The dimension is done in certain parameters of the nominal power system.*

• *The system is maximum simple and reliable. The additional items tof the installation can be a voltage stabilizer, UPS system etc.*

• *The foil elements with wide surface allow obtaining uniform distribution of the temperature field in the room. This is essential for certain types of technological equipment such as dryers, etc. The above mentioned condition can not be realized through the widespread use of spiral tubular or other heaters and domestic heating appliances standard [1], [2], [3].*

II. ANALYSES

The principal electric scheme of the photovoltaic system with foil heaters is shown in Fig.1.

[1]Assist Prof. PhD. Borislav Dimitrov – Department of "Electrical Engineering and Electrical Technologies", TU- Varna, e-mail: bdimitrov@processmodeling.org

[2]Assoc. Prof. PhD Hristofor Tahrilov – Department of "Electrical Engineering and Electrical Technologies", TU- Varna, e-mail: h.tahrilov@gmail.com

Design [1], [2], [3] of the foil heater is preceded by preliminary heat technical calculations, related with the determination of the geometric dimensions of the heat transfer surface of the heating element. As a result geometric dimensions of the equivalent smooth surface, on which the heater is placed, are defined. When designing the foil heater, the following restrictive conditions are used:

• Technological options for its implementation on a certain area, given operating temperature and power supply voltage.

• Electrical safety and electrical insulation strength.

The implementation of the proposed system requires the design of a foil heater according to the parameters of the photovoltaic system. It is necessary the classical design methodology [1], [2], [3] to be harmonized with the electrical parameters used in PV source. Overall mode of the foil heater with wide surface is shown in Figure 2.

When designing foil heaters with wide surface, according to [1], [2] [3] are set:

• P_1 and P_2 – the electric power of both heating elements must be complied with the required heating of the room;

• τ_n - temperature of heat transfer surface temperature of the heater (work temperature). It is defined at sizes c and f (Fig. 2), when working with power $P = P_1 + P_2$;

• U - power supply voltage, which is determined by the method of connecting the photovoltaic elements;

• a - thick of metallic layer of material from which the heater is made.

The design involves specifying the width of the metal bands b_1 and b_2 and the number of connected in series strips n.

According to indications from Figure 2 the following expressions for real electrical resistances of two heating elements are:

$$R_1 = \frac{n.\rho_t.l_s}{a.b_1}[\Omega]; \quad R_1 = \frac{n.\rho_t.l_s}{a.b_2}[\Omega]; \qquad (1)$$

n – numbers of metal strips;

l_s – average length of a strip - $l_s = c + \Delta$;

ρ_1 – resistivity of the metallic layer at operating temperature τ_n.

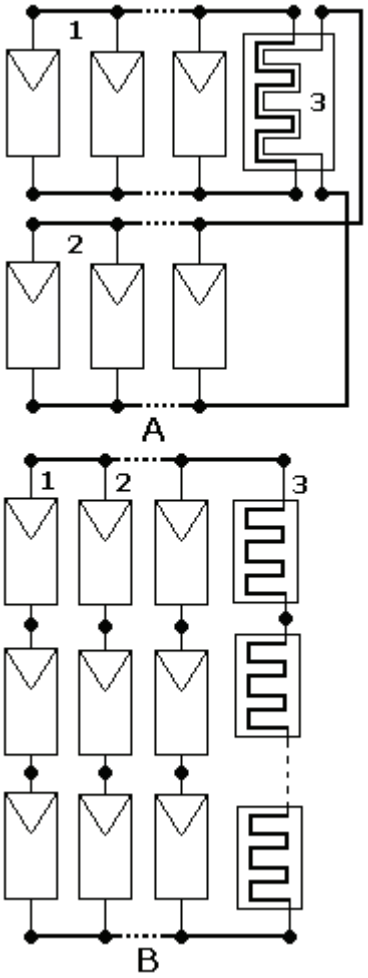


Fig. 1 The system photovoltaic strings (1.2) with a foil heater (3). A - parallel connection of the modules when the foil heater is with two parallel stripes; C –connection in series

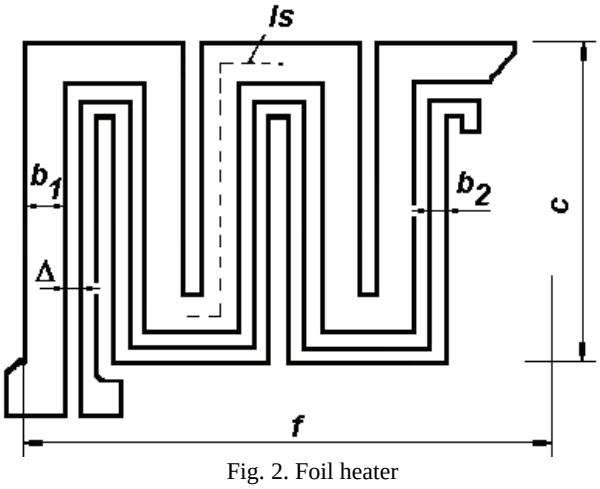


Fig. 2. Foil heater

If it should be affixed:

$$K = \frac{P_1}{P_2} \text{ it follows: } K = \frac{b_1}{b_2}; \quad K = \frac{R_2}{R_1}.$$

Provided that two heating elements operate in parallel:

$$R_e = \frac{R_2}{1+K}$$

From the expression of power $P_1 + P_2 = \frac{U^2}{R_e}$.

In such a case occurs $P_2 = \frac{U^2.a.b_2}{n.\rho_t.(c+\Delta)}$.

Full surface of the heater expressed by the geometric dimensions *c, f, n, b2* and *k* is:

$$S_{ne} = c.f = n.[2.\Delta + b_2.(1+K)].c \qquad (2)$$

The number *n* of two strips heater element is:

$$n = \frac{f}{2.\Delta + b_2.(1+K)} \qquad (3)$$

The error that is allowed by formula (3) consists of non-recording of an area between two adjacent strips, but when a large number of bands *n*, it is insignificant.
Quadratic equation in terms of a species:

$$(1+K).b_2^2 + 2.\Delta.b_2 - \frac{P_2.f.\rho_t.(c+\Delta)}{U^2.a} = 0 \qquad (4)$$

If: $A = \frac{P_2.f.\rho_t.(c+\Delta)}{U^2.a}$

The equation for determining the width of the stripe b_2 of the heating element with power P_2 is:

$$b_2 = \frac{1}{K+1}\left(\sqrt{\Delta^2 + A(K+1)} - \Delta\right) \qquad (5)$$

For complete dimensioning of two heating elements it is necessary to allocate and b_1 and *n*.

Coefficient of filling K_z of the heating element is determined as a ratio between the surface, occupied by the metal stripe on the heating element, and the total area of the heater, determined by the sizes *f* and *c*:

$$K_z = \frac{(c+\Delta).b_2.(1+K)}{c.[2.\Delta + b_2(1+K)]} \qquad (6)$$

There are three cases [1],[2],[3]:

Both heating elements are with the same power: $P_1=P_2=P$; K=1

$$b_1 = b_2 = b = \frac{1}{2}\left(\sqrt{\Delta^2 + 2A} - \Delta\right) \qquad (7)$$

The number of stripes ***n*** is:

$$n = \frac{f}{2.(b+\Delta)}; \; K_z = \frac{b.(c+\Delta)}{c.(b+\Delta)}$$

In this case they can be connected in parallel with the power of a common string or a separate power Fig.2. The

second method is applicable for fixed roof where both strings are with different orientation.

Heater consists of a heating element with power P:

$$b = -\frac{\Delta}{2} + \sqrt{\frac{\Delta^2}{4} + \frac{P.f.\rho_t.(c+\Delta)}{U^2.a}} =$$
$$= \frac{1}{2}\left(\sqrt{\Delta^2 + 4A} - \Delta\right) \quad (8)$$
$$n = \frac{f}{(b+\Delta)}; K_z = \frac{b.(c+\Delta)}{c.(b+\Delta)}$$

Instead of sizes *c* and *f* their relationship is set:

$$m = \frac{f}{c}; \quad S = c.f \quad (9)$$

The width of the stripe:

$$b = -\frac{\Delta}{2} + \sqrt{\frac{\Delta^2}{4} + \frac{P.\rho_t.S.(1+\Delta.(m/s)^{-2})}{U^2.a}} \quad (10)$$

Number of stripes:

$$n = \frac{\sqrt{m.s}}{b+\Delta} \quad (11)$$

Proposed equations (1-11) show the relation between the electrical parameters of the system (photovoltaic string - foil heater) and the geometric characteristics of the heating part. Thus allows sizing of the installation area according to required surface of the heater for the specific case under consideration.

It should be noted that the structure is possible and with wind and / or hybrid systems, but proposed realization is with photovoltaic modules. The main advantage is the possibility of obtaining the different electrical parameters of the supply system and the opportunity for sizing heaters with different characteristics. The design of PV system is carried out by established methods [4], [5].

Experimental study was done over photovoltaic system with 36 modules, 2,16 kW, 1000W/m^2 and 25^{0}C (STC). The location of the systems is 43.12, 47.55 - Varna. The proposed figures show:

- Fig.3. - Current and voltage of the system for five days - October;
- Figure 4 - current and voltage of the system for five days - August. The heating system is tested and during the summer months for comparison purposes.
- Fig.5 - output voltages (October) in organizing the different strings, which depends on foil heater sizing:
- Graph 1 - 2 (in module string) / 18 (string);
- Chart 2 - 3 / 12;
- Figure 4 - 6 / 6;
- Chart 3 - 4 / 9;
- Fig. 6 - output current in the organization of different strings. The name of the graph is as Fig. 5.
- Fig. 7 - Power system working day of October.

III. CONCLUSION

Performed theoretical analysis and experiments confirmed the above mentioned advantages of using folic heating technology to build an autonomous heating system powered by photovoltaic modules.

The proposed study shows that the parameters of photovoltaic systems provide the opportunity for a heating system with the foil heaters. The design is performed in a wide range of electrical parameters, which depend on the method of connection and building strings. Determine the number of heating elements with appropriate size and power is available as required by the heating facility. Thus heating system that meets the specific temperature requirements for distribution is realized. Opportunity for direct connection of heaters and photovoltaic elements creates conditions for increasing the efficiency of utilization of solar energy.

REFERENCES

[1] Panayotov Design of electric heaters for low-folic-type heaters. Ph.D. dissertation. Varna, 1986.

[2] Panayotov S. I. Kontrov. Summary methodology for the calculation of surface electric heater - foil type with deeply wavy shape. Scientific Session TU- Gabrovo, 1981.

[3] Panayotov S. I. Kontrov. Optimization of some design parameters of heat-convections foils heaters. TEHMA 88, MIC-Varna, 1988.

[4] Markvart T., L. CastanerPractical Handbook of Photovoltaics: Fundamentals and Applications. Elsevier Science Inc. 2003.

[5] Messenger R. J. Venture. Photovoltaic System Engineering. CRC Press LLC 2004.

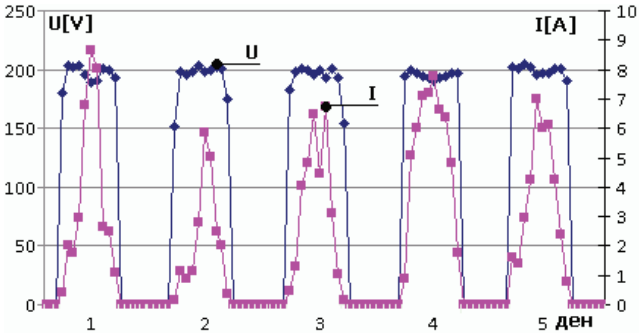


Fig. 3. Voltage and current for five working days of October

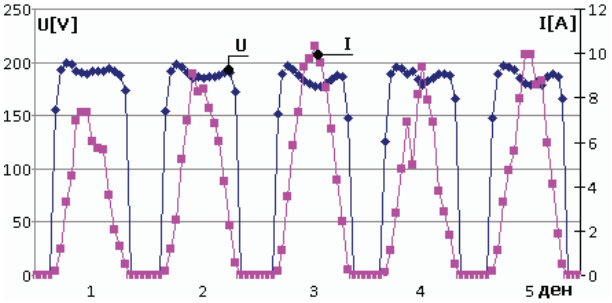


Fig. 4. Voltage and current for five working days of August

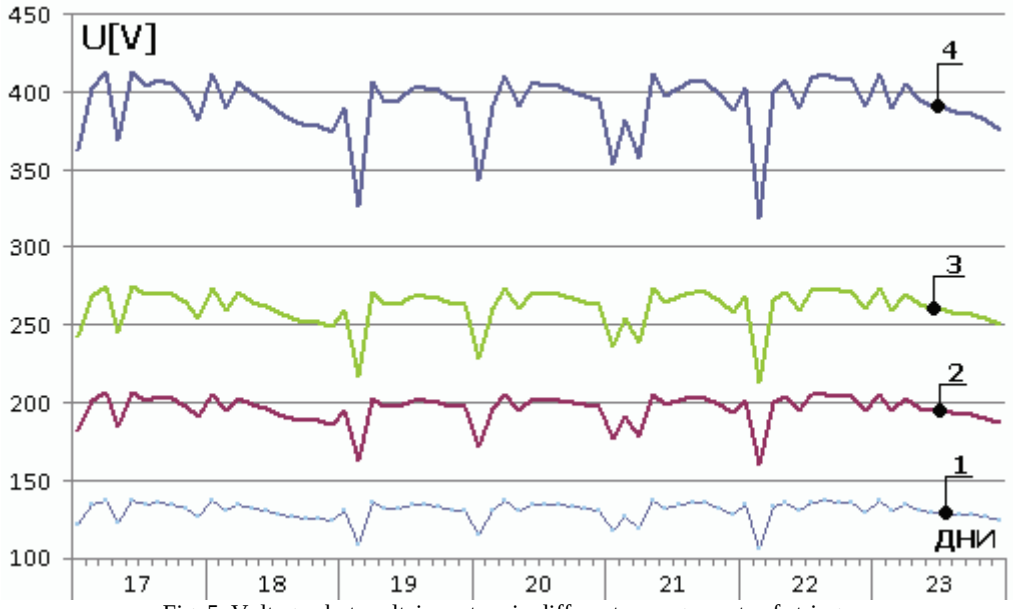


Fig. 5. Voltage photovoltaic system in different arrangements of strings:
1-2 (in module string) / 18 (string) 2 - 3 / 12, 3 - 4 / 9, 4 - 6 / 6

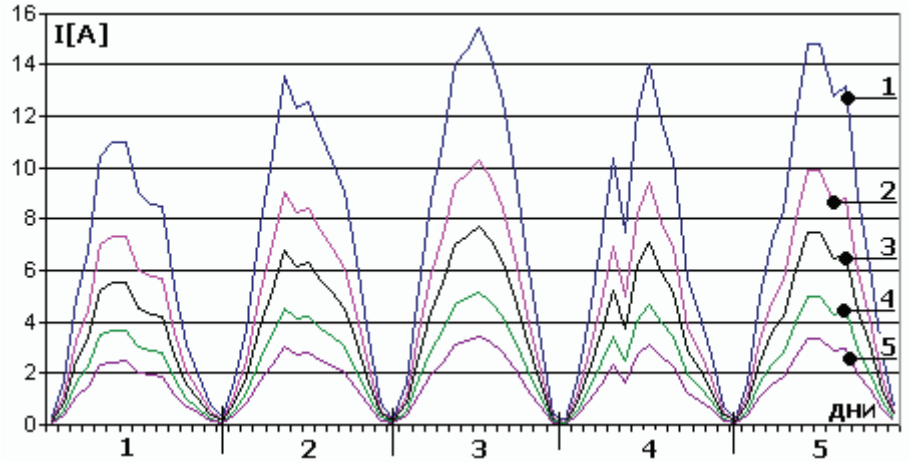


Fig. 6. Photovoltaic power system with different arrangements of strings:
1-2 (in module string) / 18 (string) 2 - 3 / 12, 3 - 4 / 9, 4 - 6 / 6

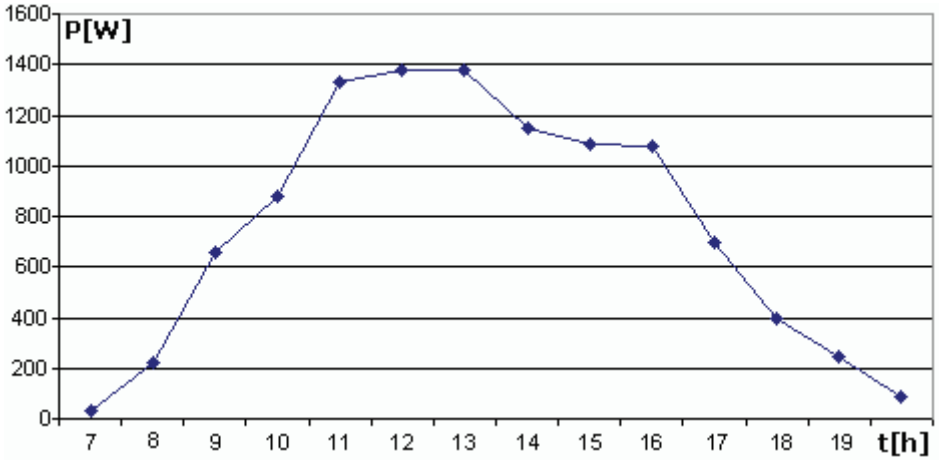


Fig. 7. Photovoltaic power systems within one working day of October

Silicon Sensors for Systems for Assessment of Sun Potential

Irina P. Pavlova[1], Ilia I. Hadzhidimov[2] and Emilian B. Bekov[3]

***Abstract* – As a renewable source of energy, the Sun produces low-calorie energy. The transformation of the Sun energy into electricity can be done by photovoltaic modules. Their projection, construction and installation depend mainly on the available Sun potential. Consequently, the Sun radiation for every region is essential. The present article aims to outline and shows the features of the devices measuring Sun radiation. An accent is put on the silicon cells for systems for assessing the Sun potential.**

***Keywords* silicon radiation sensor, pyranometer.**

I. INTRODUCTION

The sunlight occurs to be a widespread source of energy in the contemporary way of living for improving the life comfort. A positive change in people's attitude towards using the Sun energy as a renewable source is observed. A proof for that is the introduction of the photovoltaic modules, transforming the Sun light directly into electricity. This energy depends on the equality of the Sun radiation, the climate, the presence of fogs, and the air pollution (presence of dust).

The following factors influence on the right construction of a photovoltaic sun system: location, relief and orientation in the definite terrain. The duration of the periods of the Sun shining and the great irradiation with sun rays make the Sun energy an economically grounded alternative. All of these factors are a condition for the easy and accessible service of the system. Hence, having reliable data of the available Sun potential is significant. It influences the planning and the best location of the PV modules directly. The right choice of the slope of the receiving surface leads to full utilization of the Sun energy, irrespectively of the location on Earth the observed region is in. The available Sun potential influences on the power of the foreseen and desirable PV system. The right assessment and the exact defining of the Sun potential have an impact on the costs of the projecting, the installing and the exploitation of the Sun installation. Therefore the high initial investment expenses are still the main reason for the slower entering and usage of the PV system by the mass consumer.

The intensity of solar radiation is needed in the design as well as in exploitation of equipment which uses solar energy.

[1]Irina P. Pavlova,Technical University in Varna, 9000, Bulgaria, E-mail: irinappavlova@yahoo.com

[2]Ilia Iv. Hadzhidimov, Technical University in Varna, 9000, Bulgaria, E-mail: i_hadzhidimov@tu-varna.bg

[3]Emilian B. Bekov, Technical University in Varna, 9000, Bulgaria.

The purpose of this article is to demonstrate that the expensive pyranometer with high level of accuracy can be successfully replaced by cheap silicon cells. They may measure the intensity of solar radiation with sufficient precision. A major problem in the use of silicon cells as solarimeter is the dependence between volt-ampere characteristic and temperature. The increase of temperature of the cell leads to efficiency decrease of silicon cells. This leads to strong nonlinear dependence of the intensity of solar radiation from the output voltage of the cell. Standard cells depending on their characteristics have a limit at which the voltage remains the same regardless the increase in the intensity of the sun. To avoid this, the cell's output voltage and the electricity are controlled by electronics. On the one hand this offsets the dependence on temperature and on the other hand, the output voltage decreases below the limit for the cell at maximum intensity of the Sun 1200 W/m^2.

II. EXPERIMENTAL DATA

Developed silicon cell type DIY, shown in Figure 1 has the following characteristics:

- Size 60x60x3 mm;
- Rated voltage 4,5 V;
- Short circuit current 50 mA;
- Waterproofing (polycarbonate).

Experiments to measure the solar radiation have been carried out with the developed silicon cell and a standard pyranometer Kipp & Zonen type CM 11 B, shown in Figure 2 with the following characteristics:

- Spectral range - 305-2800 nm;
- Sensitivity - 4-6 μV/W/m^2;
- Impedance (nominal) – 700-1500 Ω;
- Response time - 15 sec.;
- Non-linearly - $\prec \pm 0{,}6\%(\prec 1000 W/m^2)$;
- Temperature dependence of sensitivity - $\prec \pm 1\%(-10 to + 40^o C)$;
- Directional error - $\prec \pm 10 W/m^2$ (beam$1000 W/m^2$);
- Tilt error - None;

Fig.1. Silicon cell of the DIY type

Fig.2 Pyranometer Kipp & Zonen of the CM 11 B type[1]

- Zero-offset due to temp.changes $\prec \pm 2 W/m^2$ at 5K / h
- Operating temperature - -40°C to + 80°C;
- ISO-9060 Class - Secondary Standard;
- Dimensions WxH- 150.0mmx91,5mm;
- Weight - 850 grams [1].

For calculation of the solar irradiance the following formula has be applied:

$$E_{Solar} = \frac{U}{S} \qquad (1)$$

where

E_{Solar} – Global radiation , W/m^2;

U – Output voltage, μV;

S – Sensitivity, $\mu V / W / m^2$ [2].

The CM11 B pyranometer is intended for high accuracy total global, or diffuse sky, solar radiation measurement research on a plane/level surface.

Its usage is due to the following advantages:

- low dome thermal offset error;
- excellent cosine/directional response;
- excellent long term stability of sensitivity;
- excellent linearity performance.

The non-linearity error is shown on Figure 3.

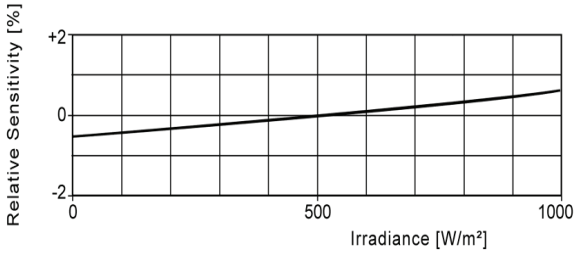


Fig.3. Non-linearity (sensitivity variation with irradiance) of Kipp & Zonen pyranometer CM 11 B.

The temperature dependence of the sensitivity is an individual function. For a CM 11 B the curve is somewhere in the shaded region of Fig. 4.

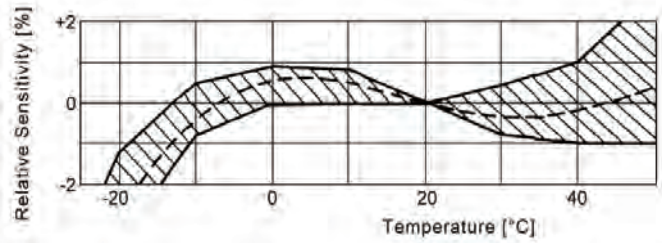


Fig. 4. The curve of relative sensitivity variation with instrument temperature of a pyranometer CM 11 B is in the shaded region. A typical curve is drawn [2].

The directional error is the summation of the azimuth and zenith error and is commonly given in W/m^2. Figure 5 shows the maximum relative zenith error in any azimuth direction for the CM 11 B.

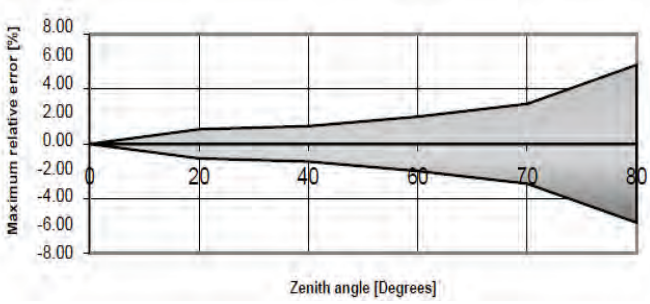


Fig. 5. Relative directional error [2]

Experiments to measure the solar radiation have been carried out with the developed silicon cell and a standard pyranometer Kipp & Zonen type CM 11 B. The experiments

took place within the Technical University - Varna for a period of one month for 500 minutes per day. The time of memorizing the results from the experiment is one minute, which makes it possible to trace in detail the changes of the solar radiation intensity. Processing the results from the measurement of intensity of solar radiation with the developed silicon cell is performed with parabolic dependence:

$$E(x) = a.x + b.x^2 , \qquad (2)$$

where a = 6,55 b = -0,01 (for this cell) are regression equation coefficients.

For standard pyranometer CM 11 B the intensity of solar radiation is calculated as follows:

$$K\_Z(x) = \frac{x - 0,2}{0,8}.1600 \quad , W/m^2. \qquad (3)$$

The standard square deviation is calculated with formula (4), and the error of measurement between the readings of the two instruments is calculated with formula (5):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{500} (\delta_i)^2}{500}} \qquad (4)$$

$$\delta = |K\_Z - E(Cell)|. \qquad (5)$$

Amongst the performed experiments, selected and displayed measurements are obtained during two sunny days (20.10. and 29.10.2008) and two cloudy days (24.10. end 21.10.2008).

The results are shown in Fig.6, Fig. 7, Fig. 8 end Fig.9.

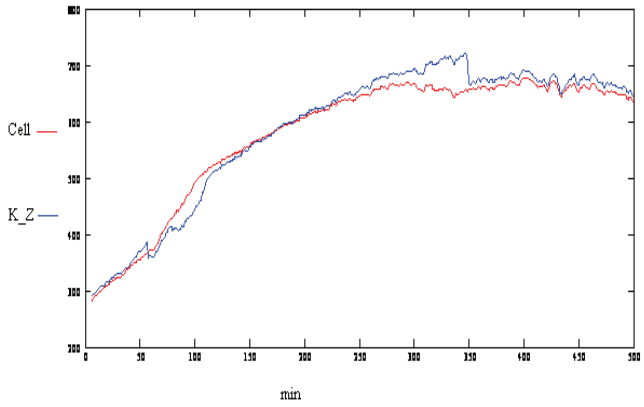


Fig. 6. Results of the intensity measuring of the Sun radiation on 20th October, 2008

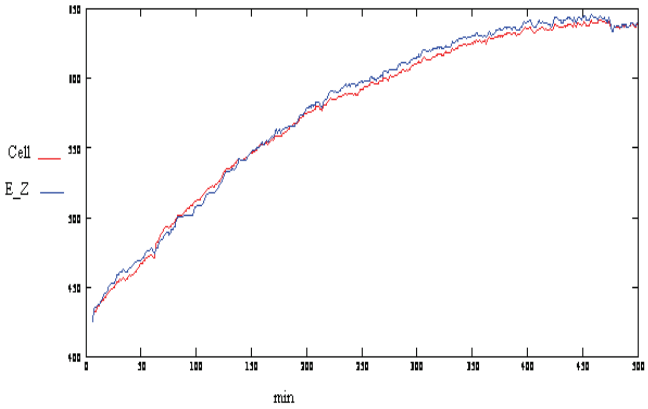


Fig. 7. Results of the intensity measuring of the Sun radiation on 29th October, 2008

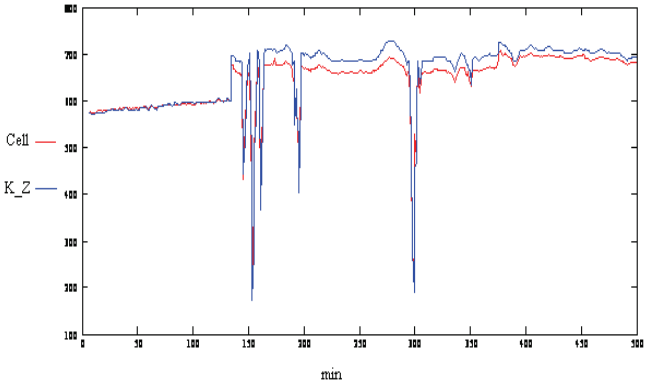


Fig. 8. Results of the intensity measuring of the Sun radiation on 24th October, 2008

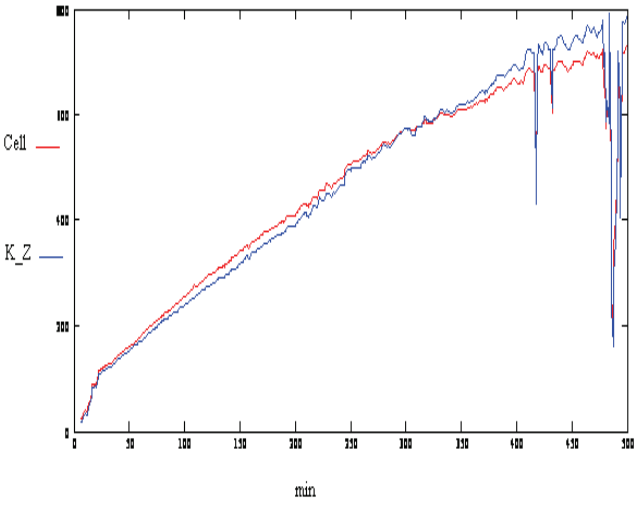


Fig. 9. Results of the intensity measuring of the Sun radiation on 21th October, 2008

Considering the results of measurements of intensity of the solar potential - the blue line (corresponding to the data of the standard pyranometer) and red line (solar silicon cell) show a good degree of matching. At two sites (Fig. 6) the deviations can be seen, caused by a shift in the position of the silicon sensor in azimuth in order to analyze the possible error by a change in the assembly position of the sensor. After the initial

removal of the sensor, it is returned to starting position and the difference in readings with the standard was decreased. Figs. 8 end 9 compares the readings of the two solarimeters in a cloudy surrounding and a several peaks at high eclipse.

The error of measurement with silicon sell versus the standard pyranometer is shown on Fig.10, Fig.11, Fig.12 and Fig.13.

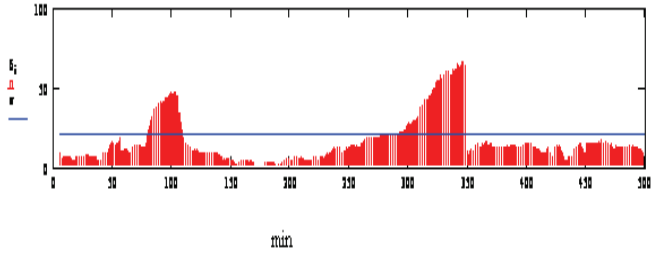


Fig. 10. Mistake of the measuring – 20th October 2008

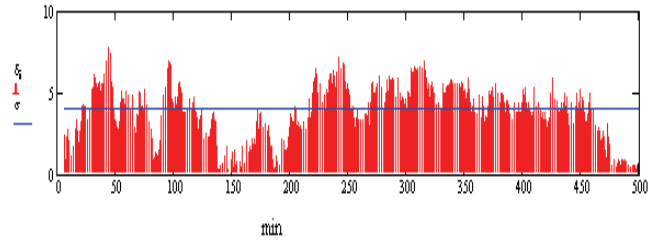


Fig.11. Mistake of the measuring – 29th October 2008

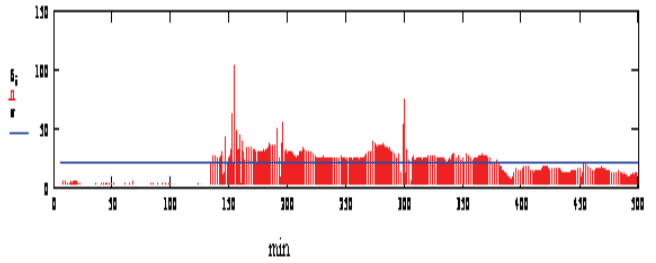


Fig. 12. Mistake of the measuring – 24th October 2008

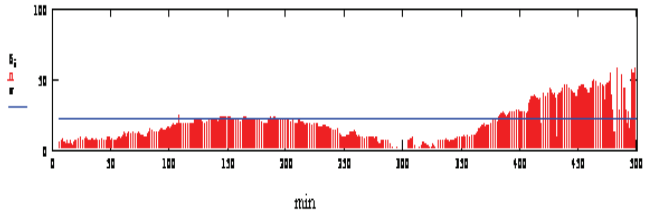


Fig. 13. Mistake of the measuring – 21th October 2008

According to formulas (3) and (4) and Fig.10, Fig.11, Fig.12 and Fig.13, the acentric square standard deviation and the maximum error obtained the following values:

$\sigma = 21.34604$; $\max(\delta) = 65.37476 W/m^2$;

$\sigma = 4.0774$; $\max(\delta) = 7.6761 W/m^2$;

$\sigma = 21.56339$; $\max(\delta) = 104.60971 W/m^2$;

$\sigma = 22.49228$; $\max(\delta) = 58.92863 W/m^2$.

III. Conclusion

The obtained data shows that the developed Sun sell can be successfully used for assessing the density of the Sun radiation. Such sensors are used even more frequently in the practice due to the following peculiarities:

- low cost;
- stability over atmospheric influences;
- compactness and simple geometrical form;
- chance for easy installment in different positions;
- acceptable accuracy in assessing of the Sun potential.

ACKNOWLEDGMENTS

The authors would like to extend their gratitude to Fund "Scientific Research" of the Bulgarian Ministry of Education for the support within Project "Development of specialized scientific infrastructure for investigation of wind and solar energy potential", No.D 002-48/10.12.2008.

REFERENCES

[1] http://meteostanice.agrobiologie.cz/figs/cm11.pdf

[2]http://www.skypowerinternational.com/pdf/Radiation/7.1415.01.000_cm11-14_bed-anleitung_engl.pdf

Increasing the Role of RES and Correct Assessment the Solar Potential of Bulgaria

Zheko D. Kiryakov[1] and Iliya I. Hadzhidimov[2]

Abstract – In recent years the issue of scale utilization of renewable energy sources places great importance as recognizing that this is one way to prevent global warming and reduce dependence on countries with limited resources of organic fuels from imported ones.

Keywords – renewable energy sources, global warming

I. INTRODUCTION.

A. Increasing the role of RES.

In recent years the issue of scale utilization of renewable energy sources places great importance as recognizing that this is one way to prevent global warming and reduce dependence on countries with limited resources of organic fuels from imported ones.

The utilization of renewable energy in the world, including Europe and in Bulgaria passed the stage of emotional enthusiasm and can say that now is the stage of assessing the technical and economic realities.

EU energy consumption is growing steadily. Question arises to meet future needs, while increasing requirements for environmentally friendly energy. To meet the challenges of the future in 1997 was admitted to the White Book on renewable energy resources, which contains a strategy for their development. In 2000 this document was followed by the Green Book relating to security of energy. Both documents highlighted the importance of developing renewable energy.

From 2007 to the Member States of the EU have set the following targets.

* Energy consumption in Europe by 2020 to reduce by 20% in energy consumption in 2006

*The share of renewables in total consumption of energy sources to 12% in 2010 and 20% in 2020

*Share of biomass reaches 9% in the total energy in 2020.

In this context, assessing the achievable levels of utilization of each available RES. The targets are very high and achieving them is possible with intensive development and improvement of energy technologies transformation and is available on the market only a highly efficient, profitable and environmentally friendly systems and products are compatible with existing and future architecture, construction and industrial technologies.

[1]Zheko D. Kiryakov is with the Technical University of Varna, 9000 Varna, Bulgaria, E-mail: jeko_kiriakov@abv.bg

[2]Iliya I. Hadzhidimov is with the Technical University of Varna, 9000 Varna, Bulgaria, E-mail: i_hadzhidimov@tu-varna.bg

B. Policies and measures.

To reach our country to the 2010 indicative target maximum target: the share of renewables in gross domestic electricity consumption to reach 11% (Under favorable climate conditions) in terms of increasing consumption of electricity needed large complex targeted measures at the state level as well as for accelerated introduction of renewable energy and energy saving. Especially important is to limit and break the trend of absolute increase in power consumption not only through energy efficiency measures (energy efficiency), but end by redirecting consumers (especially industry) to alternative fuels and energy (at least by optimizing prices of electricity for business).

C. National policy to promote consumption.

In accordance with European legislation and in fulfillment of our commitments in the negotiation process and recommendations of the Commission in the Republic of Bulgaria Act was developed for renewable and alternative energy sources and biofuels. The main objectives of the law relating to the diversification of energy supplies, reduce the cost of imported energy resources and energy, increasing the capacity of small and medium producers of renewable and alternative energy sources and biofuels and other renewable fuels, environmental and creating conditions for achieving sustainable development at local and regional level.

The objectives will be achieved by promoting development and use of technologies for production and consumption of energy produced from renewable and alternative energy sources and promote the development and use of tehnoloogii production and consumption of biofuels and other renewable fuels.

This law achieved full harmonization with Directive 2001/77/EC on the promotion of electricity produced from renewable energy sources in the internal electricity market and Directive 2003/30/EO to promote the use of biofuels or other renewable fuels for transport. The Act regulates the public relations to promote the production and consumption of electricity, heat and / or cooling energy from renewable and alternative energy sources, as well as production and consumption of biofuels. The provisions of the law is introduced for the production and consumption of biofuels,

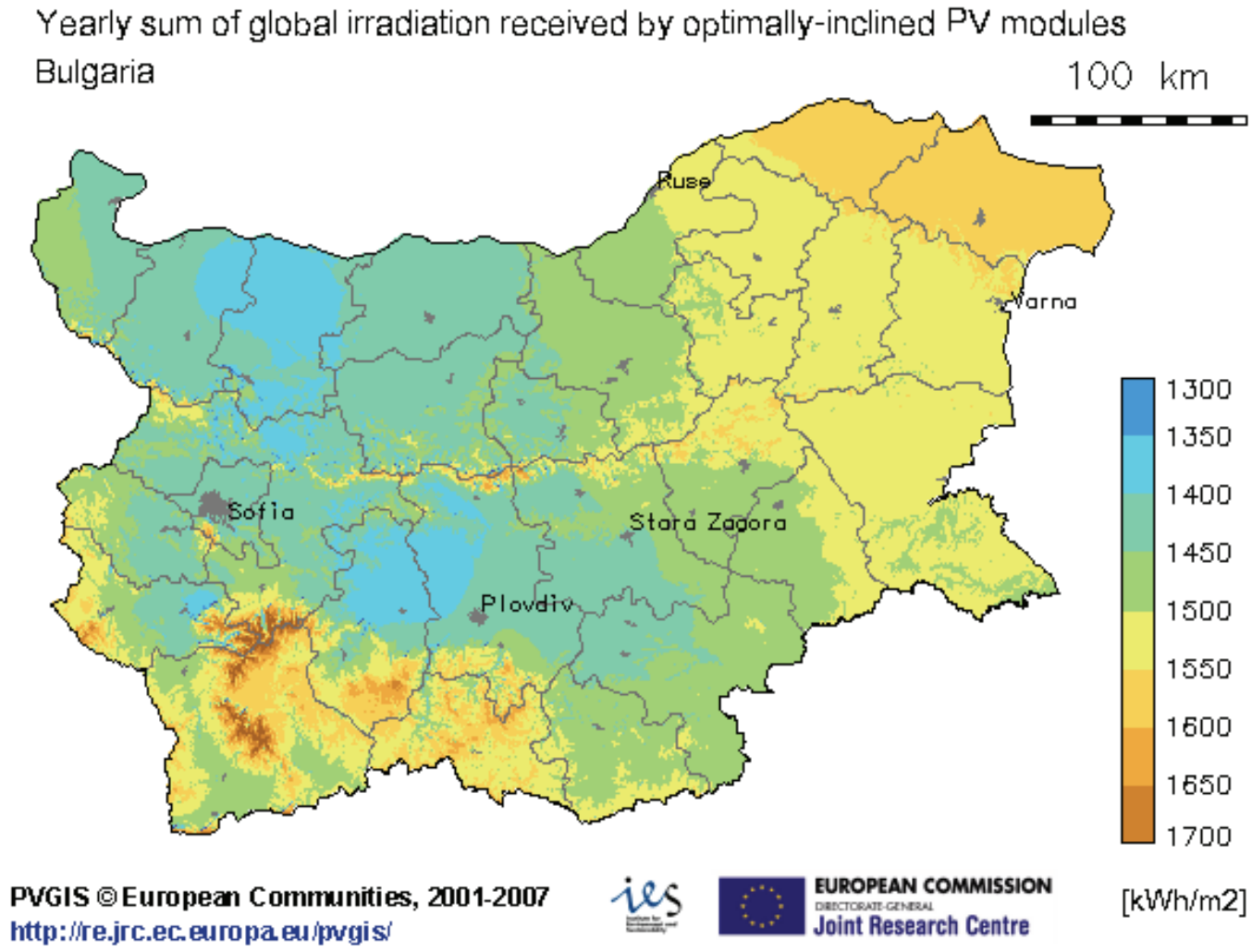


Figure 1: Annual total distribution of total solar radiation of PV module at the optimum slope 2001-2007

which will reduce emissions from the transport sector and the use of conventional fuels.

The Act provides for the development of national targets for:

- Promoting the consumption of electricity produced from RES.
- Promoting use of biofuels in transport sector.

II. Potential of Solar and Wind Energy in Bulgaria.

A. Wind Energy.

According to recent forecasts of the European wind energy association, a trend of increased use of development of wind energy in Europe. Expected installed capacity of 28 400 MW in 2003 to reach 75 000 MW in 2010. and 180 000 MW in 2020. In 2020 electricity generated from wind turbines, will cover the needs of 195 million Europeans, or half the population of the continent. According to EUROSTAT estimates of consumption of wind energy in the EU in 2010 will reach 10 000 ktoe. In Bulgaria there is little wind energy contribution to gross electricity production in the country. In 2001, the wind produced 35 MWh (3 toe), in 2003 -63 MWh (5.4 toe), and in 2004 .- 707 MWh (60 .8 ktoe). This shows that development of wind energy in Bulgaria has accelerated.

B. Solar energy.

The theoretical potential of solar energy is defined as the average amount of solar heat for one year on one square meter of horizontal surface and is expressed in kWh/m2.

Affordable solar energy potential is determined after taking into account of several key factors: the uneven distribution of energy resources of solar energy in different seasons; physiographic features of the territory, restrictions on construction and operation of solar systems in specific areas, such as nature reserves and other military facilities.

World Energy Council (WEC) identified as a possible potential of solar energy worldwide 1 575 EJ / yr.

III. The Distribution of Solar Radiation in Bulgaria and Analysis of Solar Potential.

Due to its southern exposure, and because of its relatively small clouds, Bulgaria receives a significant amount of total (direct and diffuse) radiation. In June, where is the longest day and the highest altitude of the sun, complete with serene radiation reaches approximately 9181h (m2.day). The average annual values in 5927 ranged from the most northern parts to 6160 Wh / (m2.day) for the south. The values for the total radiation in the cloudless sky are minimal in December - from 2441 to 2789 Wh (m2.day).

The average annual values of total radiation in the actual conditions of clouds in different regions of the country have ranged from 3835 to 4649 Wh (m2.day). The values of total radiation are greatest in July. Although in June the length of day is the greatest height and the sun reaches its maximum value due to the considerable cloudiness values of total radiation is less than in July.

According to data from the system PVGIS the European Union and Bulgaria data in Figure 1 and Figure 2 presents the distribution of total solar radiation on the surface of PV collector placed at the optimum angle corresponding to the geography of the area.

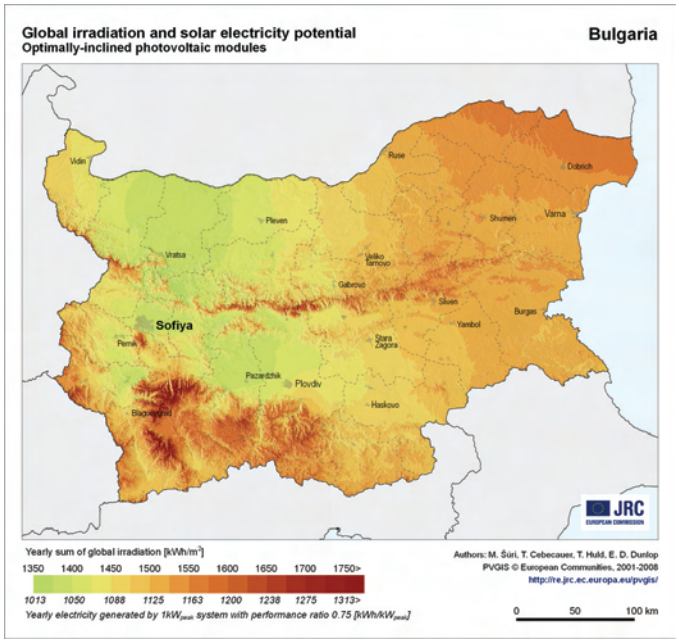


Figure 2: Annual total distribution of total solar radiation of PV module at the optimum tilt, 2001-2008

Figure 3: Annual average solar radiation on horizontal surfaces and on optimal angle

IV. Assessment the Potential of Solar Radiation in Bulgaria

The average amount of sunshine in Bulgaria is about 2 150 hours, and average solar radiation resource is 1 517 kWh m2. Generally receives total theoretical potential of solar energy incident upon the country for one year in the range of 13.10 ktoe. Available as an annual potential of solar energy utilization can indicate approximately 390 ktoe. As an official source for assessing the potential of using solar energy project under the PHARE, BG9307-03-01-L001, "Technical and economic assessment of renewable energy in Bulgaria. At the heart of the project are set out data from the Institute of Meteorology and Hydrology, BAS, all derived from 119 meteorological stations in Bulgaria for a period of over 30 years. After analysis of the databases is made regionalization of the country of solar potential and Bulgaria is divided into three regions, depending on the intensity of sunshine (Figure 3). Other potential map can be made by METEONORM, (Figure 4).

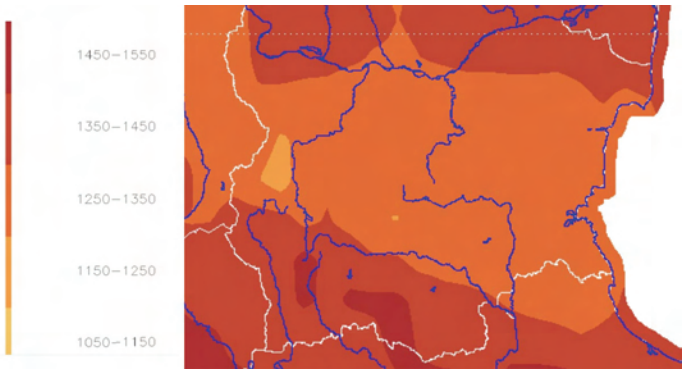


Figure 4. Potential data by METEONORM in kWh / m2 / Year

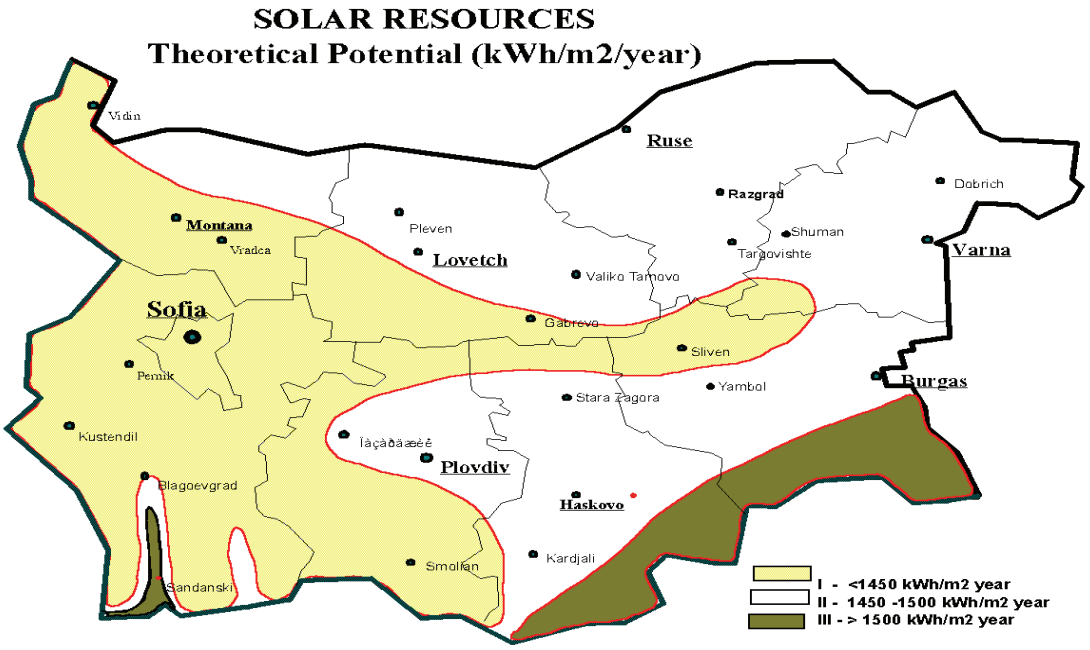


Figure 5. Theoretical potential of solar radiation in regions

On Figure 5, a theoretical distribution of the solar radiation has been performed. On Figure 6 an attempt to analyze the annual average solar radiation (Figure 3) has been performed. A nonlinear regression is obtained with 2 parameters (latitude, longitude) and 30 coefficients for an equation of 7-th degree. As seen there is not much coincidence with the other solar maps as seen on figures 1,2,4 and 5.

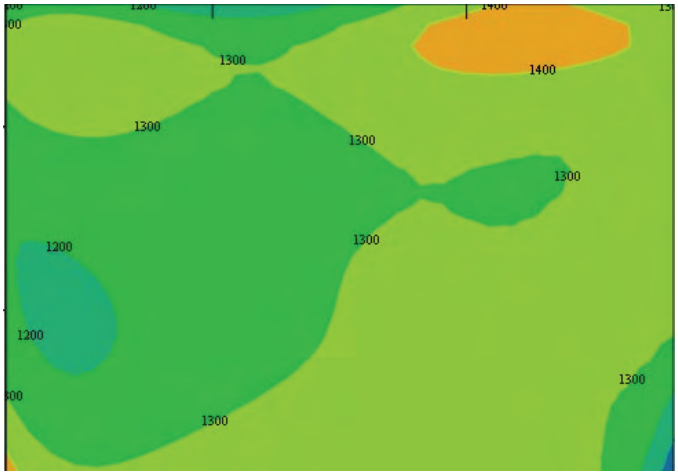


Figure 6 : Statistical data of solar radiation measurement, treated with regressional equiation

V. CONCLUSION:

As seen above, the data from the sources differs each other. Therefore it is necessary to create actual map of solar potential. The problem could be solved by at least annual measurement at different sites of the country. Also a data base must be performed in order to analyze the global changes during the years. This is necessary from the current point of view, targeting the era of the clean energies, where the solar energy takes one of the first places. Creating a detailed solar potential map will help the investors to take a fast and correct decision about building solar plants at different places. Also it is very important to have in mind the problems of shadowing, so the solar maps could be combined with topographic.

ACKNOWLEDGMENTS

The authors would like to extend their gratitude to Fund “Scientific Research” of the Bulgarian Ministry of Education for the support within Project “Development of specialized scientific infrastructure for investigation of wind and solar energy potential”, No.D 002-48/10.12.2008.

REFERENCES

[1] National Long-term Energy Efficiency Program until 2015, 2005.
[2] PVGIS, European Community, 2001-2007.
[3] PVGIS, European Community, 2001-2008.
[4] METONORM, 2009.

Investigation of an Autonomous Hybrid System for Direct Supply of an Induction Heating Device

Vencislav Valchev[1], Hristofor Tahrilov, Borislav Dimitrov

***Abstract* – The paper presents an autonomous renewable energy hybrid system for supplying an induction heating device. The system realizes direct supply of the inverter device by the converted wind and solar energy, thus omitting the necessary 50 Hz DC-AC converter for energy injection in the grid. The discussed approach is applicable for other industrial systems. The realized DC-AC converter operates at increased frequency and combines the requirements of both the supply RES and the induction heating load.**

***Keywords* – Wind and solar energy conversion, induction heating.**

I. INTRODUCTION

Renewable energy sources (RES) are the alternative for more and more depleting fossil fuels. A specific application of RES is for supplying induction heating systems. This combination, described in [1, 2, 3], based on transistor converter with increased frequency, does not require the usage of a specialized inverter. In this case it is necessary to match the power and the voltage of the source and the load.

The goal of this article is to investigate an autonomous hybrid system, built from wind generators and photovoltaic panels, for direct supplying an induction heating converter. The advantages of the system should are:

- Removing of the standard inverter. This simplifies the whole system and reduces its price. This requires realization of technical solutions which allows direct supplying of the converter form the RES system.
- The output parameters of the converter (*I, V , P, cosφ*) are according to the requirements of the used inductor. The working frequency is selected according to contour resonant frequency.
- The converter is built for each specific case with the following capabilities: sinusoidal output voltage is not obligatory and no system for synchronization and parallel work with the power grid is required.
- Additional blocks of the system – thermo sensor, temperature controller, frequency synchronization system, automatic control and others – are part of the converter.

[1]Vencislav C. Valchev - Technical University – Varna, Varna "Studentska" №1, Bulgaria, E-mail: vencivalchev@hotmail.com

II. INVESTIGATION OF THE PROPOSED SYSTEM

Block diagram of the test system is shown in Fig.1a and the detailed scheme in Fig.1b. The system is build upon the classical topologies [4, 5, 6]. Two energy sources are used: wind generator (WG) and photovoltaic modules (PV). They generate two different voltages V_1 и V_2, connected to particular DC/DC converters. They should fulfil the following requirements:

- Output voltage stabilization to a specified value;
- Proportional distribution of the loading according to the power capabilities of each energy source;
- Protection in non nominal modes
- Control from a single clock generator, because of the problems.

Taking into account the specific characteristics of each energy source the schematics of the converters are chosen:

- The wind generator is a construction that does not allow modification of the schematic (correction of the output voltage) and a full bridge converter topology is used.
- The photovoltaic modules allow random way of connection (thus different output voltages and currents) and a topology with center tap transformer is used.

Fluctuations of the supplied voltage require the transformer calculation to be done according to the maximum value of the voltage. The following equation is used:

$$V = 4 k_f f w B_r S_{fe}, \quad (1)$$

$$w S_{fe} = \frac{V}{4 k_f f B_r} \quad (2)$$

where: V – voltage; w – number of turns; S_{Fe} – cross section area of the magnetic core; f – frequency; B_r – magnetic induction; k_f (in this case $k_f=1$).

The ratio between w and S_{Fe} is selected by technology and construction requirements and in compliance with the nomenclature of the manufactured magnetic cores and copper wires.

Equation (2) is used to design all transformers in the system. As a result the filter capacitor C_f is calculated by:

$$C_f = \frac{1}{2.m.f.R_d.K_n} \quad (3)$$

where: m– number of phases; R_d – load resistance, K_n– pulsation coefficient.

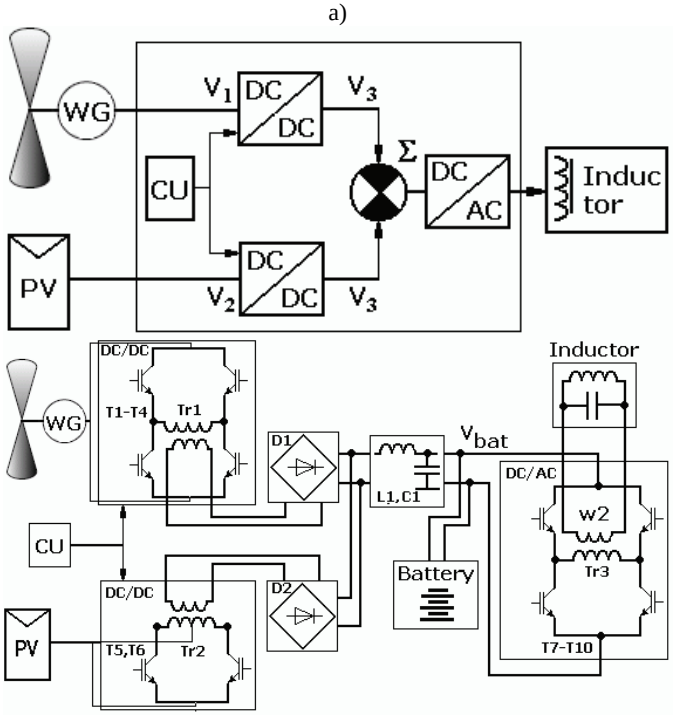
The sum of both voltages is fed into the input of the DC/AC converter. This converter operates within the frequency range of 25-35 kHz. This frequency is selected by the specific technology requirements of the heating process.

In the test system IGBT transistors are used for the DC/DC and DC/AC converters. A standard UPS system working in on-line mode is built. The investigation of the influence of power change of the load is done by changing the number of turns of the output winding w_2 of the transformer *Tr3* in Fig. 1b at a fixed input voltage (from the energy sources) V_{Bat}=180V at maximum power $P_{RES} = 2000$ W.

According to power of the load and the power of the RES, there are three possible cases:

- The power of the load is less than the maximum power of RES - $P_{LOAD} < P_{RES}$;

a)



b)

Fig.1 Block diagrams of the induction heating converters powered by RES, a); Detailed scheme, b).

- The power of the load is approximately the same as the maximum power of RES $P_{LOAD} \approx P_{RES}$;
- The power of the load is more than the maximum power of RES - $P_{LOAD} > P_{RES}$.

According to these conditions there are three operation modes:

1. When $P_{LOAD} < P_{RES}$ there are no restrictions and the operation of the two blocks RES Supply and Load are independent.

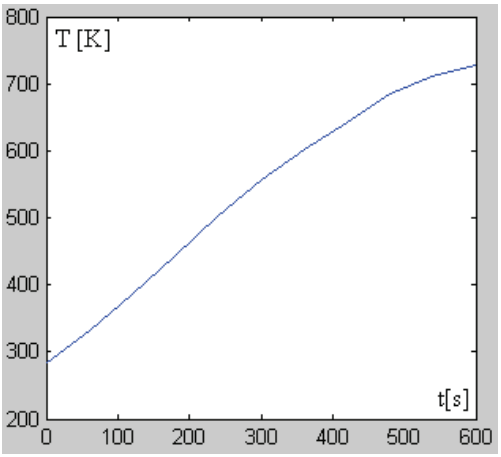
Experimental results are shown in Fig. 2 and Table I. The initial power of the load is P_{LOAD}*=1400W.*

TABLE I

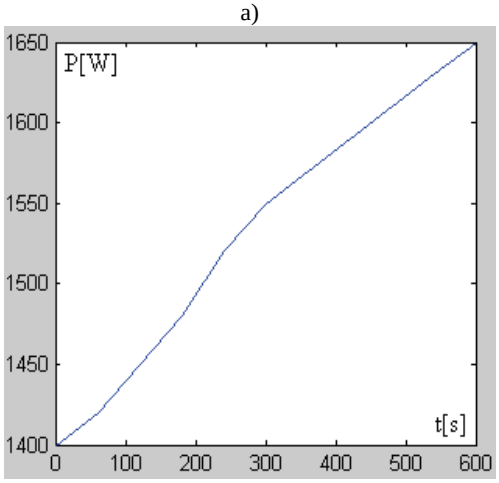
EXPERIMENTAL RESULTS AT P_{LOAD}=1400W

t	s	0	60	120	180	240	300	360	420	480	540	600
T	K	285	332	386	443	503	557	602	643	684	712	728
P	W	1400	1420	1450	1480	1520	1550	1570	1590	1610	1630	1650

2. When $P_{LOAD} \approx P_{RES}$ it is necessary to control the load because its ability to increase the power with increasing the temperature of the heated object. The experimental study of this case is done by increasing the voltage of *w2*. If the load



a)



b)

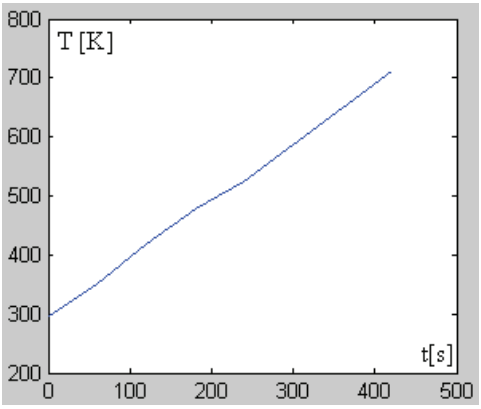
Fig.2. Transition process at induction heating of a hollow aluminum cylinder when $P_{LOAD} < P_{RES}$:
a) temperature versus time *T = f(t)*;
b) power versus time *P = f(s)*

power is $P_к = 1900 \div 2100$ W, than this overloads the RES supply. The experimental results of the temperature and time in this mode are shown in Fig. 3 and Table II.

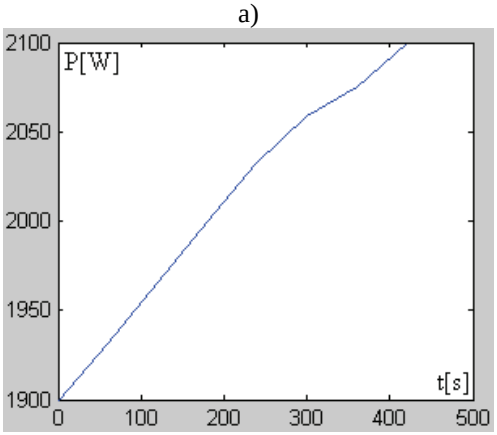
TABLE II

EXPERIMENTAL RESULTS AT P_{LOAD}=2100W

t	s	0	60	120	180	240	300	360	420
T	K	297	351	419	477	524	586	649	713
P	W	1900	1931	1966	2000	2033	2059	2075	2100



a)



b)

Фиг.3. Transition process at induction heating of a hollow aluminum cylinder when $P_{LOAD} \approx P_{RES}$:
a) temperature versus time $T = f(t)$;
b) power versus time $P = f(s)$

3. The case when $P_{LOAD} > P_{RES}$ should not be allowed for continuous operation. In repeating short time mode it is necessary to monitor the temperature of the components. In this case the goal is to reach and maintain the maximum working temperature of the components. Experimental verification should be carried out for each case.

III. Conclusion

An autonomous renewable energy hybrid system for supplying an induction heating device is realized and discussed. The advantage of the system is omitting the necessary 50 Hz DC-AC converter for energy injection in the grid by the direct supply of the inverter heating device by the converted wind and solar energy. The discussed system is realized with standard switches and components.

The approach is applicable with benefits in other industrial systems.

Acknowledgement

The paper is developed in the frames of the project 'Development of specialized scientific infrastructure for investigation of wind and solar energy potential', № ДО02-48/10.12.2008, Ministry of Education, Youth and Science, Fund 'Scientific Research', Bulgaria.

References

[1] Кувалдин А. **Индукционный нагрев ферромагнитной стали. Э**нергоатомиздат, Ленинград 1988.
[2] Слухоцкий А. **Установки индукционного нагрева**. **Э**нергоатомиздат, Ленинград 1991.
[3] Слухоцкий А. **Индукторы**. Ленинград 1989.
[4] Brown M. **Power Supply Cookbook**. Butterworth-Heinemann 1994.
[5] Ericson R. D. Maksimovic, **Fundaments of Power Electronic**, University of Colorado, Boulder 2001.
[6] Kramer. W. S Chakraborty, B. Kroposki., H. Thomas. **Adwanced Power Electronic Interfaces for Distributed Energy Systems**. National Renewable Energy Laboratory, US Department of Energy 2008.

This page intentionally left blank.

Mathematical Model for Water Use and Transfer Between Hydro Power Plants

Mile Spirovski and Arsen Arsenov

***Abstract* – In this paper optaining the basic principles for maintance continuity and amouth of movement, there are elaborated some methods for hydrograph transfer analysis throught accumulations and canals. Additional hardship creates the numerous restrictions who has to gratify the power plants like one complete unit. Like: minimal and maximal licenced expiration throught capturation constructions care for the capacity of the addictive pressure tube and drainage narrow passage, allowed velocity quickness of increase drain throught tubes, changes of the accomulations rates satisfying the given hydrograpys who with is define direct water take off from the analysed accumulation.**

***Keywords* – Hydropower plant, hydrographs, equation of…**

I. INTRODUCTION

Model of a liquid product, or that use to be called river model, is made to be able to explain the strait between two hydropower plants. There are two ways of transfer of hydrographs: Transfer of hydrograph in accumulation or reservoir and transfer of hydrograph by canal. Transfer of hydrograph by accumulation is analyze with a target to be made suffocated of the point of the hydrograph that he is subject since he gets in the accumulation. For that analyze are needed incoming data's and that hydrograph of the entrance in the accumulation and characteristic of accumulation (volume) and plants for evacuation of water from the accumulation.
Transfer of hydrograph by canal is important for fortify influence of the canal over the spade of the hydrograph ant time of travel. For that target there is a need of incoming data's for incoming hydrograph in the canal characteristic canal itself.

II. TRANSFER OF HYDROGRAPHS BY CANALS

When the hydrograph that is given travels by canal his characteristic are changing like result of the resistance of the movement that is made by the canal and his volume characteristic. In this case absence of important income through length of the canal top flow generally is suffocated. This phenomenon is called diffusion (mutually mix of fluids) and it happens in natural canals. Time difference between top (spade) of hydrograph on the entrance in the canal and on the exit of it is measure for the time of travel of the hydrograph by canal.
Under west entrance hydrograph and canal, target of the travel of the hydrograph by the canal is to be deciding the shape of the exit hydrograph (chronological phenomenon of the movement of the exit of the canal). Typical view of transferable hydrograph is given on picture 1.

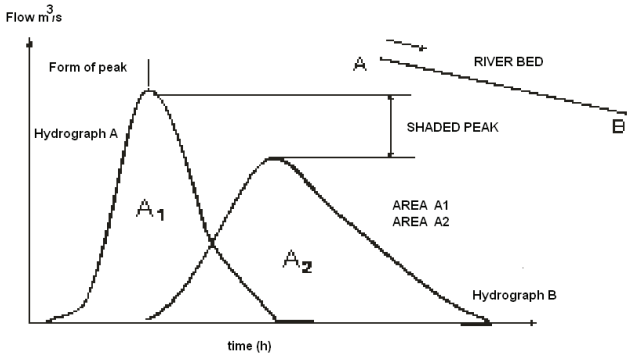


Fig. 1. Effect off transfer of hydrographs

III. SAINT VENAN - EQUATIONS

With a help of SAINT VENAN – equations, that are called and primary equations, it's describe stretch of given water wave in open canal. Water waves can be classified as: dynamical waves, gravitation waves, diffusion waves and kinematical waves in a relation how big is the number of element that are taken in the structure of the model that descries the movements of the wave. Dynamical wave, for example, it takes in consideration all the members in the structure of Saint Venan – equation for maintain of the content of movement. Under viewing of gravitation wave is underplay the effect of bow of the trough and the effect of rubbing (friction) that is develop between the water and the walls of the river. In this case the article $g\left(S_0 - S_f\right)$, the other articles of the equation for maintains of the quantity of movement are taken in consideration.

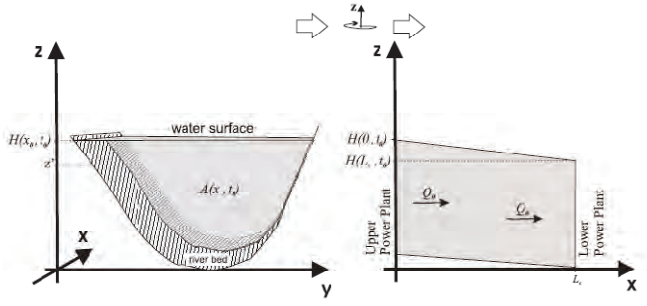


Fig. 2. Flow of water through the river

[1]Mile Spirovski Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: mile.spirovski@uklo.edu.mk,and Arsen Arsenov FEIT Skopje, E-mail: aarsenov@feit.ukim.edu.mk

Kinematical wave, again, it takes in consideration only the effects from existing of the bow river trough and the effects of the rubbing that are develop between the water and the walls of the river trough. In general form, view of Saint Venan – equations is next:

- Equation for maintence of continuity:

$$\frac{\partial Q}{\partial x}+\frac{\partial A}{\partial t}=q \qquad (1)$$

As we can presumption that step distance is infinitely small. Farther the change of the volume can be shown over first assumption of the change of the section for pass $\mathrm{A}(\mathrm{x}_0,\mathrm{t}_0)\delta x$ over the time $\delta \mathrm{t}$:

- Equation for maintains of the quantity of movement or equation for maintains of impulse

$$\frac{\partial Q}{\partial t}+\frac{\partial\left(Q^2/A\right)}{\partial x}+gA\left(\frac{\partial y}{\partial x}-S_0\right)+gAS_f=0 \qquad (2)$$

Where x is lengthwise distance under the length of the canal in m, t is time in s, A is surface lengthways section of the trough m^2; y is level of the surface of the water in the canal in m, S_0 bow on the bottom of the canal (negative); q is lengthwise income of water (continual income of water under unit length of the canal – from side) in $(m^3/s)/m$; S_f is so called bow of rubbing that can be define with a help of Manning's equation:

$$S_f=\frac{n^2Q^2}{A^2R^{4/3}}, \qquad (3)$$

and g is earth acceleration, in ms^{-2}.

In addition, R presents hydraulic radius, Q is expiration in m^3/s; n is so called coefficient of roughness or Manning's coefficient.

Equations (1) and (2) can be meet and in the next form:

Equation for maitains of coninuity:

$$B\frac{\partial h}{\partial t}+\frac{\partial Q}{\partial x}=q \qquad (4)$$

Equation for maintains of the quantity of movment:

$$\frac{\partial v}{\partial t}+\alpha\cdot v\frac{\partial v}{\partial x}+g\frac{\partial h}{\partial x}-g\left(S_0-S_f\right)=0 \qquad (5)$$

Where B is width of the mirror-surface of the water in the canal in m, and α is a coefficient of distribute of speed by given transversely section. Then with h is shown the depth of the water in the canal in m, and the rest of the marks have already the same mining.

IV. Primary equations of continuity

Calculation methods connect with the transfer of hydrographs by canal and accumulations are based on principles for maintains of mass (equation for continuity) and maintains of the equational impulse in relation with the saving of the dynamic equality.

Equation of continuity it shows meaning of saving of the mass that means that the change of the accumulation is equal of the income of the water in her minus flowing.

$$I-O=\frac{dS}{dt} \qquad (6)$$

Where t is time in seconds, and S accumulation volume in m^3, volume contain between two sections of the canal. Then, I is income in m^3/s, he is constitute from incoming income and every one additional that enters in the canal between consider two sections, O is a flowing in m^3/s, in which is exit flow and every loss of water (for example infiltration over the bottom of the canal, deduction of the water in the canal for other needs and other types of deduction).

Equation for maintains of the quantity of movement, also, called dynamic equation, is actually same with the Second Newton low under which the change of the quantity of the movement is sum of all powers that act to the view volume. That is to say,

$$\frac{d(mv)}{dt}=\sum F \qquad (7)$$

Where with m is the mass in kg, v is speed in m/s, of movement of the water volume between two sections in the canal. The sum of the powers F in self considers the pressure, rubbing and gravitation conditions.

The transfer of mass includes in self solution that consider and the equation of continiuty and the equation of moments while are made surtin simplifaction.

V. Methods for transport of water by acumulations and canals

For assignment of the movment of the water canals and waterflow when there is a word for short time intervals (daily or shorter) are more methodes, and that:

- Straddle-Stagger
- Tatum
- Muskingum
- Modified Puls,
- Modified Puls in a fuction of income,
- Working R&D,
- SSARR Time-of-storage and
- Direct import of the coefficient of transport.

Part of these methods will be present in this paper.

- Modified Plus Method

This method uses the next form of equation of continuity:

$$\left(\frac{S_2}{\Delta t}+\frac{O_2}{2}\right)=\left(\frac{S_1}{\Delta t}+\frac{O_1}{2}\right)-O_1+\left(\frac{I_1}{\Delta t}+\frac{I_2}{2}\right) \qquad (8)$$

From where is:

$$\frac{I_1+I_2}{2}-\frac{O_1+O_2}{2}=\frac{S_2-S_1}{2} \qquad (9)$$

or

$$\underbrace{I_1 + I_2 + \left(\frac{2S_1}{\Delta t} - O_1\right)}_{\text{poznata strana}} = \underbrace{\frac{2S_2}{\Delta t} - O_2}_{\text{treba da se opredli}} \quad (10)$$

or generaly,

$$\underbrace{I_n + I_{n+1} + \left(\frac{2S_n}{\Delta t} - O_n\right)}_{\text{poznata strana}} = \underbrace{\frac{2S_{n+1}}{\Delta t} - O_{n+1}}_{\text{treba da se opredli}} \quad (11)$$

Where is known dependence $O_{n+1} = f(S_{n+1})$ in the table or analytical form.

- WORKING R&D METHOD

This method uses nonlinear dependence of the volume from the flowing of the water, similarly like the method called Modified Plus Method. Also, this method allows use of accumulations with transversal section in a form of wedge, similarly like the method called Muskingum method. It shows that under linear dependencies volume in function of flowing from accumulation this method gives same result with Muskingum method. If it's work for transfer of hydrograph by accumulations that don't have a form of wedge (coefficient of Muskingum is $x = 0$), this method gives results same as they are with the Modified plus method. Working R&D method use volume of accumulation that is called "worked" volume, that is mention in the explanation of the next equations:

$$\frac{R_2}{\Delta t} = \frac{R_1}{\Delta t} + Q_h - \frac{D_1 - D_2}{2} \quad (12)$$

In adition are:

R_1 - "worked" volume of end of a time interval;

R_2 -"worked" volume of the start of a time interval;

Q_h -everage income for time of one transfer interval;

D_1 -"worked" take over of the bigining of a time interval;

D_2 -"worked" take over of the end of a time interval, that is function of R_2 shown with dependence: worked volume in a function of flowing, $D_2 = f(R_2)$.

Flowing, again, is a function from "working" discharge and is define with a help of the next equation:

$$O_2 = D_2 - \frac{x}{1-x} \cdot (I_2 - I_1) \quad (13)$$

Where there,

x -coefficient of Muskingum, with out dimension;

O_2 -flowing of end of a interval;

I_2 -income of the end of a interval.

- METHOD OF COEFFICIENTS

All methods of coefficients flowings of the end of the canals conections are calculate like linear functions from flowings of water of the entrence in the canal, to wit:

$$O_n = C_1 \cdot I_n + C_2 \cdot I_{n-1} + C_3 \cdot I_{n-3} + C_4 \cdot I_{n-4} + ... \quad (14)$$

Where there are:

O_n -ordinate of hydrograph of flowing in time interval n

I_n, I_{n-1}, I_{n-2}... -ordinates of hydrograph of income in time intervals $n, n-1, n-2$..... appositely

C_1, C_2, C_3... -coefficients of movements, like coefficients of income.

- MUSKINGUM METHOD

Muskingum method can be understand like method of storing in canals elemental parts define like sum of padding in form of prisms and wedges. At the same time, connection of the exit padding with the entrance is linear. For defining of the coefficients C_1, C_2, C_3..., under this method are use recurrent equations.

This method ignors moments equations and is only firmly based by the equations for continuity. This method is especialy used when is worked for difuse waves. Under the transfer of given hydroraph his top is suficating like a result of difusion that comes with it that is caused by the efects of egsisting volume of water. In addition, is assumped that the volume of the water in the canal depend from the quantity of the water that enters in him and the quantity of water that exits from him and can be determine on the next modus:

$$Q = xI + (1-x)O \quad (15)$$

or

$$S = KQ \quad (16)$$

Where K is a time of travel in seconds between two sections of given canal, x is a coefficient without dimension, respectively load factor that value is between 0,0 and 0,5 with which is expressing of the enter canal and exit hydrograph from the canal by the volume of water contained in the canal. In addition, if $x = 0,5$ than the volume depend equally from the entrance and exit hydrograph.

$$I - O = \frac{dS}{\Delta t} \quad (17)$$

where I is entering, O - exit hydrograph, $\frac{dS}{\Delta t}$ change of volume, and its get (for example for the second time interval):

$$\frac{S_2 - S_1}{\Delta t} = \frac{\Delta S}{\Delta t} = K \cdot \frac{\Delta Q}{\Delta t} = \frac{Q_2 - Q_1}{\Delta t} = K \cdot \frac{(I_2 - I_1) + (1-x)(O_2 - O_1)}{\Delta t} \quad (18)$$

or with experiment of that is:

$$\frac{S_2 - S_1}{2} = \frac{I_1 + I_2}{2} - \frac{O_1 + O_2}{2} \quad (19)$$

is get:

$$\frac{I_1 + I_2}{2} - \frac{O_1 + O_2}{2} = K \cdot \frac{(I_2 - I_1) + (1-x)(O_2 - O_1)}{\Delta t} \quad (20)$$

if the equation (20) is solowed under O_2, is get:

$$O_2 = \frac{\Delta t - 2Kx}{2K(1-x)+\Delta t}\cdot I_2 + \frac{\Delta t + 2Kx}{2K(1-x)+\Delta t}\cdot I_1 + \frac{2K(1-x)-\Delta t}{2K(1-x)+\Delta t}\cdot O_1 = \quad (21)$$
$$= C_1 \cdot I_2 + C_2 \cdot I_1 + C_3 \cdot O_1$$

where there are:

$$C_1 = \frac{\Delta t - 2xK}{2K(1-x)+\Delta t} \quad (22)$$

$$C_2 = \frac{\Delta t + 2Kx}{2K(1-x)+\Delta t} \quad (23)$$

$$C_3 = \frac{2K(1-x)-\Delta t}{2K(1-x)+\Delta t} \quad (24)$$

In addition, must to be sutisfed the nezt equation:

$$C_1 + C_2 + C_3 = 1 \cdot \quad (25)$$

Analogously, flowing in the third time interval bringup:

$$O_3 = \frac{\Delta t - 2Kx}{2K(1-x)+\Delta t}\cdot I_3 + \frac{\Delta t + 2Kx}{2K(1-x)+\Delta t}\cdot I_2 + \frac{2K(1-x)-\Delta t}{2K(1-x)+\Delta t}\cdot O_2 = (26)$$
$$= C_1 \cdot I_3 + C_2 \cdot I_2 + C_3 \cdot I_1 + C_4 \cdot O_1$$

or generally,

$$O_n = \frac{\Delta t - 2Kx}{2K(1-x)+\Delta t}\cdot I_n + \frac{\Delta t + 2Kx}{2K(1-x)+\Delta t}\cdot I_n + \frac{2K(1-x)-\Delta t}{2K(1-x)+\Delta t}\cdot O_{n-1} = (27)$$
$$= C_1 \cdot I_n + C_2 \cdot I_{n-1} + C_3 \cdot I_{n-2} + C_n \cdot I_1 + C_{n+1} \cdot O_1$$

where which are:

$$C_1 = \frac{\Delta t - 2xK}{2K(1-x)+\Delta t} \quad (28)$$

$$CC = \frac{(2K(1-x)+\Delta t) - 2\Delta t}{2K(1-x)+\Delta t} \quad (29)$$

$$C_2 = C_1 \cdot CC + \frac{\Delta t - 2Kx}{2K(1-x)+\Delta t} \quad (30)$$

$$C_i = C_{i-1} \cdot CC \quad (\text{за } 2 < i \le n+1) \quad (31)$$

Surely, the sum of coefficients $C_i \ (i = 1, n+1)$ is one, id est.:

$$\sum_{i=1}^{n+1} C_i = 1 \quad (32)$$

In the upper equations parts of the index have the next meaning:

Δt -value of time under transfer of water mass;

K -time of travel shown in hours;

x -coefficient pf transfer without dimension that value is between 0 and 0,5.

With a prpuse to be avoid apperance of negativ coefficients in the equations (18) and (21) till (30) the coefficient of Muskingum it has to be satisfy the3 next condition:

$$\frac{\Delta t}{2(1-x)} \le K \le \frac{\Delta t}{2x} \quad (33)$$

I has to be said that x weight factor of importing of water in the canal. If is x=0, income of water in the canal doesn't efects the change of the volume of the water in the canal (like the all tank is moved). If is, again, x=0,5 income will have maximal efect and at the same time there wont be sufication of the exit of the canal (only transfer based on K hours).

It has to be said that under use of this method critical phase is qualification of the values for K and x. This values has to be determine in precise way on the used data's for valued enter and exit hydrograph for given canal section. If that is impossible it is recommended to be taken for x values in the interval $0,2 < x < 0,5$.

After determinaton of K and x for given canal section, the procedure for calculation for getting of the exit hydrograph is next:

1. It discreet the enter hydrograph with time interval with width Δt;
2. The coefficients are calculated $C_i \ (i = 1, n+1)$;
3. It uses the equation of Muskingum for deciding of exit hydrograph on the end of the canal;
4. Step 3 is repeat until all of his intervals of the enter hydrograph are done.

VI. CONCLUSION

In this paper are shown models for transfer of water between two or more hydropower plants. It had to be said that one same plum hardly that the condition can be made respectively one owner of all hydropower plants. From hire and the need of inclusion of the method for distribution of hydrographs or use of water potential, with it is made ideal solution for use of waters (includinc all the restrictions). That means that the technic documentation, practical on ever level (only, notional proect even and fisibility study) has to be isolated with one of a kind mathematical model. While that can be formed "timetable" of the hydropower plants and with smalest consuption of the water potential to be made bigest production of electrical energy.

REFERENCES

[1] Optimization of daily regimes of work of hydraulically connection hydro electrical plants, master's work. Skopje FEIT, 2007.

[2] Esbjorn Gullhamn, "Control of water content and retention in hydropower plant cascades", Master's Degree Project Stockholm, Sweden 2004.

[3] Nils Reidar B. Olsen, "Hydroinformatics, Fluvial hydraulic and limnology", Department of hydraulic and environmental engineering the Norwegian University of Science and Technology.

POSTER SESSION PO XIV

PO XIV - Education Quality

Electronic Identification of Completing the Evaluations of Students in a Database

Rosen I. Radonov[1], Konstantin A. Zaimov[2] and Valentin H. Videkov[3]

***Abstract* – The paper examines the developed application to the existing university information system, allowing to trace the completion of students' evaluations database. The developed application relates to an electronic protocol system allowing completion of evaluations by teachers. The basic functions of the electronic protocol and its relationship to the central database are described. The algorithm for tracing the evaluation completion is presented. Some sample results of the system's utilisation are given.**

***Keywords* – Computer systems, information systems, automation, education, quality**

I. INTRODUCTION

In the education system the primary evaluation of the process is carried out based on the success of the students. That is identified by evaluations of individual disciplines. In the standard case evaluation is carried out based on discipline and students' evaluations, and rarely on the teacher of the discipline. In practice there is no university which does not maintain an electronic database with such kind of information. A significant part of these systems support the possibility for an exchange over the Internet [1]. The better systems maintain complex information, linking data on discipline, teachers, curricula, etc. One of the advantages of these systems is the ability to do analysis on various criteria [2].

An update to the existing information system is made at the Technical University of Sofia. The system is of the type described above and enables the electronic completion of students' evaluations in the database and tracing the process of evaluation.

II. TUS' INFORMATION SYSTEM

The existing University Student Information System (USIS) is based on Oracle Database [3] and is a client-server windows desktop application. The system is designed to store data for university students and to provide basic references. It is designed to store data for university students and provide reference information. The system is integrated with a second system providing data from admission campaign. The system uses TCP/IP protocol over the existing intranet for connection with the database server. The initial window after the successful login into the system is shown in Fig. 1.

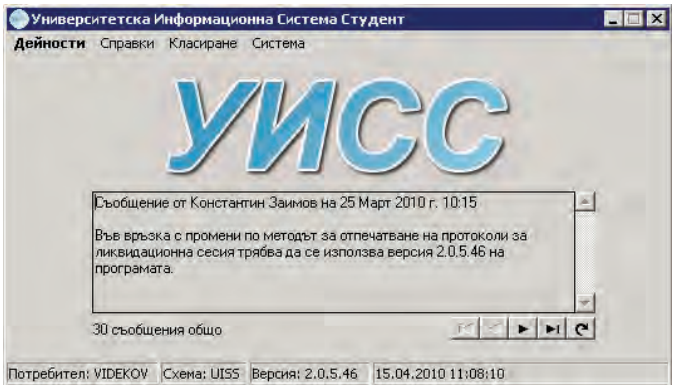


Fig. 1. The initial window of USIS.

The main sections in it are related to data input for individual students and retrieval of individual records for them.

In Fig. 2 is shown the main window with student's data. The personal data such as name, address, previous education, and more are displayed. The current status – certified semester, status, academic year and others are also displayed. Pages with other information can be opened from the navigation bar. They are related to health insurance, student status, academic curriculum in which learning takes place, etc.

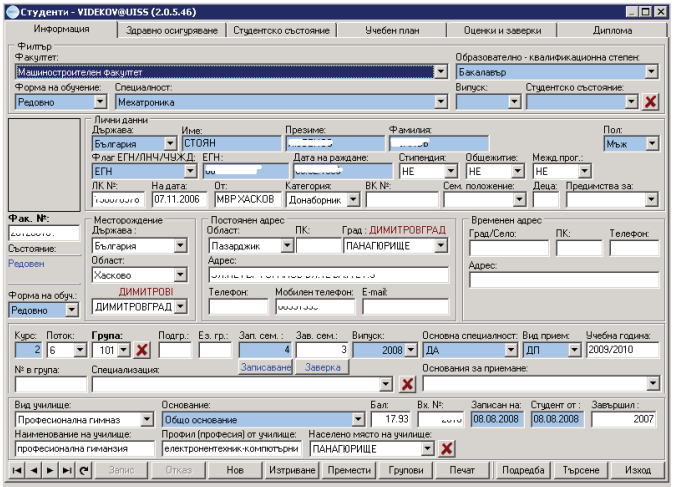


Fig. 2. Main windows – student's data.

[1]Rossen I. Radonov is with the Faculty of Electronics, Technical University of Sofia, 8 St. Kl. Ohridsky blvd, 1797 Sofia, Bulgaria, E-mail: Rossen.Radonov@ecad.tu-sofia.bg

[2]Konstantin A. Zaimov is with Information Services PLC, 2 Panayot Volov Street, Sofia 1504, Bulgaria, E-mail: K.Zaimov@is-bg.net

[3]Valentin V. Videkov is with the Faculty of Electronics, Technical University of Sofia, 8 St. Kl. Ohridsky blvd, 1797 Sofia, Bulgaria, E-mail: Videkov@ecad.tu-sofia.bg

A window displaying information from the page with student's evaluations in each discipline is shown in Fig. 3. It shows the name of the discipline, semester in which the study is carried out, the evaluation, the type of evaluation, etc.

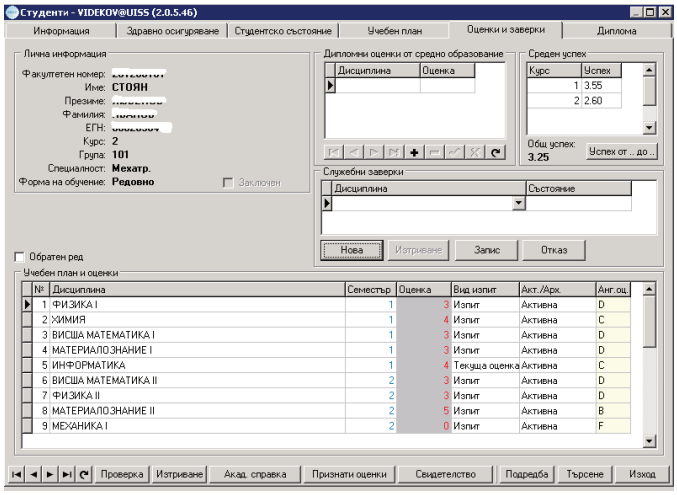


Fig. 3. Evaluations and certifications.

III. TASKS

In any management system not only the outcome is important (good or bad), but also the process of achieving that result. To achieve good results and correct management decision the process must be traceable [4]. This approach is not implemented in the current procedure for completion of the evaluation database.

What could happen and what is desirable to be known? Each system could produce an error while functioning. One simple error is to fill in a wrong evaluation. The first interested party in this case is the student. There is a possibility for the students to check their evaluation via the Internet using their personal code and faculty number. In the case of a higher evaluation it is quite possible, in practice, that there will be no feedback from the students.

In the opposite case, the employee in the student's office will inform the teacher and he or she will correct the error. This process, however, is not identified in the electronic database. Another thing is that between the evaluation in a student's e-file and the teacher stands in the office employee and that to some extent blurs the responsibility for errors.

There was no identification of the evaluation's grounds and origin in the old system – fig. 4.

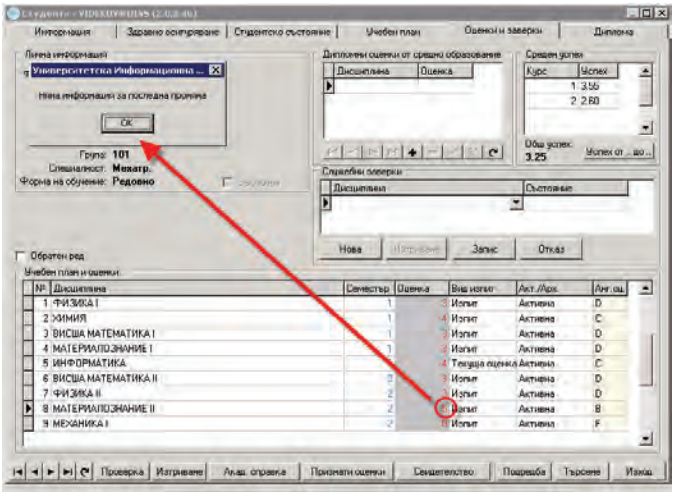


Fig. 4. Identification of the evaluation.

It is necessary that the completion of the evaluation database is not only an obligation but also a responsibility of the teachers. It is also necessary to do the completion within controlled deadlines and document the process in the timeline.

The process of error correction also has to be documented, not only because of the responsibility, but also due to the need of obtaining statistic data about errors possibility..

Having in mind the above-mentioned, the tasks are to develop a system for electronic completion of evaluation database by the teachers allowing the process to be documented and identify the origin of the evaluation.

IV. THE ELECTRONIC PROTOCOL SYSTEM

The requirements to the system are derived from the tasks to be completed. It has to:

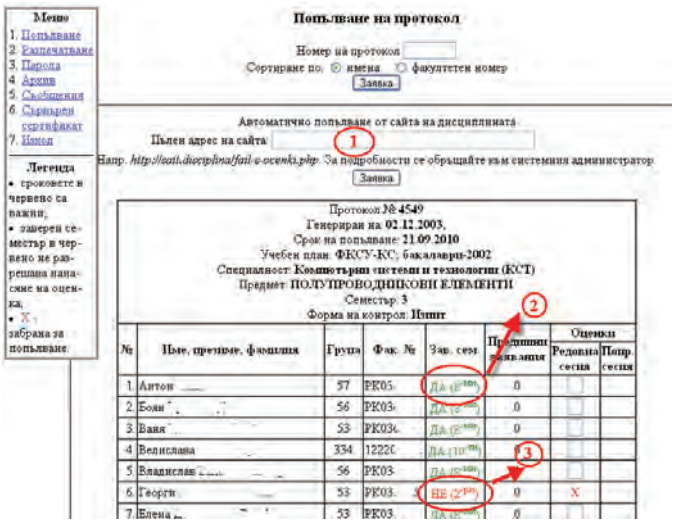
- ensure capability for electronic completion of the database with evaluation for each discipline by the relevant teacher;
- ensure protection from records manipulation;
- give information about the evaluated student to the teacher;
- identify corrections.

These requirements resulted in the development of an Internet site for completion of the evaluation database by the teachers.

The widest range of possibilities is provided by the option for using an arbitrary computer connected to the Internet. In this case special measures have to be taken in order to ensure protection of the database.

A 4-tier security system is used – username, password, digital signature and PIN code for accessing the digital signature from the USB flash memory. The digital signature can only be read from the USB flash memory.

After the successful login to the system the teacher is granted access to the menu – Fig. 5. The electronic protocol system is linked to another system – the e-management system. The last one provides automatic transfer of the evaluations from the site of the discipline to the electronic protocol – position 1 in Fig. 5.

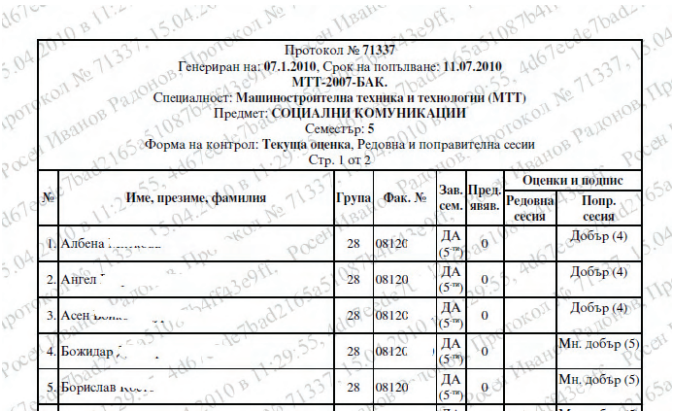


Фиг. 5. Electronic protocol – completion of the evaluations.

The status of the student is being retrieved in real time and the relative information about the last certified semester is displayed (position 2 in Fig. 5). If the student does not have the necessary certification the electronic protocol system does not allow completion of the evaluation. (position 3 in Fig. 5)

The teacher can edit the protocols as many times as he/she wants within the denoted period of the respective examination sessions. After completion the evaluation protocol is marked as ready and submitted electronically. The protocol is automatically being marked as ready by the system after the expiry of the deadline for submission.

After the submission of the protocol the teacher prints and signs it, and submits the paper copy at the respective student's office. The printed copy has a watermark which represents checksum and other data specific to the protocol – Fig. 6.


Протокол № 71337
Генериран на: 07.1.2010, Срок на попълване: 11.07.2010
МТТ-2007-БАК.
Специалност: Машиностроителна техника и технологии (МТТ)
Предмет: СОЦИАЛНИ КОМУНИКАЦИИ
Семестър: 5
Форма на контрол: Текуща оценка, Редовна и поправителна сесии
Стр. 1 от 2

№	Име, презиме, фамилия	Група	Фак. №	Зав. сем.	Пред. явяв.	Оценки и подпис: Редовна сесия	Попр. сесия
1.	Албена	28	08120	ДА (5тий)	0		Добър (4)
2.	Ангел	28	08120	ДА (5тий)	0		Добър (4)
3.	Асен	28	08120	ДА (5тий)	0		Добър (4)
4.	Божидар	28	08120	ДА (5тий)	0		Мн. добър (5)
5.	Борислав	28	08120	ДА (5тий)	0		Мн. добър (5)

Fig. 6 Paper copy of the protocol..

In some special cases the evaluation can be completed by an authorized person (e.g. an employee at the student's office). When browsing the evaluations the name of this person, protocol number, date and time are displayed – Fig. 7.

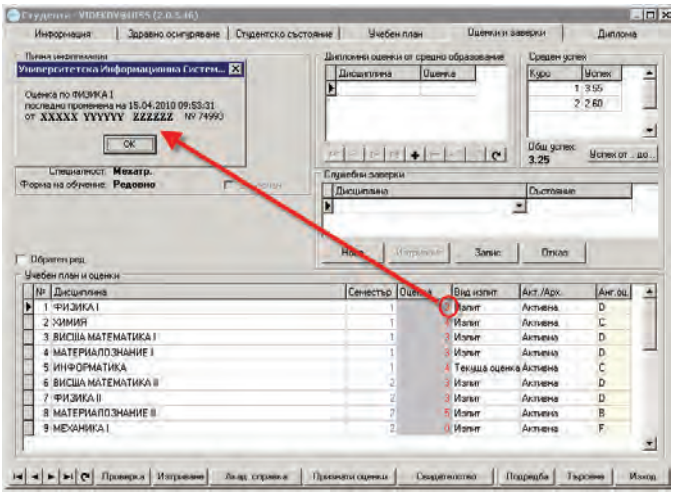


Fig. 7. The evaluation is completed by an authorized person.

When the evaluation origins from an electronic protocol, which has been completed by a teacher, only its number, date and time are displayed – Fig. 8.

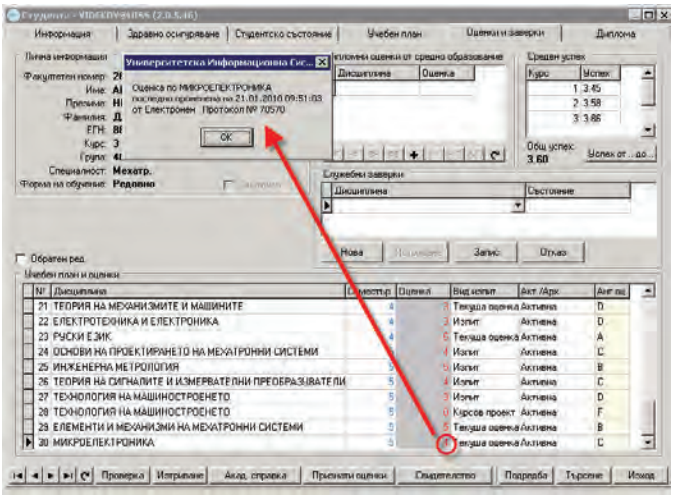


Fig. 8. The evaluation is from an electronic protocol.

The introduction of the electronic protocol system makes it possible to visualise the process of students' evaluation . Thus for example, the process of electronic protocols submission by the teachers can be traced. Fig. 9 shows the dynamics of the submission.

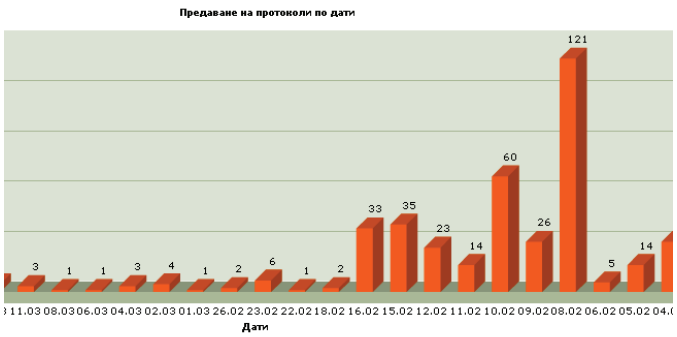


Fig. 9. Number of electronic protocols submitted per day.

The errors during completion of the database are another indicator of the evaluation process. Since the completion of the electronic protocol is the immediate responsibility of the teacher, errors could be the result of technical implementation, for example. The procedure for handling the electronic protocol is such that after submission of the protocol no changes are possible. After finding an error it has to be corrected by a separate correcting protocol. An interesting statistics can be made on the time to detect the errors after the protocol's submission and the distribution of these errors. An indicative sample is shown in Fig. 10, which shows the percentage allocation for correcting protocols as a function of time after submission of the basic protocol.

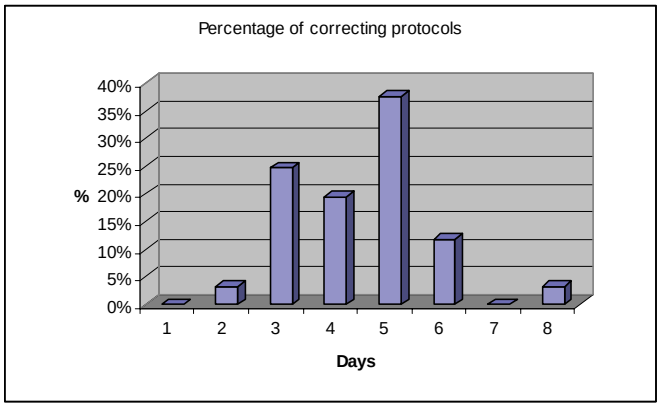


Fig. 10. Distribution of correcting protocols over time.

The results of preliminary tests of the system showed that the number of correcting protocols is less than 8% compared to the basic ones and as they concern a single evaluation their number is relatively small. In a database with over 11000 evaluations the number of corrections is below 40 or below 0.4 %. Perhaps with the acquisition of habits of teachers, this percentage will be reduced.

V. Conclusion

The Technical University in Sofia has developed an Electronic Protocol system relating to the existing database system for students that allows electronic completion of the evaluations by teachers. That speeds up the process of evaluation and the most important is that the responsibility for this process belongs to the evaluators. Its implementation makes it possible each evaluation to be identified by the time, person and on what basis is done.

A second important point is the ability to monitor over time the process of evaluating the different disciplines with the probability of errors in evaluation.

Acknowledgement

This research has been carried out in the framework of Contract Application No. 102ни196-3.

The authors express their gratitude to the USIS support team and employees at the student's offices at the faculties of Electronic Engineering, Machine Engineering and the directions in the first year of study on the generation of experimental protocols for the implementation. We thank our colleagues – the teachers, who completed the electronic protocols, as well as the authorities of the Technical University of Sofia, who supported the launch of the system.

References

[1] http://student.tu-sofia.bg – date of last access 30 March 2010.
[2] http://www.sap.com/solutions/sapbusinessobjects/ondemand/BI/index.epx - date of last access 2 April 2010.
[3] http://oracle.com - date of last access 16 November 2009.
[4] Chryssochoidis G., A. Karagiannaki, K. Pramatari, O., A cost-benefit evaluation framework of an electronic-based traceability system, British Food Journal, 2009, Volume 111, Issue 6, pp. 565 – 582, ISSN 0007-070X

Laws for E- Learning and Copyringhts

Elena Racheva[1], Slava Yordanova[2] and Mitko Mitev[3]

***Abstract* – In this paper consider the problems with Laws for E-Learning.**

Keywords – Laws, E-learning, Copyright, TU–Varna, department "Computer Sciences and Tehnology".

I. INTRODUCTION

The development of technologies and the increasing dynamics of our lives impose new demands for education and new methods of its teaching. Interest towards the means that make E-learning possible is high both from educational institutions as well as from large-scale companies and organizations from the public sector. Remote or e-education is carried out not only in the field of IT, where it's a must, but in many other fields. From different kinds of courses, on-line tests for acquiring international certificates to work with bank systems. It's undeniable that now-a-days the expenses of teaching are greatly reduced by using E-learning systems. Department "CST" is a member of the virtual department of computing "Jon Atanasov", which grew into a virtual faculty for informational and communication technologies. [1], [5]

Department "CST" carries out a series of actual projects for implementing e-education and development of its informational infrastructure. Articles on the topic of electronic education and its use in TU-Varna and more specifically in department "CST" were read at international conferences.

Department "CST" utilizes an HP ML530 server, developed by the Russian university "Ангел Кънчев", as a base for E-learning. The system uses a MySQL database server and a PHP script language for a web interface. It has the following capabilities:

1. Indicates annotations of subjects;
2. Shows school programs;
3. Lists lecturers and ways to contact them;
4. Presentation of lecture materials;
5. Presentation of lab materials and exercises;
6. An advertisement board;
7. A schedule for classes;
8. A forum for discussion;
9. Internal e-mail organization;

II. PROBLEMS

One of the major problems of E-learning is copyrighting.

The materials used in remote teaching can be classified as follows:

- literature;
- printed works;
- art;
- sound recordings;
- and others.

Each of the above categories is used in E-learning. There are rules following which the owner of the copyright can be found and its duration determined. [1], [2], [3]

The materials included in E-learning are precious mainly because they are an investment of time and money. Hence it is necessary that measures are taken against unregulated copying and misuse.

When receiving copyrights to materials for E-learning the following must be taken into consideration:

1. Where the product will be released – regions, one or more countries;
2. Copyright deadline;
3. Ways of utilizing the product – partly or the complete product;
4. Regulations needed for the distribution of the materials;
5. Stating the rights to sell the materials.

To avoid violation of copyrights when creating new applications for E-learning it is necessary to follow these rules:

1. Following the laws of the corresponding country or countries if it is released in more than one country;
2. Familiarizing the faculty with the laws regulating the copyrights;
3. Some educational institutions can use their own materials for educational purposes. They can publish materials with limited use for students and if those are used for marketing purposes, permission must first be gained. It is recommended that educational materials be protected before being distributed;
4. Juridical help may be required.

A number of institutions work on the problems if intellectual ownership. [1], [4]

III. DATA PROTECTION

Stages in data protection:

- encoding of the contents;
- adding additional information;
- personalization of the file;

Good copyright protection is provided by using MMP protocol (Multimedia Protection Protocol). This

[1]Elena B. Racheva is with the TU –Varna, Faculty of Computer Sciences and Tehnology, Studentska 2, 9010 Varna, Bulgaria, E-mail: elracheva@yahoo.com

[2]Slava M. Yordanova is with the TU –Varna, Faculty of Computer Sciences and Tehnology, Studentska 2, 9010 Varna, Bulgaria, E-mail: slava_y@abv.bg

[3]Mitko M. Mitev is with the TU- Varna, Faculty of Computer Sciences and Tehnology, Studentska 2, 9010 Varna, Bulgaria, E-mail: mitevmm@abv.bg

protocol guarantees protection of copyrights of all types of digital data. We are showing you an example of the contents of an MMP encoded file:

Name	Contents
of the file	to the MMP files
Block size	Coded : non coded information ratio
Distributer	of the distributer
Title	Title text
Author	Text containing the author's name

IV. Conclusion

E-education isn't an alternative to traditional education, it completes it. E-education realizes the idea of an affordable and easily adaptable to a person's individual needs education that lasts a lifetime. Confronting the fast changes in technologies and marketing, Bulgaria comes face to face with this new form of education, which doesn't require a significant increase in both an individual family's budget and the country's budget.

References

[1] Sl. Yordanova, V. Nau;fv, G. Kunev, "Multimedia and E-learning", Bulgaria, Sofia, 2000B. Milovanovic, Z. Stankovic, S. Ivkovic and V. Stankovic, "Loaded Cylindrical Metallic Cavities Modeling using Neural Networks", TELSIKS'99, Conference Proceedings, pp.214-217, Nis, Yugoslavia, 1999.

[2] European Commission. Education and Training. http://ec.europa.eu/education/archive/elearning/index_en.html

[3] http://www.e-learningcentre.co.uk/

[4] http://elearningtech.blogspot.com

[5] R. Steinmetz, K.Nahrstedt, "Multimedia& computing, communications and applications", Prentice Hall, 2000

Development in Education of Electrical Measurements and Circuits Theory

Rosen N. Vasilev[1], Ivaylo Y. Nedelchev[2], Vyara Y. Vasileva[3] and Miroslava G. Doneva[4]

***Abstract* – In the paper the authors present the possibilities of preparing modern, educational tools which, in the form of virtual instruments, are very helpful for educational process concerning electrical measurements and circuits theory. These dynamic instruments make it possible to use during the lectures some demonstrations which better explain the presented theory. Main platform is NI ELVIS II, leading educational platform, based on graphical orientated software for programming – LabVIEW. The virtual instruments make it possible to simulate measuring sessions realized later in the university laboratory.**

***Keywords* – Education, NI ELVIS II, LabVIEW, Electrical Measurements, Circuits Theory.**

I. INTRODUCTION

The dynamic development of computer technology requires dissemination and implementation of new information technologies in education, combining visibility and simplicity of statement with a high level of material studied. Central place is occupied by modern educational resources as electronic educational centers and methodological tools, computer models of the studied objects and processes, software and hardware automation of laboratory facilities. Large number of publications in recent years [1,2,3,4,5] seen a revolutionary development of the educational process, especially for distance tuition and for anyone who wishes to read lectures in various courses through the Internet.

For this tuition are very useful virtual instruments that better explain the contents, allowing the creation of various applications and implement new priorities in engineering education based on ability to independently solve the tasks of scientific methods. Certainly they should not be treated as a substitute for actual laboratory tasks that are required for technical training but can be used as excellent tools to prepare for practical laboratory work. In Bulgaria, this process is still in its beginning.

[1]Rosen N. Vasilev is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: rsnvasilev@yahoo.com

[2]Ivaylo Y. Nedelchev is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: ivonedelchev@yahoo.com

[3]Vyara Y. Vasileva is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: via_vas@abv.bg

[4]Miroslava G. Doneva is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: m_grisheva@abv.bg

II. STATEMENT

The Department of "Electrical engineering" of Technical University-Varna develops and implements the concept of using education platform ELVIS II of National Instruments company, LabVIEW-based graphical environment such as teaching subjects "Electrical Measurements" and "Electrical Engineering". As a rule developed educational and methodical complexes include:

- Preparation of lectures in electronic form for direct use of the Internet browser. They contain animations of theoretical models, virtual tools for solving physical problems and issues of self.
- Preparation of experimental laboratory tasks in the form of virtual instruments that can be used as preparation for real practical tasks.
- Preparation of electronic textbooks with tasks, sample solutions and rules for the layout.
- Organize a system for remote connection with opportunities for contacts and consultations with teachers.

It is up to each laboratory exercise to include the following elements:

- Admission to the program tasks, checking knowledge necessary for understanding the experiment implementation and correct interpretation of results.
- Program for visualization of the work assignment and supervision of its implementation.
- Program for visualization of experimental data obtained from real circuit or mathematical model.
- Program to help fill the experimental results in the form.
- Program to detect and display the presence of gross errors, possible reasons for their occurrence and methods to solve the problem.
- Program guidelines and methodical recommendations on the tasks and design of the report.
- Textbook program, ensuring students theoretical knowledge on the subject.

It is recommended that the use of computer labs based on virtual instruments in two modes. The first mode is basic. In him take place lectures laboratory exercises. The second mode is cognitive training featuring some of the basic restrictions and simplifications. It is possible in preparation for the laboratory exercise and it is appropriate to have Internet access. While the basic program mode can be chosen by the instructor to conduct an experiment on actual experimental setup or its mathematical model, the training mode only works with the model.

Schemes, the element parameters, issues and tasks formulated in the two versions are not identical. The basic system program must be given access only to the teacher of the realized task reports, including both the database and the results are received in their treatment. In the training version of this protocol is not needed.

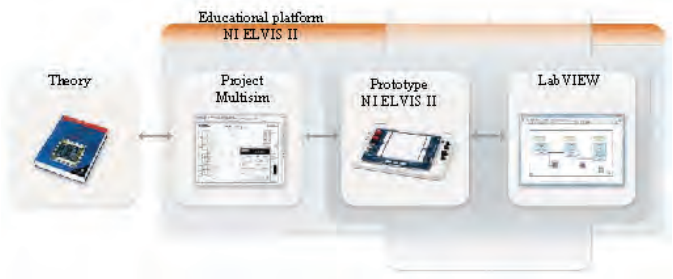


Fig.1. Educational platform NI ELVIS II

Above, at this stage is just an idea because the development of integrated hardware software complex, satisfying all the mentioned requirements is expensive process, requiring more skilled staff and additional equipment mostly global university policy. In practice, currently, the department implements the separate computer laboratory exercises based on ELVIS II, which allows the use of computer programs and hardware resources.

The essence of the educational platform NI ELVIS II is visualized in Fig. 1.

Fig.2. ELVIS II integration with LabVIEW

Its distinguishing feature is the tool of about 12 most commonly used measuring devices integrated into one compact and reliable platform providing a practical purpose of the educational process:

- Arbitrary Waveform Generator;
- Bode Analyzer;
- Digital Reader;
- Digital Writer;
- Digital Multimeter – DMM;
- Dynamic Signal Analyzer – DSA;
- Oscilloscope – Scope;
- Two-Wire Current Voltage Analyzer;
- Three- Wire Current Voltage Analyzer;
- Variable Power Supplies.

ELVIS II e closely integrated with the software Multisim 10.1 - leader in the design and simulation of electrical schemes. Students can simulate the operation of electrical circuits by Multisim to design them using the ELVIS II and explore the simulation model with real measurements through LabVIEW - Fig.2.

The ELVIS II tools are associating with express functions of LabVIEW – Express VI. On Fig.3 is depicted the function palette.

Fig.3. ELVIS II integration with Express VI

Express functions allow development of applications in an environment of LabVIEW, performing in interactive mode configuration of each instrument without a thorough knowledge of programming. With shown integrations educational platform ELVIS II allows the creation of fully ended process - the creation of a simulation model in the construction of prototype measuring the actual signals and parameters by comparing simulation and performing complex analysis - fig.4.

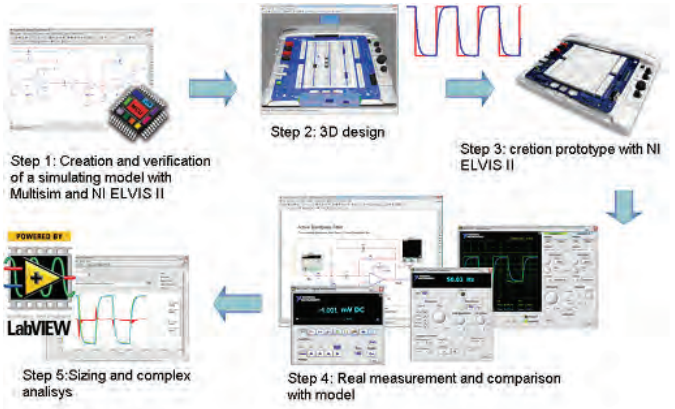


Fig.4. ELVIS II educational process diagram

Fig.5. The electrical circuit creating and analyzing.

In subjects “Electrical Engineering” and “Electrical Measurements”, students engage in examination of the basic laws in electrical engineering, creating loops, their control,

and measurement of different electrical parameters and comparing actual measurements with simulated data of the same instrument - fig. 5 and fig.6.

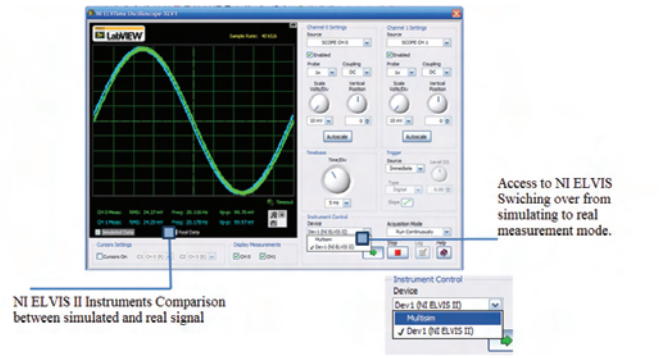


Fig.6. The electrical quantity measurement and comparing with simulating (pilot) signals.

Virtual laboratory stands are very useful for students, and they enable to practice at home the measurements technique. They are equipped with virtual instruments similar to these used in the university laboratory. The virtual stand in Fig. 7 presents principle of temperature measurements using a platinum resistance thermometer Pt100/138,5 and a thermocouple J (iron-constantan). They measure the temperature of liquid in the tank equipped with an electrical heater controlled by on/off switch.

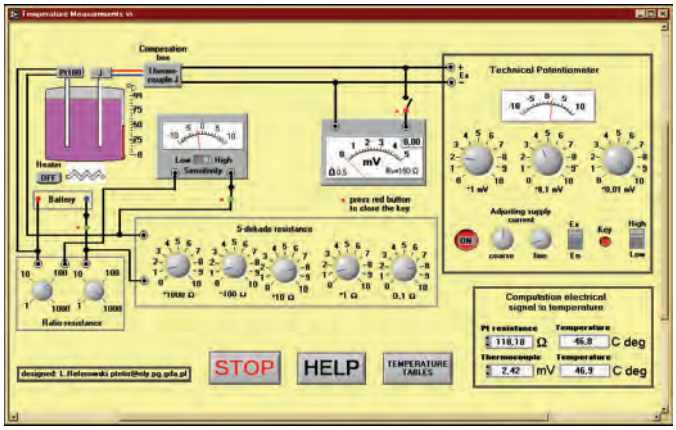


Fig. 7. Virtual stand for temperature measurements using Pt100 resistor and J thermo-couple

The resistance of platinum thermometer is measured by Wheatstone bridge. The bridge is balanced by adjustment 5-decade resistor after choosing the right values of ratio resistance. The key switch activates the galvanometer. The switch Sensitivity changes the galvanometer's sensitivity from low to high level.

The thermo-electric EMF of the thermocouple is measured by means of technical potentiometer, which is fed by an internal dry cell. The proper value of feeding current is adjusted with the help of internal standard Weston cell (Mode switch in position En). The feeding current is controlled using knobs coarse and fine. The measurement of the thermo-electric EMF of thermocouple is realized with mode switch in position Ex. The adjusted values of decade resistor at the balance state of the potentiometer determine the value of measured thermo-electric EMF in mV. The compensation box with compensation circuit between thermocouple and potentiometer compensates the influences of ambient temperature giving right value of signal for ambient temperature 0 ºC.

The influence of the output resistance of the measuring circuit on the value of measuring signal is can be shown by switching a voltmeter into the thermocouple circuit. Te value of output signal diminishes. The right value of signal can be calculated knowing the resistances of voltmeter (150 Ω) and thermocouple with leads (2 Ω).

The values of measured resistance and thermo-electric EMF are transformed into temperature with the help of standard temperature tables $R_\vartheta=f(\vartheta)$ and $E_{\vartheta\vartheta_o}=f(\vartheta-\vartheta o)$ activated by TEMPERATURE TABLES switch. It is also possible to make this calculations using Computation of electrical signal to temperature circuit. The HELP button activates window with fundamental information of measuring circuits and measuring procedure.

Fig.8 presents the front panel and the block diagram of a laboratory exercise for temperature measurement.

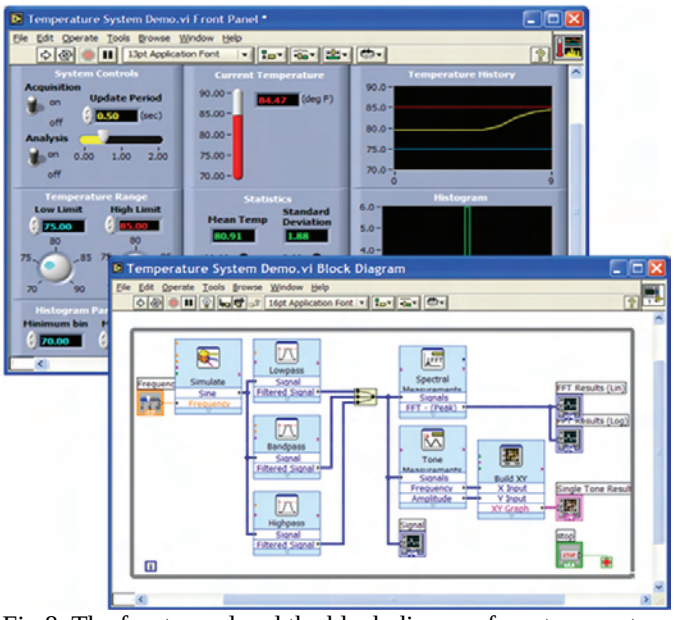


Fig.8. The front panel and the block diagram for a temperature measurement exercise

III. Conclusion

In the opinion of the authors, the advantages of application in laboratory practice of measuring complex NI ELVIS are determine by the obvious factors:

- conversion of measurements from separate points of functional dependencies (features or functions of the circuits) to the automated determination in a given range;

– possibility to use while performing a laboratory exercise in work on a comprehensive set of twelve software-implemented hardware measurement devices and hundreds NI virtual instruments libraries;

– use of such virtual tools whose hardware implementation is impossible for mass use in educational laboratories;

– the possibility of using self-developed virtual tools in the LabVIEW environment.

At the same time, NI ELVIS complex use, store all the features of the real experiment, namely:

- ability to verify conformity of theory with the real objects of interest;
- investigation the influence of parasitic parameters and deviations of measurement results;
- training in the creation of electrical circuits and systems, and instrumentation regular connection.

Despite the existing circumstances and difficulties in implementation, education platform ELVIS II is the one of the most promising instruments for the tuition organization in university education.

ACKNOWLEDGEMENT

The paper was developed in the frames of the Research Program, financed from Ministry of Education.

REFERENCES

[1] L. Referowski, D. Swisulski, “Progress in Education of Electrical Measurements”, Electronics and Electrical Engineering.,vol. 67, no. 3, pp. 25-30, 2006, ISSN 1392-1215.

[2] D. Hofmann, Measurement, Instrumentation & Test Management MTM, a Newcomer to Education & Training // Symposium “Virtual and Real Tools for Education in Measurement” Twente (NL), 2001, P. Regtien (Ed), University of Twente. ISBN 90.365.1664.1. - pp.15-22.

[3] D. Schmid, Virtual Control Laboratories and Remote Experimentation in Control Engineering // 11th EAEEIE Conference, Ulm (D), 2000, M. W. Hoffman (ed), Abteilung Mikrowellentechnik der Univertsitaet Ulm. ISBN 3 00 005965 2. - pp. 213-218.

[4] В. Лебедев, М. Марков, „Разработка лабораторного практикума по курсу ТОЕ”, 7-я международная научно-практическая конференция „Образовательные, научные и инженерныеприложения в среде LabVIEW и технологии National Instruments” 4 секция, pp.171-173, Москва, 20-21 XI, 2008.

[5] M. Halaj, E. Kurekova, P. Gabko, Multimedia Tools for Education in Measurement and Metrology // Symposium “Virtual and Real Tools for Education in Measurement”, Twente (NL), 2001, P. Regtien (Ed), University of Twente. ISBN 90.365.1664.1. - P. 3-13.

Examples for Virtual Laboratory Exercises by Electrical Measurements Based on NI ELVIS II

Rosen N. Vasilev[1] and Ivaylo Y. Nedelchev[2]

***Abstract* – ELVIS II is leading education platform, based on graphic programmable software LabVIEW. ELVIS II allows education on different technical subjects, in particular by "Electrical measurements". Technical devices and programmable software of the main kit of tools are combined in one structure. In this report are represented examples by laboratory educational tasks on "Electrical measurements". These developments consist of two main components: interactive device for control and LabVIEW virtual instrument for programming functions and control of the workstation.**

***Keywords* – Education, NI ELVIS II, LabVIEW, Electrical Measurements, Laboratory Exercises.**

I. Introduction

During last few years development of the computer technology concerns also education process. In this relation, one present decision is integrated educational platform NI ELVIS (Educational Laboratory Virtual Instrumentation Suite). This educational station is basic decision for development and creation of different laboratory exercises and educational laboratories.

The main character for NI ELVIS is its versatility. It gives possibility of the first knowledge for measurement devices and basic laws of the electrical circuits, connection methods and principle of operation, as well as it gives opportunities for students to design different analog and digital systems, to simulate prototypes of systems and devices in visual modeling environment Multisim 9, realization of the prototype over real platform and its own research with the help of integrated virtual instruments in ELVIS. In this way, students can get unique chance to pass the whole cycle of education – from modeling, through synthesis, till test real prototype over integrated platform.

The platform consists of:

1. Plate for modeling (with 2800 terminals) for independent development from students or teachers of the electrical circuits and patterns, fitting sensors and control mechanisms of the schemes. Generated signals, can be sent to the data save unit.
2. Laboratory station NI ELVIS II with versatility card for saving data PCI, for coordination of the signals, applied from data save device within the plate for modeling.
3. Virtual software, consist of 12 instruments, which are most used in the electrical measurements – oscilloscope, digital multimeter, spectrum analyzer, functional generator, variable power supplies etc.
4. Software Multisim 9.
5. Necessary drivers and group of some useful examples for LabVIEW.
6. Personal computer or laptop.
7. Multimedia projector.

[1]Rosen N. Vasilev is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: rsnvasilev@yahoo.com

[2]Ivaylo Y. Nedelchev is with the Faculty of Electrical Engineering, TU of Varna, 9010 Varna, Bulgaria, E-mail: ivonedelchev@yahoo.com

II. Statement

Virtual instruments from soft panel (fig.1) of the platform are used for conducting exercises with NI ELVIS II. In discipline "Electrical measurements" in Technical University of Varna are developed complex laboratory exercises, some examples of which are discussed in the present paper.

Fig.1. Soft panel of NI ELVIS II

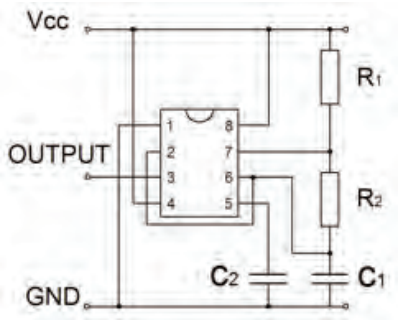


Fig.2. Connection diagram of timer LM555

In the first example are investigated quality of pulses generated from timer LM 555, upon frequency and time domain. Fig.2 shows connection diagram of the timer, and Fig.3 the real connection of the elements over platform NI ELVIS II. It is generated signal with frequency about 11 kHz. Frequency defining components are: R_1, R_2 and C_1. They determinate and the duty coefficient, which can be represented with equation:

$$K=T_1/T, \quad (1)$$

were T_1 is duty period and T is the whole time period of the signal. Using potentiometers in this positions, duty coefficient and frequency can be altered in wide borders. For research the square impulse, generated from timer, are used oscilloscope (SCOPE) and digital spectrum analyzer (DSA) from the soft

panel. In this way, by measuring upon time and frequency domain its easily to see distortion in the shape of generated square signal. Mathematically, this rectangular impulse can be represented by Fourier series:

$$f(\omega) = \frac{4A}{\pi} \sum_{n=1}^{\infty} \frac{1}{2n-1} cos((2n-1)\omega) \qquad (2)$$

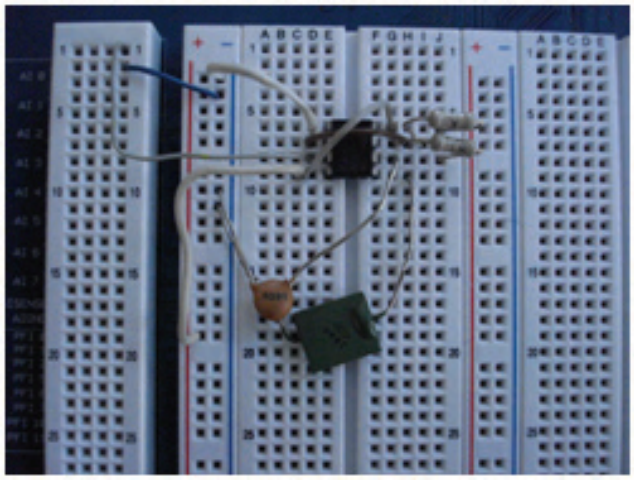


Fig.3. Screenshot of real connection of the timer

In other words signal is composed from odd harmonics with fading amplitude. Fig.4 and Fig.5 shows measurements with the SCOPE and DSA. Diagram from Fig.4 shows real form of the generated impulses.

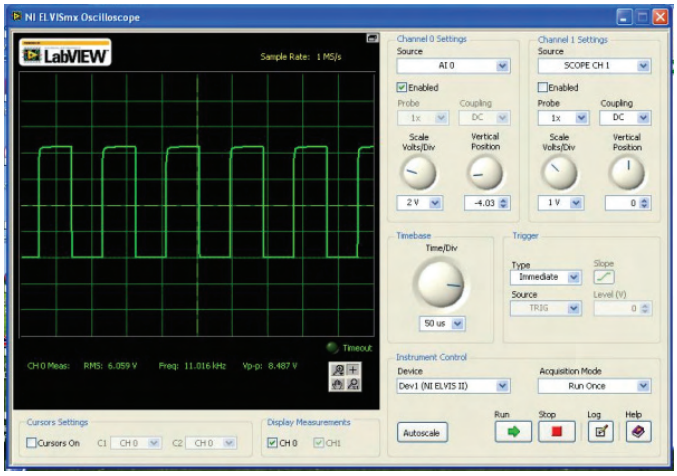


Fig.4 Time domain diagram

Asymmetry at the front of the pulse is due to transient process in the scheme and when we look carefully to the spectrum of signal, can be noticed presence of high order odd harmonics (from 15-th by 19-th) with high level and also presence of even harmonics, which shouldn't be there if consider mathematically. With appropriate filters can be find out that high order harmonics, has influence on the fronts of the impulses and low order - on the stability of the levels. Changing duty coefficient and frequency, students can meet different kinds of square signal and their spectrums, as well as the potentialities of the timer LM 555.

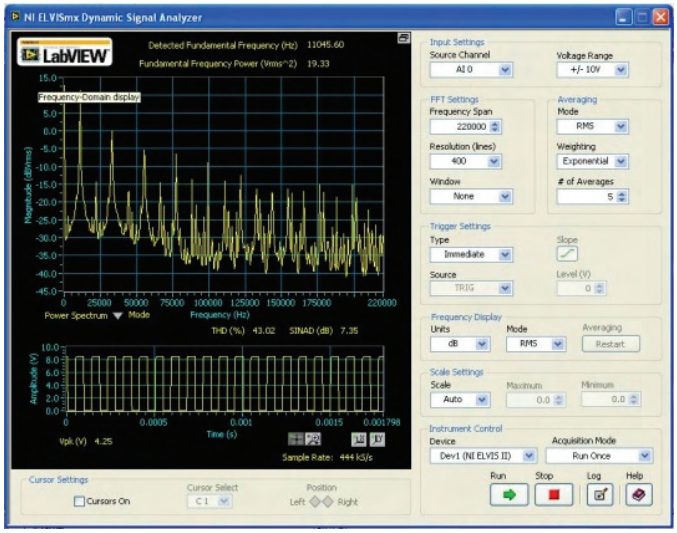


Fig.5 Frequency domain diagram

Second example shows measurement of voltage by method of comparison with the help of the compensator. Wiring diagram (Fig.6) is well known from the theory, and its realization by NI ELVIS II is shown on the Fig.7.

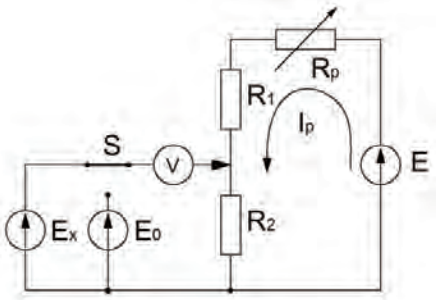


Fig.6. Diagram of the compensator.

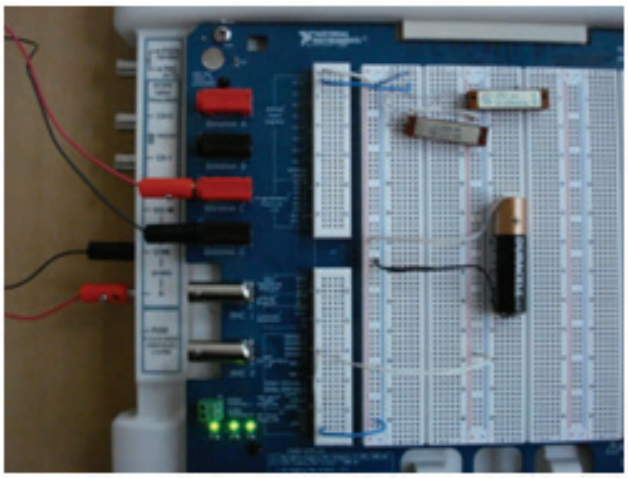


Fig. 7. Screenshot of real connection of the compensator

For zero indicator is used virtual oscilloscope (SCOPE), and by DMM, can be measured working current Ip in the compensator's circuit. Standard voltage E0 is applied from variable power supplies of the platform (VPS), which is known near hundredth, but this can be improved using real

standard element providing voltage with a high precision. First one of the helical potentiometers (Rp) is for adjustment of working current (Ip) and other one is a voltage divider (R_1, R_2) for balance the compensator during measurement. In this case is shown measurement the voltage of a small battery. After balance of the compensator SCOPE shows zero volts (Fig.8), with the help of the virtual instrument for impedance measurement (IMPED) is read the resistance of divider (R2) (Fig.9).

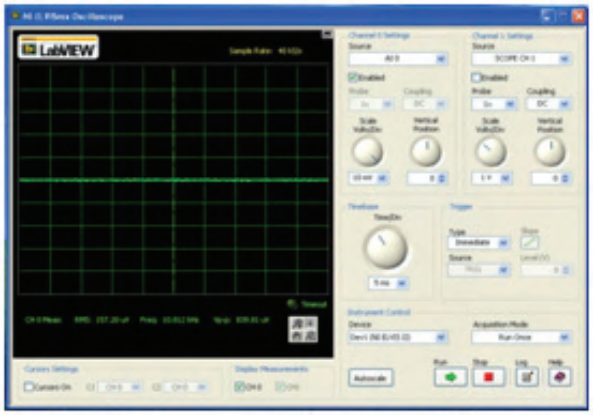


Fig.8. Reading of the Zero Indicator

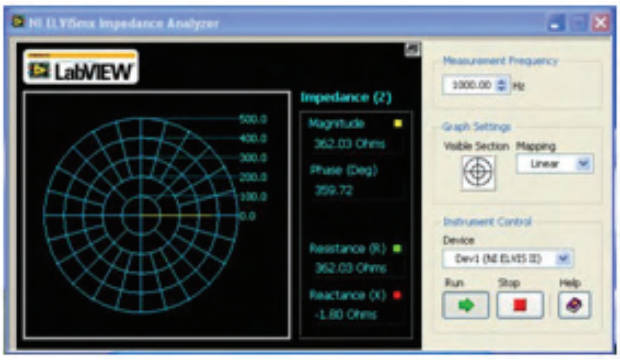


Fig.9. Reading from the IMPED

Unknown voltage can be calculated with:

$$U = Ip.R_2 \quad (3)$$

Another example for usage this device is for calibration of electromechanical voltmeters, but this case is possible if is used more accurate standard source instead of VPS.

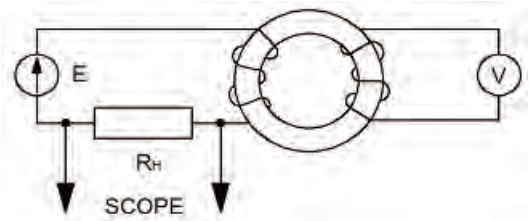


Fig.10. Examination of ferromagnetic materials by ammeter and voltmeter's method

Next example concerns investigation of ferromagnetic materials by ammeter and voltmeter's method (Fig.10). Connection over the NI ELVIS II platform is shown in Fig.11. Examined material is electrical steel with magnetic line length l=0,1 m. and section $S=5.10^{-6}$ m^2. Magnetization coil L_1 has w_1=1400 windings and measurement coil L_2 has w_2=450 windings.

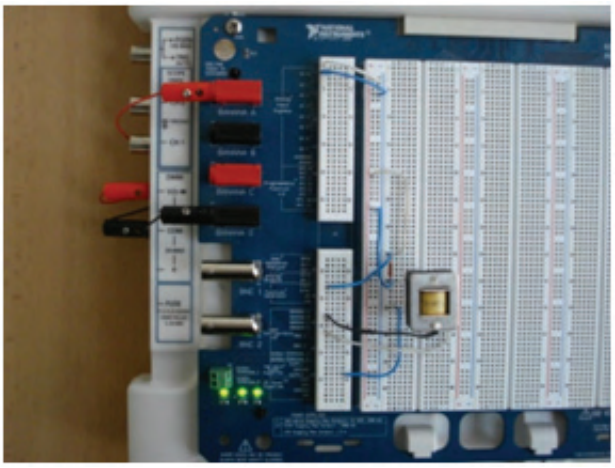


Fig. 11. Real connection of the pattern

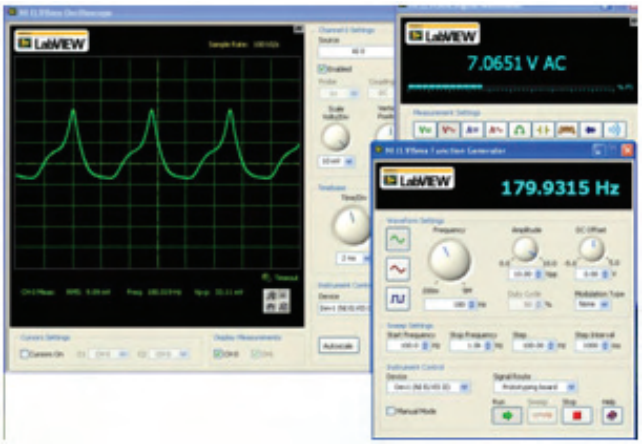


Fig. 12. Screenshot of the process of magnetization

For magnetization source is used integrated functional generator (FGEN), which can be activated from soft panel. The frequency in our case is 180 Hz. In magnetization circuit is connected resistance R_H=12Ω, for measurement drop of the voltage over it and calculation of magnetization current Im = Um/ R_H, with the help of the SCOPE from the soft panel. Non-linear character of the input current, enforces usage of the SCOPE. In the circuit of the measurement coil w_2 is connected voltmeter (DMM), which shows induced electromotive force. By this way is realized magnetization of the examined pattern with FGEN and primary coil L_1 and measurement of the induced electromotive force in secondary coil L_2 by DMM. Measurement current of the magnetization become possible with the SCOPE and its regulation is from virtual panel of the FGEN. A screenshot of this process is represented in Fig.12. By using measured electric values (Um and U_2), student can easily obtain amplitudes of the magnetic values (Hm and Bm), which correspond to the electrical, with the relations:

$$B_m = 4{,}44.f.w_2.S \quad (4)$$

$$H_m = \frac{U_2 w_1}{R_H .l} \quad (5)$$

From upper measurements and calculations is obtained one point of the magnetization curve B=f(H). By changing the level of magnetization of used pattern and reaching the saturation point, student can obtain necessary number of points of this dependence and trace entire curve out. With further calculation, can be received other main parameters like magnetic looses in the pattern and permeability.

The last example in this paper is linear Wheatstone bridge (Fig.13) and it illustrate accurately measurement of resistance in the environment of the NI ELVIS II workstation. On the Fig.14 is represented working scheme and used elements. Power of the bridge is applied from VPS, and for zero indicator is included DMM. Adjustment of the bridge can be accomplished with helical potentiometer (R_1, R_2).

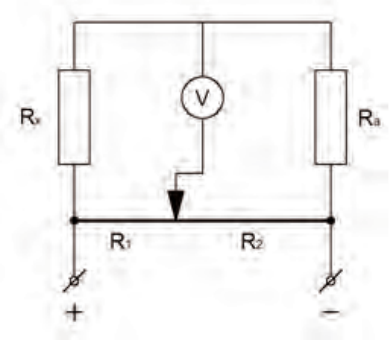


Fig.13. Wiring diagram of the linear Wheatstone bridge.

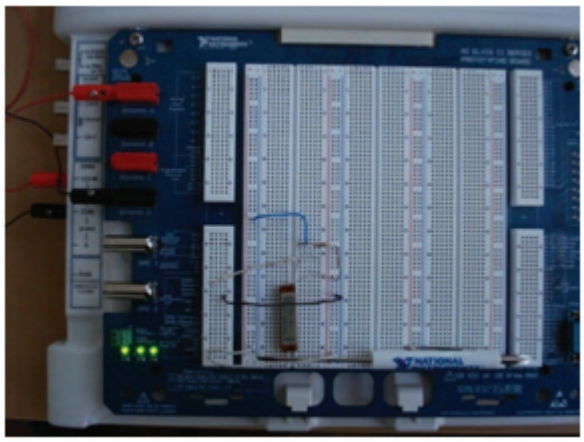


Fig.14. Real connection of the linear Wheatstone bridge.

When bridge is balanced, should be measured the resistance between endmost points of the potentiometer to the middle point, which are two of bridge's branches (R_1, R_2). These resistances are obtained with the virtual instrument IMPED from soft panel. The unknown resistance is calculated with:

$$R_X = \frac{R_1 R_A}{R_2} \quad (6)$$

III. Conclusion

The shown examples illustrate usage of the platform NI ELVIS II by construction and execution of different by kinds and complexity laboratory exercises, made in discipline "Electrical measurements". Advantages of the used virtual instruments obviously are: easy and quickly realization of the examined schemes and components, compactness by using different measurement units (almost all of the required instruments in electrical measurements laboratory are available), possibility of easy combination, reconstruction and enlargement of the investigated scheme, clarity, simplicity, and presence of built in defenses. These virtual instruments are more interesting for students and more understandable. This virtualization in laboratories reveals big opportunities for improving process of education and this new approach also lets materially to expand variation of the exercises, rising individual work with the students.

Acknowledgement

The paper was developed in the frames of the Research Program, financed from Ministry of Education.

References

[1] B. Paton, "Introduction to NI ELVIS., Part Number 323777C-01, NI Corporation, October 2005.
[2] R. Vasilev, I. Nedelchev, Electrical Measurements – Laboratory Exercises, Technical University of Varna, 2008.
[3] www.ni.com/elvis/

Framework for Mapping Learner and Gamer

Milen Petrov [1] and Adelina Aleksieva-Petrova[2]

***Abstract* – How to make education more effective while increasing the learner's are very important issues. Using educational games or so-called game elements in learning is one possible approach to achieving that goal. The main idea of the paper is to put on a convergence between learning and gaming styles including various types of user characters. In order to achieve that goal, we propose the framework which the learner and gaming style was included into a new learmer (learner and gamer) model. That helps in answering question how to choose appropriate games and how to apply them to achieve the learning objectives best suited to the learner as a gamer.**

***Keywords* – learner style, gamer style, learner, learning games, e-learning**

I. INTRODUCTION

How to make education more effective while increasing the learner's are very important issues. Using educational games or so-called game elements in learning is one possible approach to achieving that goal. An educational game can be defined as an instructional method that can provide students with real-time response, thus enabling them to gauge how well they retain the main concepts presented in the instruction session [1]. Then it is necessary to create the framework for evaluation of the game and game elements and how to use them in e-learning.

Business games are already well proven mean for effective learning the entire life cycle of products and processes for all stakeholders involved into business. Games similar to the PRIME game typically should model complex sets of real-life industrial processes and enable the creation and evaluation of specific Virtual Business Environments [2]. Thus, such games can help for example gaining practical experience in strategic manufacturing through intensive and fascinating gaming and e-learning.

Caillois defines four different types of games: those that involve competition; those that involve chance; those that involve simulation; and those that involve what he terms vertigo, such as fairground rides [3].

This paper discusses the main problem of how to choose appropriate games and how to apply them to achieve the learning objectives best suited to the learner as a gamer.

In the next section of this paper a methodology of research is presented which we will follow in modeling games and gaming styles. The third section gives definition of gaming styles which will be used in further research. The next forth section of the paper we propose the framework which map the gamer and learning style in a new learmer model. The last section makes out conclusions and outlines future work, as this paper is only a starting point for further research into building a software framework for successfully incorporating games and game elements into e-learning programs that can adapt to the learner according to his or her needs.

[1] Milen Petrov is from Department of Software Technologies, Faculty of Mathematics and Informatics, Sofia University, Bulgaria, James Baucher 5 Blvd, E-mail: milenp@fmi.uni-sofia.bg

[2]Adelina Aleksieva-Petrova is from Faculty of Computer Systems and Control, Technical University - Sofia, Bulgaria, E-mail: aaleksieva@tu-sofia.bg

II. METHODOLOGY

The methodology followed in this paper is depicted in Fig. 1 below.

This paper's research falls into follow blocks above: "3. Research model", "4. User model" and partially into block "5a. Select game(s)".

In order to build content for the user as both a learner and gamer in the data represented by extensible modelling language (XML) or other technical format of choice which will be made in later stage, we first need to build a user model in a modelling language.

All research stages, most of which are out-of-the-scope of current paper, nevertheless important to oversee the whole picture are: first - (1) *review of games and game elements*, next (2) *review of learning styles*, then combining (1) and (2) we received (3) *research model* which leads to (4) *user model*. As (4) outcome from one hand is (5a) *select game(s)* and (5b) *evaluate* those *selected games*, from the other hand (4) user model and (5a) selected games(s) makes interface to external system (in our case adaptive e-learning system) - shown on top of the figure. And last, but not least building block is (5c) which is *refining of game(s) or models* and returning feedback on (3) research model, in order to improve it.

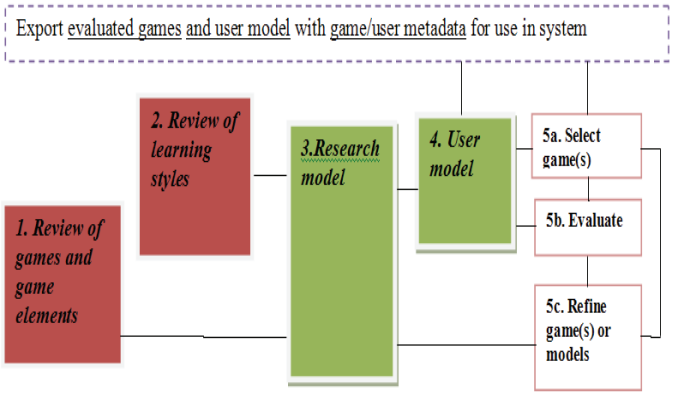


Fig. 1. Dependencies between learning style, resources and methods with gaming styles and game goals

In the next third section we start with a conceptual model of games and game elements and gaming style related to e-learning, which is referenced above as research model.

III. GAMING STYLE

According the above review we can summarize the following gaming styles:

A. Logician

Logician gaming styles are those in which gamer likes logic challenges and used spatial awareness, verbal skills, numeracy skills and spelling.

B. Competitor/Shooter

Competitor and shooter gaming styles we combine in order to characterize gamers that –enjoys action and shooting and focussed on the competition itself. That gaming styles uses not only shooting but also various instruments in a sports game.

C. Strategist

Strategist – is attracted by resolving complex problems within a game.

D. Dreamer

Dreamer gaming style is dedicated to a person who likes playing roles and thrives in the fantasy world of avatars.

E. Other - gaming styles and learning goals.

As many gaming can be defined according to different criteria, keeping styles is limited to mentioned four styles in order to keep research feasible.

Gaming style, learning activity and possible games are given below according to style and activities as outlined in Table 1.

TABLE I
GAMER STYLE, LEARNING ACTIVITY AND POSSIBLE GAME TYPES

Gamer style	Learning activity	Possible games
Logician	logic, spatial awareness, verbal skills, numeracy skills, and spelling	puzzle quiz
Dreamer	problem-solving and lateral thinking, collaborative skills, social interaction, negotiation, management of complex systems,	role-playing games adventure games
	strategy and working through scenarios	
Competitor	hand–eye coordination, planning and strategizing, problem-solving, teamwork and the ability to think quickly	sports platform shooter
Strategist	Strategy games can be used to teach planning, decision-making, testing hypotheses, strategic thinking, and management skills and seeing the consequences of actions taken.	strategy games

For mapping learner style - there are used four learning styles, proposed by Honey and Mumford - activist, theorist, reflector and pragmatist [4].

Classification of gamers' style which we relate to learning activities in the table below leads to following important according to our research remarks:

1. Learning style, which characterizes the *activist learner* can be mapped and used with *competitor or shooter* style of gamer. Typical for gamers with that style are open minded for new ideas and tend to experiment. They also like teamwork and prefer to be active.

2. On the opposite of activist is *theorist learning* style. That learning style can be mapped in higher degree to *logician style* of gamer. In that case is valid logical thinking and step-by-step solution of the problems.

3. Third learning style is known as *reflector learning* style. That style can be mapped to gamers' style named *dreamer* in the classification above. Usually that style of gamer prefers to stay aside and to analyse problem from different birth-eye views.

4. Last learning style we used is *pragmatist learning* style. That style can be mapped to strategy games and corresponding gamers' style of *strategist*. Pragmatists tend to describe and conceptualize the things which can be applied to their work.

Examples of research on different games and games elements are many in the literature - for example one research is done in [5], where are outlined three possible games - HangMan, Crosswords and Sudoku, which can be classified to first gaming style - logician. Deep investigation of classification of different appropriate games to learning and gaming styles as map proposed above is out of the scope of the paper.

Beside that there are needed strong indicators in order to classify different games as appropriate for one or another learning or gaming style. To facilitate that need in the next section of the paper we proposed framework for games and game elements components, as needed in e-learning systems and particularly in adaptive e-learning systems - which are target of our research.

IV. Conceptual Framework for Games and Game Elements Components in E-Learning

Fig. 2 outlines the dependencies and relationships between different key concepts (represented in the figure as blocks). The figure is divided into three horizontal layers and one upper right building block (game indicators).

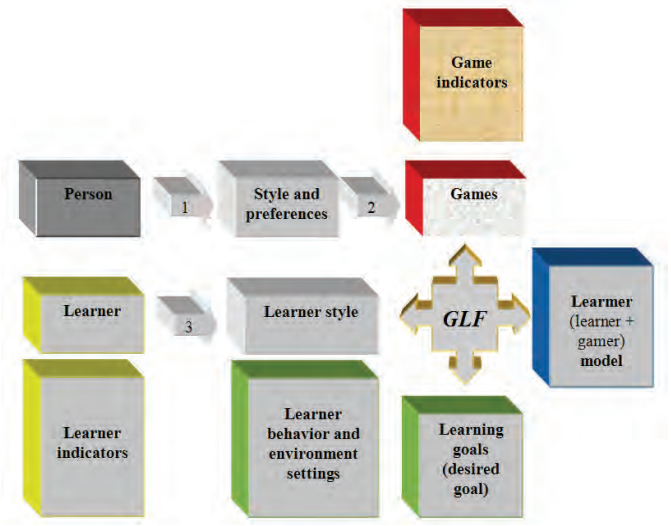


Fig. 2. System used to evaluate the relationship between learner and gamer to learning goals

In the *first horizontal layer* a person (or gamer) as an actor is represented, with its own personal style and preferences (represented as second building block), and the last block is selecting games without any care or knowledge of what exactly this game will teach him or her. This kind of learning without intention is known as informal learning [6].

Each game has its own characteristics, referred to in the figure as indicators. Some of these indicators are related to what the person learns, while others are not directly related to learning. We call useful indicators from the learning perspective *game learning indicators* (GLI).

The union between "pure" entertainment indicators and educational (game) learning indicators makes up general game indicators. In our framework we must track both types of indicators, as the entertainment indicators are very important and will retain attention of the learner.

The *second horizontal layer* represents the person as a learner. Any learner has his or her own learning style or blend of styles. Learning styles are outside the goals of this paper.

Learner model stores information about the final user, receiver of the e-learning content, such as personal data, preferences, goals, level of knowledge, performance shown during assessment, etc.

Nonetheless, learning styles are certainly an important aspect of this research. One of the most popular learning style classifications includes four styles: activist, reflector, theorist and pragmatist (as mentioned above). Other learning styles are reviewed in more detail in [7].

The *third horizontal level* represents different characteristics related to the learner: learning indicators represent learning style, while learning behaviour and environmental settings represent the influence of the existing environment on one's learning style.

Last, there are three very important concepts in this current research: mapping function and the relation between learning goals (selected before starting the learning process), learning style (which already incorporates environmental settings and behaviour) and goals achieved (on the figure this is the block on the far right) and games selected (achieving both formal and informal learning), which incorporates game learning indicators. The mapping function which is defined above is called the *game learning function* (GLF). Investigation of the GLF is beyond scope of this paper.

V. Conclusions

The primary goal of this paper is to classify games and related game learning indicators as a first step in modelling and developing a software framework that evaluates games according to their ability to support and enable learning.

The second goal of this paper is to classify games according to the classification system described in relevant literature. That will be of use in further research into building and investigating the GLF and to help classify gamer characteristics according to game indicators and personal style and preferences.

The learner and gaming style was be included into a new learmer (learner and gamer) model, which will be used for managing adaptation within the adaptation model. Thus, the development of modern learning systems that are adaptable to one's own learning characteristics is often oriented toward the integration of game elements and their application in ways that take into account the learner's unique learning style [8].

Acknowledgement

Research is supported by project ADaptive techn0logy-enhanced Platform for eduTAinment (ADOPTA), started 01.2009, no. D002/155 funded by funded by the Bulgarian Ministry of Education and Science.

References

[1] Leach G., T. Sugarman, "Play to win! Using games in library instruction to enhance student learning", Research Strategies, 2005, Volume 20, Issue 3, Pages 191-203.

[2] Boytchev P., Bontchev B., Nikolov R., Oliveira M., and Koychev I., 2006. Designing a Toolset for the PRIME Virtual Business Environment. 10th Workshop on Experimental Interactive Learning in Industrial Management; IFIP WG5.7 SIG, Trondheim, Norway.

[3] Caillois, Roger., 1961. Man, Play and Games, trans. Meyer Barash. New York: Free Press.

[4] Honey P, A.Mumford, 2000. "The Learning Styles Helper's Guide", Peter Honey Publications, September, ISBN: 978-1-902899-28-2

[5] Petrov, M. "Assessment with Game Elements", Fifth International Conference - Computer Science'2009 , 05 - 06 November 2009, Sofia, Bulgaria, (in-print).
[6] Eraut M, 2004. "Informal learning in the workplace", Studies in Continuing Education, Vol. 26, No. 2
[7] Coffield F., Moseley D., Hall E., Ecclestone K., 2004. "Should we be using Learning Styles? What research has to say to practice", Learning and Skills Development Agency, London.
[8] Vassileva D., Bontchev, B., 2009. "Management of storyboard graphs for adaptive courseware construction", Proc. of 3rd Int. Conf. on Communication and Inf. Technology (CIT'09), Athens, Greece, ISBN: 978-960-474-146-5, pp. 39-44.

Using Student Feedback in Measuring the Quality of Teaching in Higher Education

Danijel Mijic[1]

***Abstract* – European standards for quality assurance in higher education define important role of students in quality assurance process. As students are directly involved in teaching process, their perception of teaching process quality could be used as one of the valuable indicators for measuring the quality of teaching. Students' opinions about various aspects of teaching process are collected using evaluation instruments. One way of getting feedback about teaching process quality is evaluation using students' questionnaires in paper or online form. This paper describes some practical experiences and results in students' evaluation of teaching quality at University of East Sarajevo, Faculty of Electrical Engineering.**

***Keywords* – quality, assurance, higher education, evaluation, teaching.**

I. INTRODUCTION

Quality assurance plays important role in higher education. In order to retain and improve their position at the educational services market, institutions of higher education need to pay close attention to the quality of their services. One of the processes that is crucial for improving overall quality is a teaching process. Measuring teaching process quality involves getting information about various aspects of teaching process and usually includes measuring quality of teachers, quality of individual course units, and quality of complete academic programmes.

According to the published literature on this subject, there are many instruments used for measuring quality indicators in higher education, [1], [2], [3]. Some of them are focused on measuring quality of individual course units and teachers, while others tend to provide information about quality of complete academic programmes or institutions. One common thing for both of the mentioned types of instruments is that they are mainly used for getting feedback from students in order to get information about students' perception of various aspects of teaching process. As students are directly involved in teaching process, their evaluation of teaching quality plays very important role in measuring overall teaching quality.

This paper describes an instrument for measuring teaching quality that is developed and used at University of East Sarajevo, Faculty of Electrical Engineering, from the academic year 2006/07 to the present day. The paper presents some practical experiences and results using the instrument at Faculty of Electrical Engineering.

[1]Danijel Mijic is with the University of East Sarajevo, Faculty of Electrical Engineering, Vuka Karadzica 30, 71123 East Sarajevo, Bosnia and Herzegovina, E-mail: danijel.mijic@etf.unssa.rs.ba

II. INSTRUMENTS FOR GETTING STUDENT FEEDBACK

Students' role in measuring teaching quality is a well known and established in some regions in the world. This is especially true for North America and Australia, where standardized and proven instruments are used for getting students' feedback, like SEEQ (Student Evaluation of Educational Quality) and CEQ (Course Experience Questionnaire).

One of the best developed and most widely used student feedback questionnaires in the USA is the SEEQ [4]. Unlike instruments such as CEQ the SEEQ is not based on student learning research but on psychometric analysis. A consequence of this is that while the constructs underlying the SEEQ are less well supported by learning theory, the psychometric characteristics of the questionnaire are developed to a high degree. The intellectual rights and copyright in the SEEQ belong to Professor Herbert W. Marsh of the University of Western Sydney, Macarthur. There is a series of publications describing research, methodologies and the SEEQ instrument, [5], [6], [7], [8]. The intellectual rights and the copyright in the CEQ belong to Professor Paul Ramsden, the Graduate Careers Council of Australia and the Australian Commonwealth Department of Education, Training and Youth Affairs. The CEQ instrument is widely accepted at Australian universities. It is also adopted for the other purposes different from the original one, with some modifications and additions according to requirements and the application, as described in [9], [10].

In the European higher education area students' role in quality assurance is defined in Standards and guidelines for QA, published by European Association for Quality Assurance in Higher Education (ENQA). Many higher education institutions have developed their own instruments for getting feedback from students, but there is not much published research evidence on proven reliability and validity of results of using these instruments like it is the case with the SEEQ and CEQ.

Students' role in measuring teaching quality is important for the several reasons [6]:

- Diagnostic feedback to teachers about the effectiveness of their teaching,
- A measure of teaching effectiveness to be used in administrative decision making,
- Information for students to use in the selection of course units and teachers,

- An outcome or process description for use in research on teaching.

Traditional way of getting students' feedback is by using paper forms with a list of statements to which students indicate their level of agreement. The statements are chosen to appropriately reflect various aspects of teaching. The level of agreement is mostly indicated using five point scale with descriptive optional answers, like "Strongly agree" to "Strongly disagree", or similar scales that enable paper forms to be scanned and processed for getting results. The paper forms are distributed to students participating the evaluation usually in two ways. One is the class environment where students are asked to fill out the questionnaires and return them immediately. The other way is sending questionnaires to students by classical postal system and expecting them to return them in the same way. However, when larger groups of students are involved in the evaluation, this means of getting students' feedback becomes time-consuming and inefficient when it comes to collecting and processing results.

An alternative solution to traditional one mentioned earlier became available with the increased use of information technologies in higher education. The paper form questionnaires are transformed into the online forms with the similar content. In this way results are automatically stored into electronic form suitable for further processing and analysis. The online questionnaires have many advantages over the traditional ones. The main advantage is improved efficiency. A few examples of instruments for online evaluation are described in [9], [10], [11].

III. Method and the Evaluation Instrument

Students' evaluation of teachers and course units at University of East Sarajevo, Faculty of Electrical Engineering, was formally introduced in 2002. At first, the traditional way of getting students' feedback was used. At the end of each semester, during the evaluation period of two weeks, students filled out a printed questionnaire which contained statements and questions for evaluation of teaching quality of individual teachers and quality of individual course units in the current semester. Students are required to respond to statements and questions using predefined answers in the form of rates from 5 to 10, or using Yes/No answers. It was also possible to write a free comment for every teacher or course unit in the questionnaire. The scale 5 to 10 is selected because the same scale is used in local higher educational system for rating students at exams and assessments.

This form of students' evaluation was used for several years and achieved positive results with respect to getting feedback from students about their perception of teaching process quality. However, this means of getting students' feedback shown itself as inefficient and very time-consuming when it comes to processing and analyzing results. Since all of the active students are required to fill out the questionnaire at the end of each semester, there was significant number of paper forms to be manually sorted and processed in order to get the results. The sample size was usually about 200 to 250 active students involved in the four year undergraduate programme. Every student had to evaluate teaching performance of each teacher and for each course unit in the semester. Even with the relatively small sample, the work that had to be done for processing the results was significant and could not be done in the short time. This was the main reason for initiating development of application for online evaluation of teaching quality.

The application for online evaluation of teaching quality was developed in the form of multi-user web application. It enabled students to fill out online questionnaires having similar content to the traditional ones, but it also enabled students, teachers and administration to see the results of students' feedback. The application could be accessed from any place equipped with Internet connection by using standard web browser.

The online questionnaires used four scales for evaluation of course units and teachers, overall students' course satisfaction and evaluation of institution's resources used as a support for teaching. Scale for evaluation of course units and teachers consists of three questions with Yes/No answers and a free comment field, while scale for evaluation of teachers consists of eight statements which are rated on the scale from 5 to 10 and one free comment field. Scale for rating overall students' course satisfaction and scales for rating each of the institutional resources used for supporting teaching consist of one statement with rates from 5 to 10 and a free comment field.

The application was developed using open source technologies on the LAMP platform (Linux, Apache, MySQL, PHP). One of the main reasons for choosing these technologies was their free availability, no additional financial expenses for software licenses. The other important reason was the existing information system used at University of East Sarajevo and all of its organizational units. The DBMS used in the information system was also MySQL. The application uses much of the existing data from the existing information system, especially data about students, teachers and course units.

IV. Results

Since the beginning of use in second semester of academic year 2006/07, the application collected feedback from 1015 students about 86 individual teachers and 172 individual course units. The overall sample size was 1226 students and overall response rate is 82.79%. Response rates in specific academic years and semesters are given in Table I.

TABLE I
RESPONSE RATES THROUGH ACADEMIC YEARS

Academic Year	Semester	Sample	Response	Response rate
2006/07	2	208	186	89.42%
2007/08	1	236	215	91.10%
2007/08	2	170	115	67.65%
2008/09	1	186	155	83.33%
2008/09	2	186	144	77.42%
2009/10	1	240	200	83.33%

A. *Students' Reflections on Using the Instrument*

Using the instrument was considered convenient by most of the students. They were not limited to time and place for filling out the questionnaire, instead they could do it any time during the evaluation period from the privacy of their homes or other places having Internet connectivity. They were not under pressure to fill out questionnaires in the class environment, so they felt more comfortably. Response rates shown in Table 1 are well above response rate of 50% that is mostly considered to be a minimum value for accepting results as representative from the given sample [1].

One negative impact on using the instrument was a small level of skepticism about staying anonymous in the process of electronic evaluation. Since the students used their students' ID and a password to access online evaluation system, some of them suspected that this ID is used to track their responses and they were afraid of having negative consequences on their future studies in the case they respond in negative context to some statements or questions for evaluating teachers. Of course, this was not the case, the instrument used students' IDs for getting information about course units and teachers associated with the course units for the particular student. In order to eliminate negative impact of this issue on response rates and response grades, the process of authentication is changed in the new version of application.

B. *Teachers' Reflections on Using the Instrument*

Teachers' reflections on the instrument were generally positive. In some minor number of cases they greatly depend on the results obtained from the students' feedback. Some teachers complained that students are not competent to evaluate their work, or some said that teachers' rates depend on the course unit complexity.

One of the most important advantages of using this instrument, as with the other similar instruments for getting students' feedback about teaching quality, is the process of self-regulation. Students' perception of teaching quality helps teachers to improve quality of their teaching. Since the teachers are able to see results of students' evaluation for the other teachers, they are additionally motivated to give their best in teaching process and compete with other teachers in order to get higher ratings.

C. *Improving Quality of Teaching*

Students' evaluation of teaching quality is not by itself enough to improve teaching quality. Having a collection of students' evaluations of teaching quality doesn't guarantee the quality improvement. This could be true for the several reasons, but one of the most important is lack of institutional policies that regulate interpretation, analysis of results and necessary actions based upon results of evaluation in order to improve quality. As with students' evaluations of teaching using SEEQ instrument, there is little evidence that the collection of student feedback using the CEQ instrument in itself leads to any improvement in the perceived quality of programmes of study [1].

In the case of University of East Sarajevo, in the present situation there is no procedures defined for acting upon results of students' evaluation of teaching quality. The only influence on quality is done by the process of self-regulation. Teachers with lower ratings can identify segments of their work that are to be improved in order to improve teaching quality. The same applies to individual course units and the other elements of teaching process that are subject to evaluation. As quality assurance in higher education is becoming more and more important in Bosnia and Herzegovina and the region, it is likely that the institutional policies will very soon incorporate appropriate mechanisms for taking into account results obtained by students' evaluations of teaching quality.

Fig. 1 shows trend of overall course satisfaction of students over the several years and semesters. The ratings are displayed on a scale from 5 to 10.

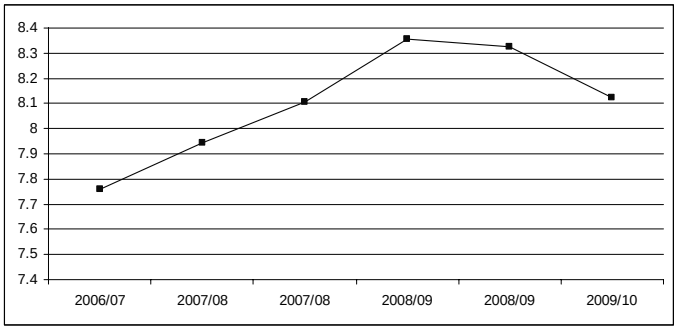


Fig. 1. Trend of students' overall course satisfaction

Based on results shown in Fig. 1 it could be concluded that the students' overall course satisfaction has positive trend from second semester of academic year 2006/07 to second semester of academic year 2008/09. Next two evaluations resulted in small negative trends. The results shown could be interpreted as a quality improvement, but there is no evidence that this is due to some actions taken in order to improve quality. Results of all students' evaluations are stored in a database and can be accessed in order to analyze trends of specific quality indicators. Since the application is in function from the second semester of academic year 2006/07 there is not enough data to establish long term trends of quality indicators. This is especially true for evaluation of teachers.

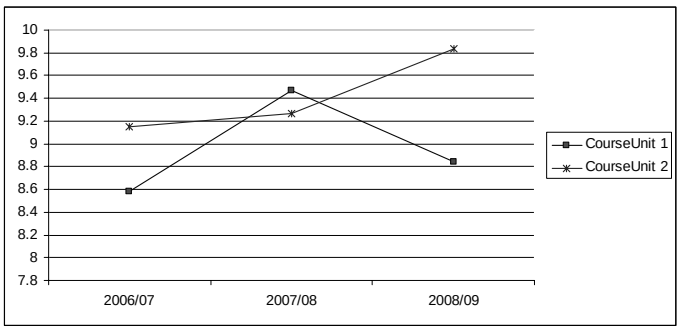


Fig. 2. Trend of teacher's ratings for two course units in the second semester

Fig. 2 shows trends of overall teaching performance of randomly selected teacher on two course units for the second semester in academic years 2006/07, 2007/08 and 2008/09.

Fig. 3 shows trends of overall teaching performance of the same teacher on two course units for the first semester in academic years 2007/08, 2008/09 and 2009/10.

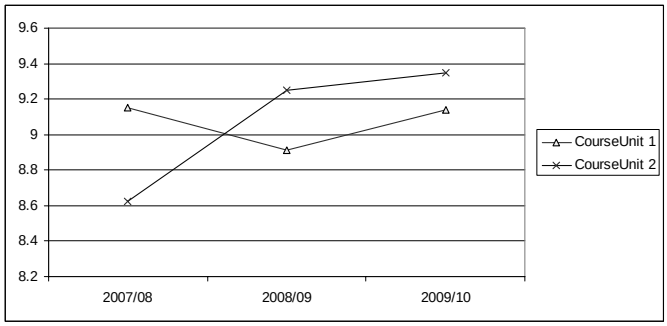


Fig. 3. Trend of teacher's ratings for two course units in the first semester

V. CONCLUSION

The instrument used for measuring the quality of teaching at University of East Sarajevo, Faculty of Electrical Engineering has many benefits when compared to the previous methods of obtaining students' feedback. The main benefit is its efficiency in getting students' responses and processing results. In the previous solution using traditional means of getting feedback from students it took weeks to collect, process and publish results of the students' evaluation of teaching quality. Using the new instrument for online evaluation this time is greatly reduced and completely eliminated. In the end of evaluation period results are already available and ready for publishing and analysis. Partial results could be available even during the evaluation period.

Another important benefit is the availability of results of evaluations in electronic form which is convenient for further processing and analysis. Based on the results from the previous evaluations one could conclude about trends of specific quality indicators. This way it could be detected if the evaluation is having impact on the quality improvement.

The results could also be aggregated and transformed in form convenient for advanced reporting and analysis which is required for administration staff as a support for decision making. Using OLAP tools data could be analyzed through different dimensions by persons with limited IT expertise.

New model of the application that is used as evaluation instrument is currently in the development phase. It should improve some of the identified limitations of the existing application in order to use it at the University level. By using the application the University management will have more detailed insight in teaching quality indicators when compared to current solution which is very limited. This model should offer enough flexibility in order to apply it on the other similar institutions having different organizational scheme. It will offer integration with other existing proprietary systems for the purpose of using the existing data where the required data is already available.

For the better success of the students' evaluation of teaching quality it is necessary to motivate relevant population of students and to achieve high response rates. Students should be aware that their response will be considered and taken into account and they can improve their courses by being more involved in the process of quality assurance.

Teachers also have to be motivated to react in response to students' evaluation. Self-regulation is not always sufficient and higher education institutions should take more care about defining relevant policies and procedures in order to improve overall teaching quality.

REFERENCES

[1] Richardson, J. T. E., "Instruments for obtaining student feedback: a review of the literature", Assessment & Evaluation in Higher Education Vol. 30, No. 4, pp.387-415, 2005.

[2] McInnis, C., Griffin, P., James, R., Coates, H., "Development of the Course Experience Questionnaire (CEQ)", Centre for the Study of Higher Education and Assessment Research Centre, Faculty of Education, The University of Melbourne, 2001.

[3] Hung, W., Smith, T. J., Harris, M. S., Lockard, J., "Development research of a teachers' educational performance support system: the practices of design, development, and evaluation", Association for Educational Communications and Technology, 2007.

[4] Coffey, M., Gibbs, G., "The Evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK Higher Education", Assessment & Evaluation in Higher Education, Vol. 26, No. 1, 2001.

[5] Marsh, H, W., "A Longitudinal Perspective of Students' Evaluations of University Teaching: Ratings of the Same Teachers over a 13-Year Period", Paper presented at the Annual meeting of the American Educational Research Association, San Francisco, CA, 1992.

[6] Marsh, H, W., "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research", International Journal of Educational Research, 11, pp.253-387, 1987.

[7] Marsh, H, W., "Students' Evaluations of College/University Teaching: A Description of Research and an Instrument", Annual Meeting of the Australian Association for Research in Educational, Sydney, Australia, 1980.

[8] Beatty, B., Marsh, H. W., "Students' Evaluations of Instructional Effectiveness: Research and a Survey Instrument", University of California, 1974.

[9] Tucker, B., Jones, S., Straker, L., Cole, J., "Course Evaluation on the Web: Facilitating Student and Teacher Reflection to Improve Learning", New Directions for Teaching & Learning, Issue 96, pp.81-93, 2003.

[10] Tucker, B., Jones, S., Straker, L., "Online student evaluation improves Course Experience Questionnaire results in a physiotherapy program", Higher Education Research & Development, Vol. 27, Issue 3, pp.281-296, 2008.

[11] Harrington, C. F., Reasons, S. G.:, "Online Student Evaluation of Teaching for Distance Education: A Perfect Match?", The Journal of Educators Online, Volume 2, Number 1, 2005.

Program Library for Development and Research of Distance Learning Courses

Mitko M.Mitev[1] and Dimitar Georgiev[2]

Abstract: **A library of tools for development and research of distance learning courses is represented. The library includes generators of terms and relationships between them, program modules for the purpose of structuring of the educational materials, grouping of terms, establishing of input and output relationships between the disciplines and others. The tools possess graphic interface and can be used independently or in graphic environment for design and research**.

Keywords: **program library, graph properties in terms of theory of sets, optimization algorithms, grouping, structuring, modelling.**

I. INTRODUCTION

One of the most progressively developing types of educational processes is the distance learning. The advantages and the disadvantages of distance learning are well known [6,7]. The major problems that have to be solved are connected with the following directions:

- Development of program tools for organization and maintenance of the educational process, based on the principles of distance learning.
- Development of theoretical concepts and methodical support for design, research and organization of group and individual education [1].
- Development and research of practical courses for different fields of education.

The present paper concerns the development of library of tools for development and research of courses for distance learning.

II. FORMAL CONCEPT OF INSTRUMENTAL TOOLS

The instrumental tools are based on the following concept[2-5]:

- The educational material which is to be included in the process of distance learning is presented as a logical sequence of interconnected terms subsequently situated on the axis of time.
- Every term is categorized according to its difficulty of learning measured in time
- A term which is necessary for the clarification of another term is regarded as a basic term, and the other one is regarded as its consequence.
- The basic terms have to be put on the axis of time before their consequences.
- The distance between the terms is measured by the total time necessary for the clarification of the terms situated between them on the axis of time.

The formalization of the concept is put down to its presentation as a finite directed graph *G(X,U,P,W,F)* [2] where:

$X = \{x_i\}$, $|X| = N$ – the set of terms described by the discipline.

$U = \{u_{ii}\}$ – the set of connections between the terms, where x_i is the basic term of the term x_j.

$W = \{w_{ij}\}$ – the set of weighting factors of the terms $X \leftrightarrow W$

The ordered set *G(X,U,P,W,F)* can be interpreted as a finite directed graph, where $P(x_{ij}, u_{ij}, x_j)$ is an incidentor which is a predicate with value true, if x_i is a basic term of x_j.

On that basis there are many algorithms used for structuring the educational materials as logical subsequences of interconnected terms, there are algorithms for restructuring, for reduction of the consequences of wrong arrangements, for bringing the interconnected terms closer in terms of time, for grouping and arranging the terms connected with the previous/next discipline of the educational plan, for grouping the terms for individual learning, for carrying out tests, for adapting the teaching rate in accordance with the level of knowledge, for statistical research etc.

III. PROGRAM LIBRARY OF INSTRUMENTAL TOOLS FOR DEVELOPING AND RESEARCH OF COURSES FOR DISTANCE LEARNING

The program implementation of the algorithms is presented as a program library of instrumental tools for developing and research of courses for distance learning. The library is structured to be used in two program environments (Fig.1.):

- *Library of major classes without graphic interface*. These tools can be applied to already developed projects with graphic interface or to the graphic environment of the visual studio.

[1]Mitko M. Mitev is with the TU- Varna, Department of Computer Sciences and Engineering, Studentska 1, 9010 Varna, Bulgaria, E-mail: mitevmm@abv.bg

[2]Dimitar Georgiev is with the TU- Varna, Department of Computer Sciences and Engineering, Studentska 1, 9010 Varna, Bulgaria, E-mail: teamarni@yahoo.com

- *Library of upgrades of the major classes.* The graphic interface uses the already defined classes in the library which are using now custom controls for visualization.

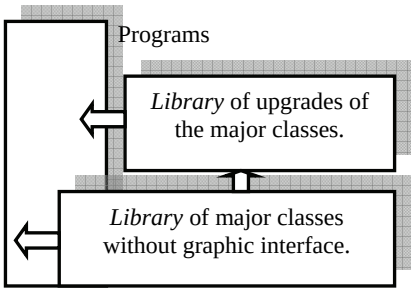


Fig.1.The main structure of the library

The classes in the library are grouped in the following way (Fig.2.):

1. *Classes used for generation of graphs.* Random number generators are used for this purpose. The generated values are interpreted as numbers of the vertices (terms), weight factors of each vertex (time for understanding/learning), weight factors of the edges and others.
2. *Classes used for structuring of the educational material.* The classes are based on various algorithms for arranging the vertices on the axis of time. For this purpose the dynamical weight factors of the vertices are defined which are the sum of the contiguous vertices multiplied by the static weight factor of the edge. The edges having a negative direction increase their dynamical weight factor by a negative coefficient. The classes can be executed with or without the negative coefficient used for the edges with negative directions.
3. *Classes used to define the starting and the final package.* These classes analyze the interdisciplinary connections between the terms. The purpose of the analysis is to separate the terms in groups (starting and final). Applying classes of the previous group will optimize their arrangement on the axis of time. These additional options can rearrange the packages depending on the interconnections between the terms. As a result the starting and the final terms are situated optimally close both to the outside and to the inside terms.
4. *Classes used for grouping of the terms.* The purpose of these classes is the formation of educational units for distant learning organization. The initial grouping is connected with previously defined limiting conditions and is used during process of course development. In the process of study the grouping is done under the conditions of individual learning. This helps for the adaptation of the learning process to the individual specifics of each student. The grouping criteria represent a strong inside connectivity (and a weak outside connectivity) of the terms. The time direction has to be taken into account (edges in straight and opposite direction).
5. *Classes for analysis of testing results.* The process of examination of the acquired knowledge and practical skills is presented as definition of subgraphs. The terms which haven't been learned present a subgraph connected with the previous and the next terms. This raises the question of their rearrangement on the axis of time. In case of an insufficient or critical rate of learning starting and final packages can be formed and analyzed.
6. *Classes used for statistical work.* The test data are saved into a database. The statistical process is connected with determining of the mathematical dispersion, asymmetry. With famous criteria of Fisher and Student the statistical difference in the results before and after using one of the algorithms is proved.
7. *Interface classes.* These classes are mainly used for input of real data from the examined educational subject. Mainly these are classes for text information analyses, identification of the used terms and definition the synonyms and the homonyms. The classes have also user interface. Their future development is strongly connected with the adaptation to the library of instrumental tools for design and research of the courses for distant learning.
8. *Additional classes.* These classes are used for support of the different presentations of finite oriented graphs and for adaptation to different data bases and program environments.

IV. CONCLUSION

A fully functional system of instrumental tools for developing and research for courses of distance learning is developed.

The library is designed in two levels: one for use in a program environment and the other for use in a graphic environment.

In the first case the major components of the library can be implemented in program systems for distant learning in order to improve their functionality. Such applications require software developers to work together with the teachers. Such systems are characterized by high performance but limited exploitation.

The second type is based on graphic interface. This allows the system to be assembled mainly by teachers not specialized in computer systems. Because of the user interface the system

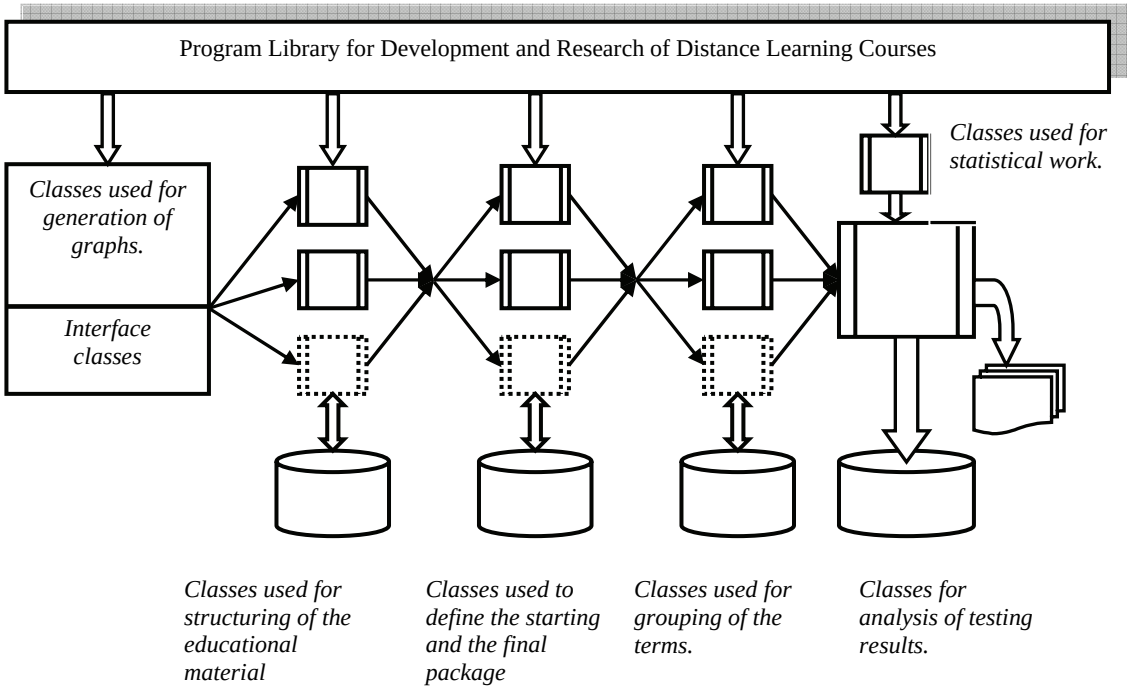


Fig.2.The extended structure of the library

provides high flexibility in structuring and easy achievement of various variations of solutions.

Some components of the library can be used separately for solving of partial problems.

REFERENCES

[1] Elena Racheva, Mitko Mitev, Eleonora Kalcheva - "Genetic algorithm for structuring the educational content for e-learning purposes", "Computer sciences and technologies" magazine, 1/2008, Technical University – Varna ,pp. 15-20 (*in bulgarian language*)

[2] Mitko Mitev, Elena Racheva - "Developing an algorithm and a program for educational material structuring", "Computer sciences and technologies" magazine, 2/2008, Technical University – Varna , pp. 15-20 (*in bulgarian language*)

[3] Mitko Mitev, Elena Racheva, Ginka Marinova - "Algorithm for Defining of Initial Package of Notions for Purpose of E-Learning", International Conference on Computer Systems and Technologies - CompSysTech'2009, Ruse, Bulgaria, 2009,IV 9-1 IV 9-5

[4] Mitko M. Mitev – "Algorithm for defining final package of notions in e-learning", Collection of works from the Technical University - Varna, pp. 124-128 (*in bulgarian language*)

[5] Mitko Mitev, Elena Racheva, Ginka Marinova - "Graph-analytical approach for developing courses in e-learning education", Fourth International conference "Modern information technologies in education for all: innovative methods and models", Kiev, 2009,pp. 107-116 (*in russian language*)

[6] Kaciniene I.,V.J. V.,S. Daukilas, D. Vaisnoriene - "Factors that Impact Quality of e-Teaching/ Learning Technologies in Higher Education", issue: 05 / 2008, pp. 132-151, www.ceeol.com

[7] Volungeviciene A., M. Tereseviciene –"Quality Assessment Dimensions of Distance Teaching/Learning Curriculum Designing ", 05 / 2008, pp. 32-53, on www.ceeol.com

[8] Learning Technology Standards Committee, Draft Standard for Learning Object Metadata. IEEE Standard 1484.12.1, New York: Institute of Electrical and Electronics Engineers.

This page intentionally left blank.

The MPLS Network Simulators in the Computer Network Education

Veneta P. Aleksieva[1]

***Abstract* –In this paper are discussed the MPLS network simulators as teaching tool. It is presented some criteria of the simulators comparison. The main goal is to find the best tool for teaching innovation.**

***Keywords* – simulator, MPLS, network education, networking tool**

I. INTRODUCTION

Next generation communication networks are migrating towards a unified network architecture where both wired and wireless network segments will co-exist. This is accompanied by the growing demand on networks to provide QoS, due to the rise in popularity of real time and multimedia applications. Future networks should accommodate a variety of services, satisfying different traffic types, providing support for user mobility, and will be able to guarantee QoS.

MPLS/Multi protocol Label Switching/ is a connection oriented technology that arises to palliate the problems that current networks have related to speed, scalability and traffic engineering [1]. In fact, MPLS packets forwarding is based on labels and not in the analysis of encapsulated data from upper levels. It is a multi protocol technology that supports any network protocol as well as any technology in lower layers (link or physical). MPLS is used as a traffic engineering tool to direct traffic in a network in a more efficient way than original IP shortest path routing. Path in the network can be reserved for traffic that is sensitive, and links and router that are more secure and not known to fail can be used for this kind of traffic.

MPLS recovery provides different levels of service, based on their service requirements. It should give the flexibility to select the recovery mechanism, choose the granularity at which traffic is protected and choose the specific types of traffic that are protected in order to give operations more control over that tradeoff. [2]

Actually, there is a growing need for end-to-end QoS mechanisms with mobility support to address the requirements of heterogeneous network environments. The wireless segments in the next generation networks are likely to be based on multi-hop ad-hoc or infrastructure-less networks. Also, in wired networks, paths once established, hardly change. On the other hand in mobile networks, calculated routes have limited lifespan since the network connection structure changes due to nodal mobility. The topological changes affect the resource availability directly. Further, wireless resources and topological connectivity are even harder to track accurately because of the nature and transmission effects associated with the shared wireless media. Consequently, adhering to the negotiated QoS guarantees is even more challenging than in static network topologies. It will be good to analyze and test these cases with simulation, before applying in the real networks.

The recovery of the MPLS network is based on the algorithm that is applied in order to detect the faults and route the data flow in an alternative path. There are various algorithms. However, each algorithm employs only one of the two basic techniques:

- **protection switching**, where a recomputed alternative path, which is usually disjoint from the working path, is set up for every flow
- **rerouting**, where an alternative path is dynamically recomputed after a fault is detected.

For both techniques, the alternative path can be either global or local. The protection of data flows in case of link or router failures is very important, especially for real time services and multimedia applications.

In this research a comparative between some common teaching simulators is made. Some researchers have already made proposals in this way [2,3,4,5]. It is given an overview of the available network simulators with MPLS implemented, and it is described by which criteria must choose MPLS simulator as teaching tool. The primary purpose of this research is to find a simulator that enables to simulate various MPLS applications without constructing a real MPLS network.

II. CHOICE OF CRITERIA OF MPLS SIMULATORS COMPARISON

When surveying various simulation tools which could be used in evaluation studies of service availability and resiliency mechanisms in MPLS network, their following features should be taken into account:

- Modelling capabilities
- Credibility of simulation models
- Credibility of simulation results
- Extendibility
- Usability
- Cost of licenses

Each potential simulator will be used in a simple simulation of MPLS recovery model, to gain practical experience and to

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

assess the level of user-friendliness of simulators considered [6].

In this research the main goal is to be found an MPLS simulator that allows designing and setting MPLS domains and their components and give different MPLS recovery mechanism opportunities. When the simulation is run, it must perform statistical analysis of its results, from all educational points of view. Their main features must:

- be a teaching simulator that allows an elementary statistical analysis of the network traffic;
- have a visual editor to design scenes;
- be a multiplatform software;
- be free-of-charge;
- be easy to install;
- have not been designed to work in real MPLS networks using current manufacturers' components.

In the research[2] it is used some additional criteria of comparison of MPLS simulators:

- Interactive simulation
- Multilanguage
- Free software
- QoS simulation
- Real environments applicability
- Installation and execution easiness

Moreover, it can be used four criteria which are very important for the performance of the restoration mechanisms in MPLS networks:

- fault recovery time,
- packet loss,
- packet reordering,
- tolerance of multiple faults.

Simulation results must indicate the performance advantages of the proposed MPLS recovery algorithms, when compared with other restoration mechanisms, based on these criteria. In the research [7] are used some new criteria:

- Latency
- Full restoration time
- Vulnerability
- Quality of protection

In this research it is used combination of all these criteria, because in the computer network education of MPLS it is important to present advantages and disadvantages of different MPLS recovery schemes.

III. COMPARISON OF MPLS SIMULATORS

In this section it is given a short overview of the available network simulators with MPLS implemented. Each simulator is individually assessed and mutually compared.

- **J-Sim** (Java Sim) [8] is a component based network simulator developed entirely in Java, by Hung-ying Tyan and some other people at the Ohio State University. J-Sim is an open source simulator. There has been one MPLS module contributed to this simulator, developed by the Infonet Group of the University of Namur. This model consists of two components: a forwarding table component and a MPLS component. It associates an IP prefix or an incoming label with an outgoing interface and an outgoing label. The MPLS component forwards packets according to the configuration of the forwarding table. This model does not include any label distribution protocol so LSPs must be setup statically.
- **OMNeT++ [**9] is a discrete event simulation environment programmed in C++ and developed by András Varga. The simulator is open source and free to use for academic and non-profit users. The simulator is component-based and there are many models in development for this simulator. The MPLS model in OMNeT++ was originally developed by Xuan Thang Nguyen from the University of Technology in Sydney, but now the model is maintained by András Varga [10]. The model includes components for MPLS forwarding, LDP, CR-LDP and RSVP-TE. This model was developed to study the forwarding mechanisms in MPLS so there where some simplification done in the way that the Label Switched Routers was implemented. This made this model, as it is today, not suitable for failure recovery simulation. The forwarding tables for all routers in the network is implemented in one table, which means that all routers have the same copy of the network topology at the same time. If a failure happens in the network than all routers will be aware of the failure at the same time. This is not what happens in a real case scenario where topology changes like a failure has to be indicated to nodes by sending routing update information from the point of failure. There were also some problems with the implemented TCP and UDP for IP on time. RSVP-TE was implemented, but after some testing of the included MPLS model and the transport protocols, it is noticed that a lot of the implemented functionality in the model had to be rewritten. It has a professional technical support, growing customer base, hardware requirements are moderate and relatively easy to learn. It has some weakness – not so good documentation, model design GUI is not detailed enough to be useful, simulation results reporting is not adequate.
- **GLASS-** GMPLS Lightwave Agile Switching Simulator is a Java based network simulator [11]. The simulator uses the basic framework of the Scalable Simulation Framework (SSF) with its extension SSFNet. Although this is as the name implies a simulator that is developed for GMPLS simulations, it can be used for MPLS simulations as well. The simulator has implemented MPLS forwarding and label distribution with LDP, CR-LDP and RSVP-TE. It is designed for studying IP+MPLS network and multilayer resiliency, and growing customer base. But documentation is not adequate, a relative new widely used simulation tool, smaller library of rebuilt modules and protocols than in other simulators, simulation result reporting is not adequate, poor post-technical support.
- **OPNET** [12] – This simulator has large library of simulation models of communication protocols and equipment, professional support and good documentation. The main disadvantages are relatively high price and it doesn't support on-line analyses if simulation output data.
- **NS-2** - (Network Simulator 2) is a discrete event simulator targeted at networking research [13]. The simulator has been in development since 1989 and there are many

different contributed models available for the simulator. This simulator is also open source and there are a lot of users that have contributed to make this simulator. The main implementation is made by the VINT project at BL, Xerox PARC, UCB, and USC/ISI. NS-2 is coded in C++ and TCL. The simulator has one contributed MPLS module for NS-2. This module is called MNS (MPLS Network Simulator) and was developed by Gaeil Ahn. This model had implemented MPLS forwarding and label distribution by LDP and CR-LDP. The model has also some functionality implemented for recovery mechanisms, like the ability to setup a backup path and associate it with a working path. Unfortunately this model did not include an implementation of RSVP-TE. The main advantages here are: well known and widely used, powerful and flexible scripting and simulation setup, many protocols implemented, available source code, easy to extend. But some things are no so good: some protocols and features are not well documented, not well known technical support, patching extensions in is not easy.

The results of comparison are presented in Table1.

TABLE I
COMPARISON OF MPLS SIMULATORS

	J-Sim	**OMNeT++**	**GLASS**	**OPNET**	**NS-2**
VoIP protocol stacks	missing	satisfy	missing	satisfy	satisfy
MPLS-TE and RSVP signaling protocols	poor	poor	excellent	excellent	satisfy
Models of typical failures and different routing protocols	satisfy	satisfy	satisfy	excellent	excellent
Models of optical layer components	satisfy	satisfy	excellent	satisfy	poor
Models of different teletraffic scenarios	poor	satisfy	excellent	excellent	satisfy
QoS mechanisms	satisfy	satisfy	satisfy	excellent	excellent
Different network architectures	excellent	excellent	excellent	excellent	excellent
Analysis of typical performance measures	poor	poor	satisfy	excellent	satisfy
Credibility of simulation models	missing	excellent	missing	satisfy	satisfy
Quality of sources of randomness	poor	excellent		poor	excellent
Analysis of simulation output data		excellent		poor	excellent
Ability of extending or adding new simulation models of protocols	excellent	excellent	excellent	excellent	excellent
Existence of GUI, which be used as input/ output interface	excellent	excellent	excellent	excellent	poor
Existence of a good manual and an introductory tutorial	poor	satisfy	poor	excellent	excellent
The cost of license	excellent		excellent	poor	excellent

As it is possible to see in the Table1, none of these simulators satisfies all requirements. This problem could be solved by designing a custom-made simulator. The best start of research to this new simulator is with OPNET,OMNET++ and free of charge NS-2.

IV. CONCLUSION

Multi Protocol Label Switching (MPLS) is being used in many corporate networks and public infrastructures and as a backbone technology of many Autonomous Systems. Because of its importance, what is needed is to find out simulators able to simulate MPLS networks whose results reflect the real environment as much as possible. It is possible to check the level of reliability provided by the simulator. These results could be taken into account later to simulate real-time critical services over MPLS networks.

Simulation models are credible if they are valid and verified. A model is valid if it represents a given system accurately, at the required level of details. Verification is concerned with determining whether the simulation model has been correctly translated into a computer program. If simulator is free of charge, nobody takes responsibility for offering valid models encoded in verified programs.

Assuming that a network simulator uses credible simulation models, credibility of the final results it produces depends on the quality of its sources of randomness of failures or statistical representativeness of traces of real traffic and statistical accuracy of the final simulation results.

Extendibility of simulator with new features to existing simulation models is important, because the amount of work or time which is needed to extend the existing simulation models must be less.

A user-friendly simulator should be equipped with GUI, because during simulation the GUI could be used for showing evolution of simulated processes and intermediate values of analyzed performance measures.

In this research are presented simulators, which are not suitable to satisfy all these requirements. This problem could be solved by designing a custom-made simulator. This

simulator will be supporting tool to research projects related to MPLS as well as in teaching subjects related to this technology.

ACKNOWLEDGEMENT

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

REFERENCES

[1] M. Kodialam. T. V. Lakshman. Restorable Dynamic QoS Routing. IEEE Communications Magazine, Vol 40, Issue 6, June 2002, pp 72-81.

[2] M. Domínguez-Dorado, F. J. Rodríguez-Pérez, J. Carmona-Murillo, J. L. González-Sánchez "Educational Improvements Applying an MPLS Network Simulator: A Technical Approach",IEEE Multidisciplinary Engineering Education Magazine, VOL. 2, NO. 4, DECEMBER 2007, http://www.ieee.org/edsocsac

[3] Gaeil Ahn, Woojik Chun, "Overview of MPLS Network Simulator: Design and Implementation", http://heim.ifi.uio.no/~johanmp/mpls/MNS_v1.0_arch.pdf

[4] V. Sharma, Ed. Metanoia, Inc. ,etc. "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", http://www.faqs.org/rfcs/rfc3469.html

[5] Maria Hadjiona, Chryssis Georgiou, Maria Papa, and Vasos Vassiliou "A Hybrid Fault-Tolerant Algorithm for MPLS Networks", Wired/Wireless Internet Communications, Volume 5031/2008, Tuesday, May 20, 2008, http://www.springerlink.com/content/q01x1q865337t8t0/

[6] http://www.cosc.canterbury.ac.nz/research/reports/TechReps/2006/tr_0605.pdf

[7] http://www-entel.upc.es/xavierh/mpls (07/05/2007).

[8] http://www.j-sim.zcu.cz/

[9] http://www.omnetpp.org/

[10] http://www.tdx.cesca.es/TESIS_UPC/AVAILABLE/TDX-0323104-125205//Chapter2.pdf

[11] http://www.glasssimulator.com/simulators.html

[12] http://www.opnet.com/solutions/

[13] http://www.isi.edu/nsnam/ns/

E-content for Computer-Aided Design of Communication Circuits with PSpice Simulator

Galia I. Marinova[1]

Abstract – The paper presents the e-content of a Knowledge base with algorithms for analyzing different classes of communication circuits in PSpice simulator. Algorithms are developed for 22 classes of circuits and there are more than 200 circuits in the Knowledge base.

Keywords **– PSpice, Communication circuits, Algorithms for analysis, Knowledge base, E-content.**

I. INTRODUCTION

The PSpice simulator is a basic tool in university classes of Electrotechnics, Electronics and Communications. Different teaching environments integrating the PSpice software are presented in [1, 2, 6, 7, 8].

PSpice simulator is one of the tools used to teach Computer-aided design – a third year bachelor degree course in the Telecommunications Faculty of Technical University – Sofia. The goal of the course is to give students the abilities to design communication circuits and systems integrating knowledge from several previous and parallel courses:

- Circuit theory methods from the course of Electrotechnics;
- Analog and digital filter design from the course of Communication networks;
- Analog and digital circuits design from the courses of the same name;
- Radiocommunications and Radiocommunication systems courses.

TABLE 1

Class of circuits	Number of circuits	**Circuits types or examples**
Filters with operational amplifiers	35	Law pass, High pass, Band pass and Stop band filters with different approximations
Passive LC filters	10	Example: Band pass filter for communication antenna system
Digital filters	10	FIR filters with one or more stages [9,10]
Equalizers	2	Examples for 4 and 10 frequencies
Filters with quartz crystals	10	Cohn crystal ladder filters with 3 or 4 stages [12]
SC filters	2	Low pass filters, Integrators [5]
Filters with transmission lines	15	Realized with different number of transmission line elements [9]
Generators	15	Square wave, triangle wave and sine wave realized with transistors, 2 or 3 operational amplifiers
Generators with quartz crystal stabilization	4	Sine wave generators
Arithmetic circuits with operational amplifiers	20	Integrator, Differentiator, Comparator, Summing and difference amplifier, Logarithmic and exponential amplifier, Square root circuit
Voltage regulator circuits	10	Linear voltage regulator circuit with BJT at the output, Voltage regulator circuit with output BJT in switch mode, Voltage regulator circuit based on direct converter principle , with output BJT in switch mode, Voltage regulator circuit with output MOS transistor in switch mode [2]
Resonance amplifier	3	Realized with transistors
Amplifiers	5	Realized with transistors and with operational amplifiers
Programmable gain amplifiers	3	Realized with switches
Circuits with timer LM555	25	PWM circuit, Schmitt trigger, Basic astable oscillator, Monostable timer [11]
Digital combinatorial circuits	10	Logical circuits
Arithmetic digital circuits	10	Adders and multipliers
Digital circuits with memory	5	Latches, Flip-flops
Radiocommunication circuits with transmission lines	10	Mixers, Modulators, Demodulators, Wilkinson power divider /adder, Directional coupler, Hybrid ring, Balun [1]
VCOs	2	Realized with OA and with timer LM555 [11]
Mixers, balance modulators	2	Realized with transistors
Frequency converters	2	Realized with transistors

[1] Galia I Marinova is with the Telecommunications faculty, Technical University – Sofia, Sofia-1612, 8, bul. Kliment Ohridski, E-mail: gim@tu-sofia.bg

The course of Computer-aided design is structured in lectures, laboratory exercises and individual or group projects. All these aspects are evaluated and then students are estimated.

In order to give students a global view, closer to real applications and the ability to solve a large number of design problems, a Knowledge base with algorithms to analyze different classes of communication circuits is developed.

II. E-CONTENT IN THE KNOWLEDGE BASE

The e-content serving the Knowledge base in the Learning environment is structured following the general methodology for circuit design with PSpice and gives the specific step-by-step algorithms for each circuit class.

The main steps in the algorithms are:

- Choice of circuits from a circuit class suitable for the desired application.
- Determination of circuit element parameters for meeting the design specification and constraints. For some classes of circuits on-line calculators are available and they can be used, for example for crystal filters [14] or timers with LM555 circuit [11].
- Creation of project, designs and schematics for the circuits in OrCAD Capture.
- Choice of the input sources – voltage or current stimulus in analog or digital form.
- Definition of the set of analyses to be performed in order to obtain the representative curves for the circuits.
- Definition of the set of curves to be visualized by the graphical postprocessor Probe.
- Definition of the set of parameters to be extracted from the curves and from the output file.
- Definition of the criteria for estimation of circuit properties and for comparison.
- Conclusions for the circuit performance, its area of application, the restrictions for its application, optimization advices and advices for multisolution synthesis.

Algorithms are developed for 22 classes of circuits and a set of circuits taken from books [3, 4] and internet sites [12, 13] is considered in each class. The Knowledge base contains more than 200 circuits. The e-content of the Knowledge base is presented in Table 1.

III. ILLUSTRATIVE EXAMPLES OF E-CONTENT IN THE KNOWLEDGE BASE

Three examples from the e-content of the Knowledge base are given on Fig. 1, 2, 3:

Fig. 1 presents a double crystal filter in radio receiver from [13] – the original circuit, the circuit with RLC models of the crystals, then AC curves and the parameter values for the resonance frequency, the low and high cut off frequencies and the frequency band.

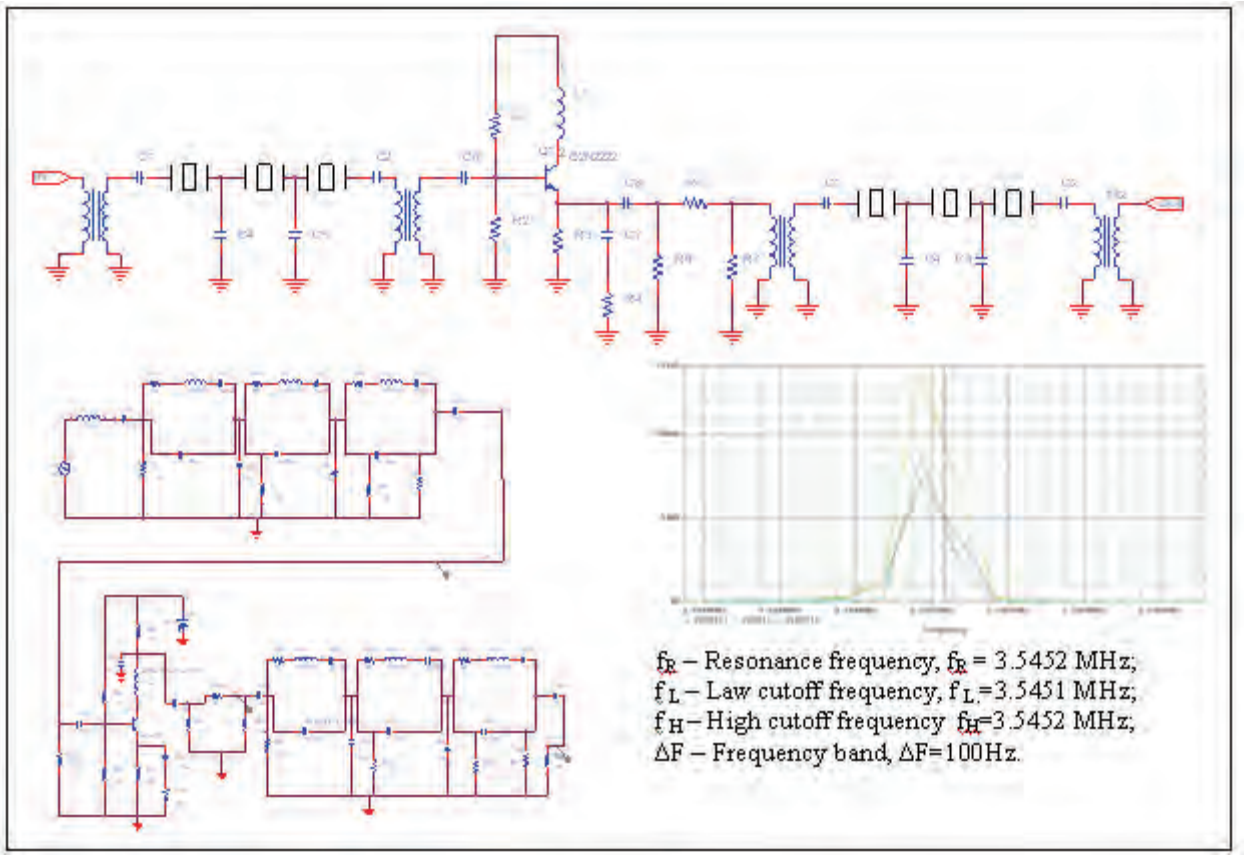


Fig. 1. Double crystal filter in radio receiver – original circuit, circuit with RLC model of the crystals, AC curves and parameters

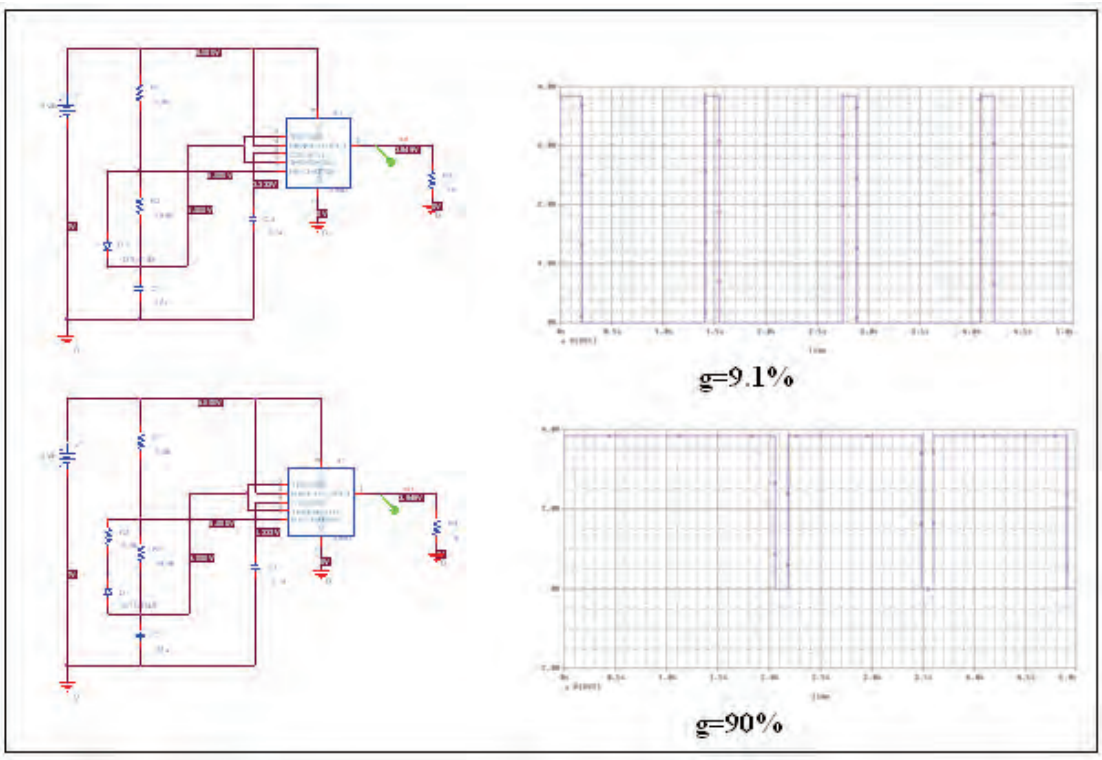


Fig. 2. Pulse generator circuits with short and large ON time. Circuits, output voltages and output duty cycle g=Ton/T, T- pulse period

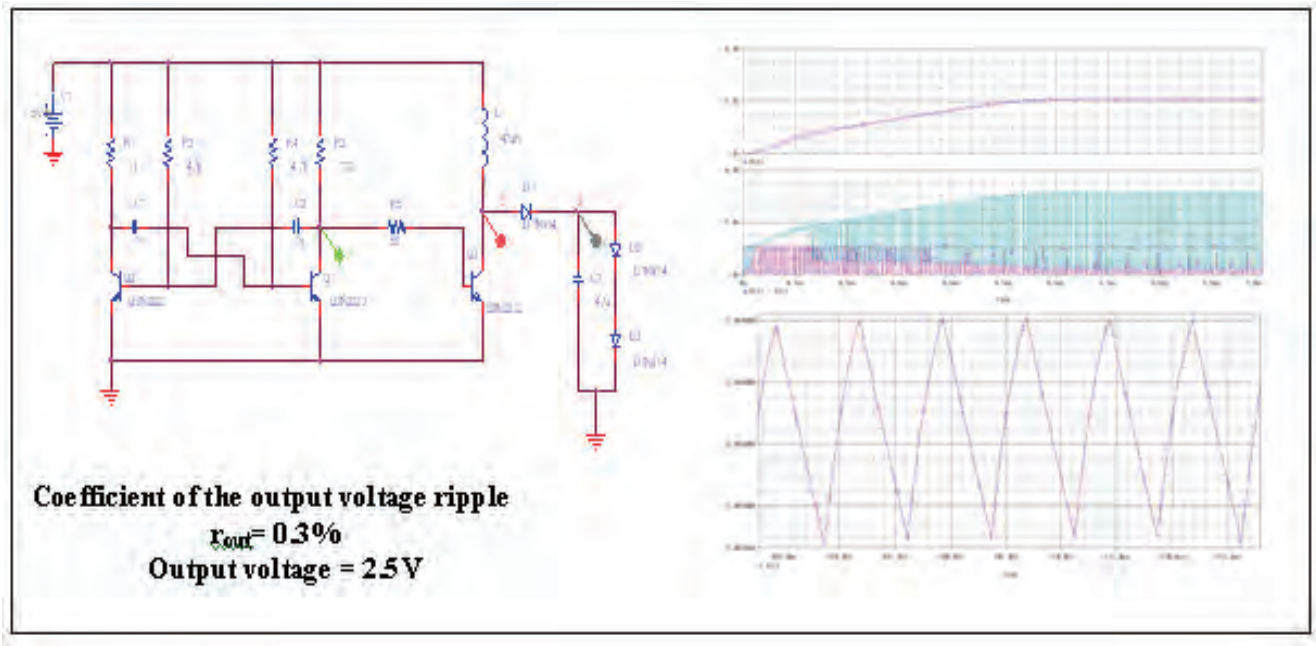


Fig. 3. Boost voltage regulator – circuit, voltages in nodes 1, 2 and 3.
Circuit parameters: output voltage and coefficient of the output voltage ripple.

Fig. 2 presents pulse generator circuits with short and large ON time - circuit, output voltages in time domain and output duty cycle coefficients

$$g=Ton/T,$$

where T is the pulse period.

Students are expected to give proposals for the possible applications of the two circuits.

Fig. 3 presents a boost voltage regulator – circuit, voltages in nodes 1, 2 and 3 in time domain and circuit parameters - output voltage and coefficient of the output voltage ripple.

Students are expected to conclude whether the circuit meets the restriction for voltage regulator circuits:

$$r_{out}<1\%$$

They should also check whether the boost regulator circuit can aliment a LED with 1.5V battery.

During the regular course students simulate about 10 different circuits to get familiar with the PSpice design methodology and then they work on individual or group projects chosen on a random principle from the Knowledge base.

IV. Conclusion

The e-content of the Knowledge base presented in the paper helps solving a great variety of design problems using PSpice simulator which is useful for students and engineers. It serves as well to illustrate effectively the lectures in the course of Computer-aided design.

Besides the PSpice simulator, students use also OrCad Capture and OrCad Layout tools [15] for schematic capture and PCB design of the different circuits in the Knowledge base.

In forth year they have an additional optional course for Computer-aided design of digital communication circuits with VHDL where they study design on programmable logic with tools as WARP 6.2 [16], ISE WEBPACK [17] and QUARTUS II [18]. Additional e-content is developed for these software tools applications in a larger Learning environment for Computer-aided design in communications.

References

[1] A. Rusek, Barbara Oakley, "PSpice Applications in the Teaching of Wireless and High Frequency Electronics", *Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition*, American Society for Engineering Education, 2001

[2] G. Marinova, D. Dimitrov, "Learning optimal synthesis of voltage regulator circuits through comparative study in PSpice", *XV International Symposium on Theoretical Electrical Engineering*, 22-24 June 2009, Lubeck, Germany, pp. 211-215

[3] G. Marinova, *Guide for computer-aided design systems*, Filvest, Sofia, 1999

[4] E. Gadjeva, T.Kumdjiev, S. Farhy, *Computer modeling and simulation of electronic and electric circuits with OrCAD PSpice*, Meridian 22, Sofia, 2001

[5] E. Shoikova, I.Pandiev, "Behavioral PSpice macromodel of universal SC-filter LTC1060(MF10)", *Electronics*, Sozopol, 2001

[6] K.A. Connor and J.Braunstein, *Fields and waves I*, 2.19.2002/31.01.2003, Rensselear Polytechnic Institute, Troy, New York, USA, ESCE-2100/Spring 2002

[7] Kh. S. Al-Olimat, "Shifting Teaching and Training to Online Learning: PSPICE through CAMTASIA"**,** *Proceedings of the 2008 ASEE North Central Section Conference*, American Society for Engineering Education, 2008

[8] P. Tobin, "The role of PSpice in the engineering teaching environment", , *International Conference on Engineering Education – ICEE 2007*, Coimbra, Portugal September 3 – 7, 2007

[9] P. Tobin, *PSpice for Filters and Transmission Lines*, Book 2, Morgan Claypool publishers, Mar 2007.

[10] P. Tobin, *PSpice for Digital Signal Processing*, Book 5, Morgan Claypool publishers, Mar 2007.

[11] http://home.cogeco.ca/~rpaisley4/LM555.html

[12] http://www.cliftonlaboratories.com/cohn_crystal_filter.htm,

[13] http://www.pan-tex.net/usr/r/receivers/sfilter.htm#dual

[14] http://www.giangrandi.ch/electronics/crystalfilters/xtalfilters.shtml

[15] www.cadence.com

[16] www.cypress.com

[17] www.xilinx.com

[18] www.altera.com

On-line system for Testing and Certification

Milena N. Karova[1] Julka P. Petkova[2] Emil U. Kerekovski[3]

Abstract – **The on-line system is designed to close the cycle of e-learning, providing the opportunity to test out individually and remotely. Teachers in different disciplines create test questions on an issue, then activate them on an individual student, as determined during the activity (time in which to be completed). The certificate includes information about the test - title and total number of questions for the student data - name and number of faculty, evidence of completion - time and time initial evaluation and assessment in points, rates and assessment words and numerals in the six-grade system assessment.**

The system is programmed using ASP.NET, SQL Sever 2005 and C#.

Keywords **–on-line test, e-learning, certificate, test questions, validation, database.**

I. INTRODUCTION

One of the main objectives of the global network is easier for consumers to search information. As present the Internet very quickly goes out of its original close-oriented application. It is very important to use Internet for studying, to access information on mobility, especially if e-learning offers. E-learning provides many benefits: student mobility, the opportunity for individual pace of preparation, training at convenient times for student work-based learning; significantly cheaper cost, continuing education, complied with the requirements of students with disabilities. Specially created data base allows 24-hour access to learning materials, resources, documentation and Internet resources and selected bibliography.

This on-line system is designed to close the cycle of e-learning, enabling the test to be conducted individually and remotely. Teachers in various disciplines create test questions on an issue, then activate them individually for each student, as determined during the activity (time in which to be completed) [2]. After activating the test, students completed them and immediately after their skills are assessed by the system. At the request of teachers, the time of each test may be limited. After assessing the test, students and teachers can download a certificate for passing the test in PDF format. The certificate includes test data - title and total number of questions for student data - name and number of faculty, completing data - initial time and during the assessment and evaluation in points, rates and assessment in words and figures on the six-grade system assessment. This system stores all data for students' tests. Each student can see the results of all their tests and to download certificates at any time. Teachers also have access to the students' results.

[1]Milena N. Karova is with the Department of Computer Science, Technical University Varna, Bulgaria, E-mail: mkarova@ieee.bg

[2]Julka P. Petkova is with the Department of Computer Science, Technical University Varna, Bulgaria, E-mail: jppet@abv.bg

[3]Emil U. Kerekovski is student master with the Department of Computer Science, Technical University Varna, Bulgaria, E-mail: kerekovski@gmail.com

II. SYSTEM DESIGN

A. Main functions of on-line system

There are three important functions: user management, login and access management and create, activate and complete the tests.

The first includes registration of the user (requires the introduction of basic user data: user name, password, name, number, and faculty e-mail address) and adjust your own profile. To this module of the system includes functionality to recover a forgotten password. Also included here is the complete management of users by system administrators (adding, activation / deactivation, profile correction, and change the access level of the user).

The second function traits the three levels of access - a system administrator (users with rights to administer the data in all modules of the system), a group administrator (user with rights groups to administer the training and testing - in particular teachers); ordinary users (users' rights to use the materials in groups for training and testing to make - in particular students). The login is done by entering a user name and password. Enhanced security password is encrypted. The menu is loaded dynamically from a table in the database, which describes the access rights of different types of users to different pages.

System administrators group and administrators should be able to create tests for on-line testing and activate them for ordinary users. Viewers must complete tests and receive results immediately after completion. The results should be displayed on the screen with the option to download the certificate examinations, including the assessment thereof, in the form of a PDF document.

B. Database Design

To store and to data manage, Microsoft SQL Server 2005 is used. Microsoft SQL Server 2005 is a platform for management and data analysis, which offers enhanced security, opportunity for expansion and accessibility of data and applications for analysis of the right place at the desired time [3]. In addition, SQL Server 2005 makes it easy creation, deployment and management of applications. The name of the base and the location and name of the server are not relevant, since the details of the connection to the server are exported to the application configuration file called web.config in

connection string. The base consists of 13 tables [1]. Links between tables are constructed based on relationships using foreign keys and relations of type m:n related tables are used.

The contact between ASP.NET application and the database uses the library DBEngine classes for Microsoft. NET Framework 2.0 [4]. The library includes methods, extraction line (DataRow), rows of a table (DataTable), rows of multiple tables (DataSet). The methods to perform queries returning result of scalar type are also included. A class BaseAdapter is an object of class BaseDataProvider DBEngine library and it is inherited by classes which are formally called adapters (Adapters). These classes are divided by topic and contain methods, calling methods from the library and DBEngine responsible for specific topics [Fig. 1.]

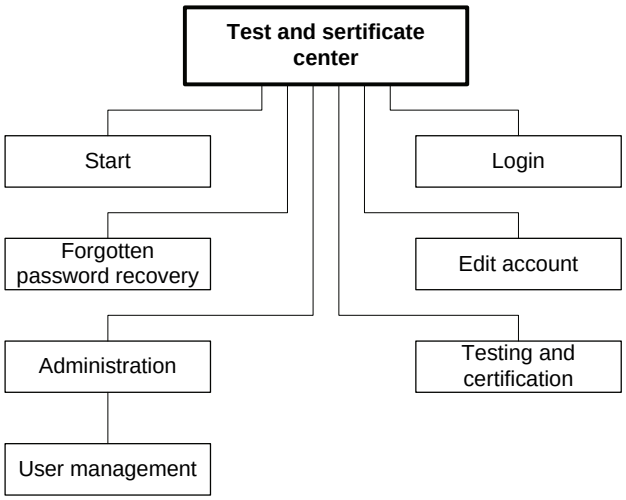


Fig. 1. Main functions of system

C. Interface Design

The login requires user name and password. For greater certainty in the database (DB) the password is stored in encrypted form and compared. It is encrypted using one-way hash function (Secure Hash Algorithm).

The recovery of password requires a entering of the correct username and the e-mail address to generate a new random password of 10 characters. It sent an e-mail address of the consumer and he can use it to login. Each user can edit his profile, which includes change data and change the password. The username is unique and not subject to change. To change the password, it is requiring the introduction of the old password and the introduction of the new password twice.

System administrators can register users, and the users themselves can fill the registration form. The difference is that an individual registered user receives the lowest level of access (regular user) and accounts are activated, while system administrators can change the access level to enable or disable accounts.

In the section testing and certification, the system administrators create and group tests. They enter the total information of the test (title, subtitle, total number of questions of questions per page and maximum time for its completion). Then they fill information about the issues one by one. Each issue is filled with text, answers their texts and points (ten responses). Tests can be edited completely while not yet filled out by students. Once they were completed, tests are also subject to revision, but rather limitations: can not change the number of questions and the number of points and their answers

Once you create a test, it can be enabled for normal users (students). System administrator can activate the test for students, regardless of which groups are connected, while the group administrator can activate your test for students connected to their groups. The test can be activated for a specified number of hours (1, 2, 3, 6, 12, 24 or 48 hours). The test can also be disabled for a user.

On the evaluation page the user is able to download a certificate of test solutions in the form of a PDF document. It contains the same information that is contained on the page. To create a PDF document it uses the library for Microsoft.NET - iTextSharp, which is distributed free for non-commercial purposes. It contains methods for creating the document and adding elements to it (text, paragraph and image). It also uses the library PDFCreator, which is in direct interaction with iTextSharp and contains methods to build the document, interpret the HTML stream (HTML Parser).

In the section testing and certification ordinary users (students) may consult the results of all tests made by them and to download the certificate of each test was evaluated. System and group administrators can also see the results of students grouped by tests, and to download certificates.

III. Creating, Change, Assesment and Certfication of On-line Tests

The management of the tests carried out by several forms, from the menu access page Testing.aspx (test). Fig. 2. shows the structure of the forms involved in this section.

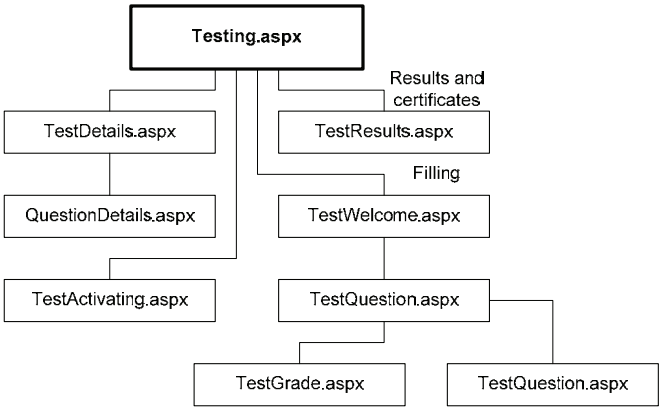


Fig. 2. Structure of system's forms

The tests can be created and edited by users who have a type system administrator or group administrator. Page Testing.aspx is pagination GridView control, where are the tests. This control has a different functionality for different types of users, system and group administrators can manage (create, edit, activate and deactivate) test and examine the results.

For editing the test it clicks the icon for editing.The next page QuestionDetails.aspx, introduces or edit data themselves questions. Index edited or added to the current issue is kept in ViewState container. On this page, add or edit the questions one by one. Once completed with the introduction of issues, all data is sent to the database in a transaction, taking an object of type TestsDataSet, which are stored and transported so far. Data are validated by the matters necessarily have at least 2 possible answers and all answers have completed points that will bring answers to fill.

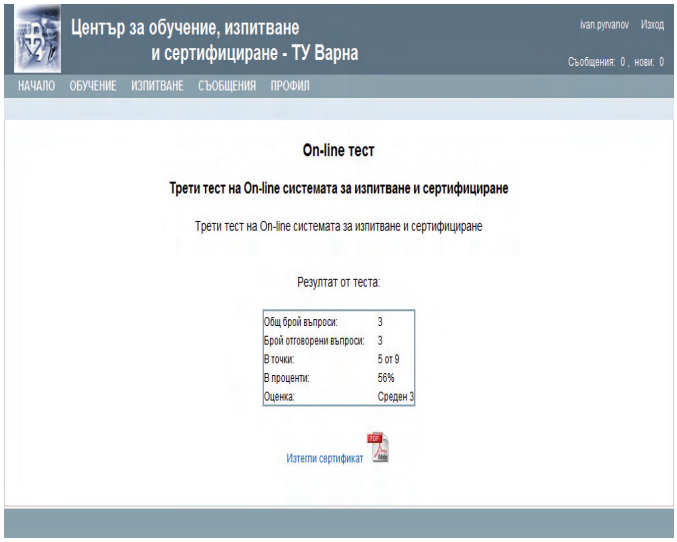


Fig 3. Test evaluation page

The test activation for a user(s) from a group starts on page TestActivating.aspx. Students can see the results and download certificates at any time [Fig. 3.]. The information remains available in the database and is not intended to be deleted. For the preparation of the certificates in PDF format is used the iTextSharp library NET Framework. This library is free to use for non-commercial purposes. It is added to another library that interprets the HTML stream and methods used iTextSharp to create PDF document. An HTML stream is made in which is based a certificate template. A class is created with only static PDFReports method TestCertificatePDF. This method derives from the database necessary data using the completion test identifier (TestTokenID). It creates a PDF document from HTML stream and return it as a result of object type MemoryStream [Fig. 4.].

All input data in the system are validated. It is used the built-in ASP.NET controls validating - RequiredFieldValidator (request an introduction of value in a given field); RegularExpressionValidator (validate value using regular expression); CompareValidator (compare with another value and check its type); CustomValidator (used in validated more complex logic).

The main part of the business logic of the system is given in stored procedures in the database. The idea is to perform all special features there and to filter the rows derived from the application. Thus, the traffic between the database and the application is minimized. It also adds to increase the system performance. All data operations, performed in Microsoft SQL Server environment, are running significantly faster than in the NET environment.

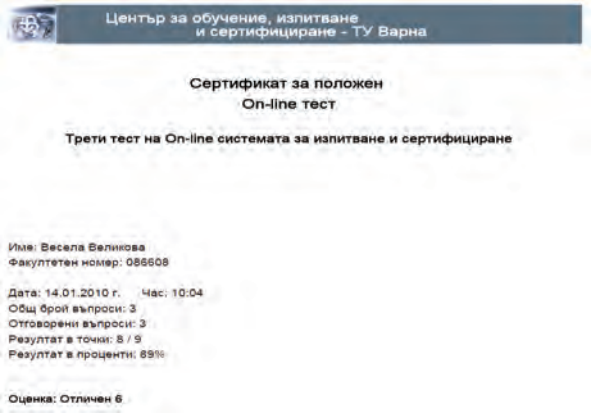


Fig. 4. Sample of certificate

IV. Conclusion

E-learning is a modern and interesting form, the result of developments in information technology and communications. Verification of knowledge during training and after its completion is a crucial stage of the training itself.

This system closes cycle of e-learning, provides an opportunity to create tests that students complete and receive certificates for their results. The features of this system are: creation of test questions that have one possible answer; possibility to select the size of the pages with the questions; possibility of limiting the time for completing tests; activation of tests for students in groups as well as selecting the time of activity; evaluation of completed tests immediately after their completion, even then the students learn their results; possibility to download the certificate affixed to a test of each student.

This system is composed to add and develop the needs of the school. Here are mentioned some of the more important aspects of such a future development:

•A multilingual education system interface can be translated into different languages to facilitate the students.

•It can provide an interface for teachers to translate the tests in different languages.

•For each test could be created set of questions (with a larger number of test requirement) and the questions are selected randomly and mixed.

•It is possible to add different systems for evaluation as required by the school for foreign language education

•It is possible to create a module that controls the content of static pages (eg home page). The content may be stored in different languages in a database and may be adjusted at any time through the system administrator interface.

The mobility and the ability of tests to be completed on-line from remote computers units are factors that greatly assist the participants in distance learning and disadvantaged people. The system has a very intuitive interface and it is easier to use, no further knowledge. No special hardware and software requirements for its use, making access to it free and available anytime, from anywhere, using any type of computers and

mobile devices. It is necessary to equipper with a browser and have Internet access network.

REFERENCES

[1] Ernandes, M., Database Design, Softpress, Sofia, 2004

[2] Gradinarova B., Knowledge Assessment Exercises in Adaptive E-Learning Systems, International Scientific Conference Computer Science'2008, Kavala, Greece ISBN:978-954-580-256-0

[3] Malik, S., Pro ADO.NET 2.0, Apress, 2005

[4] Nakov S., NET Framework Programming, Faber, Sofia, 2005-2006

Some Important factors for Designing Semi-Autonomous Learning Environments

Boyka Geadinarova[1]

***Abstract* –The aim of this research is to propose solution of the question, "in what ways do mediated learning environments support or hinder learner autonomy?"**

Learner autonomy is an important factor in the success of mediated learning environments. The central aspect of learner autonomy is the control that the learner exercises over the various aspects of learning, beginning with the decision to learn or not to learn.

The first are the motivational-intentional forces that drive the learner to apply some determination to the act of learning. They are the cognitive functions of learning and include learner initiative, motivation and personal involvement.

The second area of learner-control is the one comprising the "nuts-and-bolts" of the act of learning, such as defining learning goals, deciding on a learning sequence, choosing a workable pacing of learning activities, and selecting learning resources [5]. We will examine the implications of each of these areas of learner-control, and share our analysis of a series of interviews with cyber-learners, based on this framework of cognitive, algorithmic, semiotic and economic factors.

***Keywords* – self-directed learning, learner autonomy, educational policy, international development.**

I. INTRODUCTION

Early attempts at defining features of Distance Education have stumbled upon an interesting conundrum. Some authors [6], pointed out that printed material mailed to a distant location increased pressure on the learners to set their own schedule and to work around deadlines imposed by the teaching institution. This feature, juxtaposed to the fact that distance education shows one of the highest drop-out rates among all educational environments, led to the supposition that distance learning requires some higher degree of learner autonomy than traditional classroom instruction. Indeed, lack of autonomy was considered the main reason why students failed or discontinued their programs.

Another feature of Distance Education was identified as the constraint imposed on institutions to produce a standard learning program that will be followed by all learners in the same sequence, usually within a set of prescribed deadlines [4]. This institutional standardization, inevitably, is then passed on to the learner. In this respect, distance learning environments can be said to *constrain* the expression of autonomy among learners and instructors alike, at least when compared with traditional environments where components of the program may be modified in response to learner feedback or other considerations.

The question of whether a specific learning environment will support or hinder the expression of autonomy is an important one for educators. Contemporary literature in adult education has focused on learner self-direction as a core value associated with the notion of facilitation, rather than the dispensation of learning [7].The point here is not to retrace the steps that led to the emergence of learner-autonomy as a strongly held value among adult educators, but merely to situate our study within its context. It is our view that the quality of any learning environment is to a significant extent dependent on the degree to which that environment acknowledges the need to support learner self-direction.

Interestingly, these quotes point to very different aspects of learner autonomy, the central one being the *control* that the learner exercises over the various aspects of learning, beginning with the decision to learn or not to learn. But as Candy points out, there are other specific areas where learner-control can be exercised.

II. PEDAGOGICAL END PSYCHOLOGICAL ASPECTS OR THE AUTHORS

The question, then, is to investigate the "areas" of learner control. How many are there? How do they intersect with the specific features of D. E. environments? What are their implications for adult learning?

According to [8], the first area includes the motivational-intentional forces that drive the learner to apply some determination (or "vigour") to the act of learning. What Huey Long called the 'psychological' aspects of learner autonomy will be referred to here as the *conative* functions of learning.

The second set of elements identified by [8] as a subset of learner autonomy were the "pedagogical" aspects of learning. These involve the control over the "nuts-and-bolts" of the act of learning, such as defining learning goals, deciding on a learning sequence, choosing a workable pacing of learning activities, and selecting learning resources [5]. These elements can be grouped under the more precise heading of *algorithmic* aspects of learning. In traditional learning environments, most of the algorithms are the responsibility of a teacher or a teaching institution. Learning goals, student workload and methods of evaluation are usually stipulated at the outset and little participation in their formulation is expected from the learner. Any derogation from this approach entails devolving to the learner, on top of the expected "learning tasks", at least

[1]Boyka Zh. Gradinarova is with the Faculty of Computing and Automation, 1,Studentska Str. Varna,Bulgaria, E-mail: Bgradinarova@tu-varna.bg

some of the "teaching tasks" normally reserved for the instructor. In this sense, we can say that autonomy is directly related to the number and magnitude of the "teaching tasks" that are appropriated by the learner [10]. Most mediated learning environments require such participation from learners, albeit to different degrees and with varying results as will be described below.

A. Emerging dimensions

Just a few years ago, learner control was necessarily limited to these two sets of features, conative and algorithmic. After deciding whether, what, and how to learn, one had covered all areas where it was conceivably possible to exercise some degree of learner autonomy. Now with the proliferation of learning environments that include mediated instruction materials, exponentially available learning resources, new means of communication, and a marketplace literally exploding with learning opportunities, two other components of learning emerge as possible areas where learner control may be exercised – or impeded. We have namely identified the *semiotic* dimension learning, and the emerging *economics* of the knowledge marketplace.

Until recently, the prevalent medium for encoding, storing and disseminating knowledge was to provide access to print materials through libraries, mail-order programs, or custom-printed resources. Today, learning materials include rather diverse media which may share very few features with printed text. For example hypertext, asynchronous messaging and electronic whiteboards each possess their own set of codes and behaviors that are inconsistent with the linear quality of print. Furthermore, the manner in which each new medium is utilized by instructors and learners varies to some extent, leading to further diversification in the perception of their semantic possibilities [3]. For example, hypertext can be used as a way to link course materials to outside resources, or as an inherent part of the material to be learned, or then again as non-compulsory enrichment to the basic text such as illustrations or diagrams to be viewed when needed. From the learner's perspective, hypertext can be perceived as a convenient way to store and retrieve information, or as a bothersome irritant leading to feelings of frustration in the presence of overwhelming amounts of poorly organized data. Because each environment offers its own set of communication pragmatics and its own approach to using them, we can say that the semiotic choices made by designers and instructors are an integral part of the learner's experience, and as such offer opportunities to enhance or deter learner autonomy.

Learning is no longer the reserved province of traditional institutions such as schools or colleges. Indeed, it is now acknowledged that universities find themselves in direct competition not only with each other, but with a multitude of offerings from a thriving marketplace [9]. Today an important component of any learning environment is the perceived economic value of its knowledge in the marketplace, either as an asset for finding employment or as a means of production in the knowledge economy. Based on this consideration, learners must not only decide why and what to learn, but also where to learn it and who to learn it from. This decision will surely be based on factors like individual preference for a proposed learning environment, but ultimately the choice will rest on the perceived cost-benefit and opportunity cost which are generated by each alternative. In this context, we can observe that the *economics* of learning are emerging as an important component of learning environments.

The diagram in Fig.1 illustrates how learner autonomy can be divided in four areas of learner control: conative, algorithmic, semantic and economic. One useful feature of this representation is that it makes it possible to explore learner perceptions within different learning environments, while retaining a constant framework for analysis.

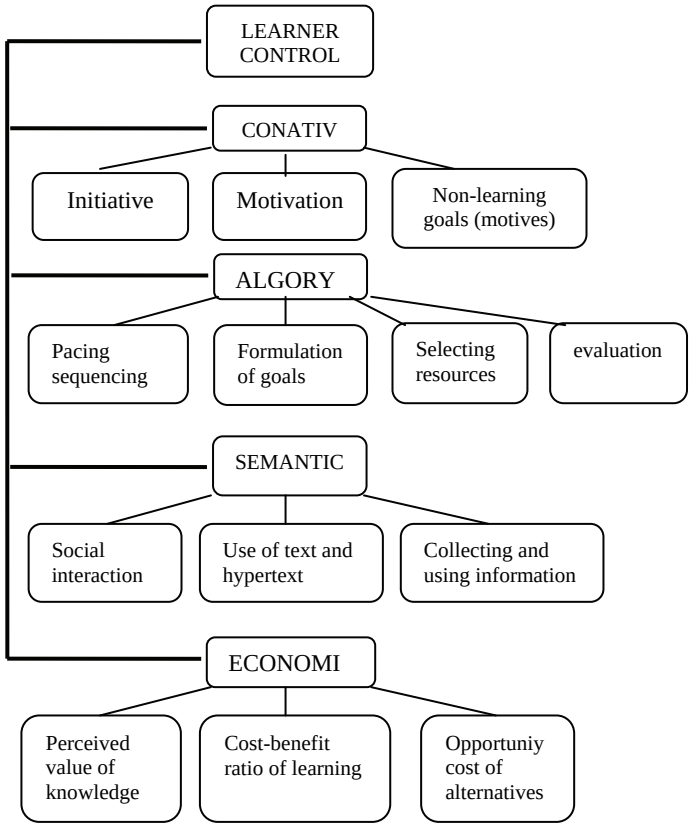


Fig. 2. Scope of learner autonomy

III. METOHODOLOGY

Open-ended interviews were conducted with 13 adult students registered in on-line courses in Psychology, Finance, Education, and Political Science. Questions were formulated to explore the four areas of learner autonomy, using everyday terminology familiar to the students.

The research question to be explored using this method was, more specifically,

> *"In what ways do mediated learning environments hinder or support the emergence and expression of learner autonomy?"*

The questions asked the informants were not as formally wordedl. We conducted semi-guided interviews generally purported to get some feedback from the students on their learning experience in on-line university courses. Interviews were taped, transcribed, coded and analysed using standard content analysis techniques. Coding was done by grouping units of meaning under tentative headings, and then combining the headings under generic titles using an emergent design method. The factors that determined the students' perception of each learning environment were categorized as: Interaction; Structure; Value; Context; and Media.

IV. FINDINGS

One of the first things that became apparent during the interviews was the diversity in the likes and dislikes of individuals concerning the various components of the learning environments. All courses were designed using online course materials, a messaging device and a textbook. Some used hyperlinks to other web-based resources, and none included classroom or face-to-face meetings. The individual preferences were polarized around the use of media, the course structure and the value of course content. Predictably, about half of the students said it took some effort to prevent inertia when facing the prospect of doing tasks online, in the absence of an imposed schedule.

A. Interaction

The portion of the course grade allotted for participation in on-line discussion groups varied between 0% and 40%. In the groups where there was less pressure to 'participate', students felt that the interaction was more meaningful and that they were more in 'control' of the environment. From the students' responses, it appears that instructors did not attach any weighting to the quality or tone of the interactions. This finding supports the notion that the *conative* aspects of learner autonomy in D. E. environments need to be further analyzed to include the subsets that are specifically linked to the characteristics of the environment itself. For example, while social interaction has been found to be one important factor in the motivation of learners, the type of interaction provided by chat-groups, e-mail and moodles need to be further explored.

B. Structurs

A few students who had a personal interest in their course topic spontaneously searched for alternative learning resources but overall, this was not a prevalent practice. All courses except one had set very specific objectives, thereby circumventing student participation in their formulation. When a learning goal was stated in general terms, mature students more readily established links with their own experiences and interests. Evaluation was done in much the same way as in classroom environments, participation in discussion groups being graded in lieu of attendance. Overall, students had difficulty evaluating their own learning, stating instead that they earned their grade simply by conforming to the course-work requirements. Students with poor performance tended to blame the "lack of clarity" of the course objectives.

Students admitted readily that they chose an on-line course because of the flexibility it afforded in their schedule. However, some found the prescribed pacing too slow, while others found it somewhat daunting, especially when assigned weekly readings – thus, scheduling became more of a problem than anticipated. The requirement to participate in online discussions was perceived as additional workload that would have been less demanding in classroom interaction. The detailed program structure found in all but one of the on-line courses was perceived to make the learning tasks more manageable, as they were relieved of any ambiguity.

This finding points to the importance of some important *algorithmic* features that are inextricably woven into the design of each D. E. learning environment. In most cases, we found that design features tended to reduce learner autonomy in very serious ways. Setting unalterable objectives, leaving all evaluation activities to the instructor, setting the same sequence of learning for all students independently of their individual needs or characteristics, all of the have detrimental effects on learner autonomy. In fact, this is the area where the most severe weaknesses were found in the designs we studied. Since there is no inherent reason why D. E. packages should limit autonomy so much, we need to ask ourselveswhy designers tend to appropriate for themselves such excessive control over the environment.

C. Value

Students can be placed in two groups according to the criteria they used to establish the value of their learning. The first group derived their estimation from the potential usefulness of their newly acquired knowledge in some immediate area of their lives, either by providing tools for better understanding world issues or financial matters, or by developing skills that apply to family relationships or the workplace. The second group was typically concerned with completing a university degree and selecting eligible courses for their anticipated convenience or easy workload. Interestingly, several students admitted opting for on-line courses assuming – wrongly they soon discovered – that they would entail a flexible schedule and a less demanding productivity. Some of the derived benefits were discovered as learning occurred throughout the courses. Others were identified as unanticipated spin-offs, such as developing better writing or computer skills. One student pointed out that his workplace offered a similar course package, featuring a better design and a lower cost, but that it could not be credited towards his university degree.

As in many institutions, the actual per-credit cost to the student is considerably higher for the on-line version of a course. Students generally accepted this fact with some resignation, but could not explain the disparity. One student realized too late that he could have learned independently, at

a much lower cost, everything that he learned in his course. Two other students acknowledged that they had chosen the institutional avenue in order to access the university's sophisticated computer labs. In light of these findings, we can say that the *economics* of D. E. are often poorly understood by institutional designers. In some instances, the cost associated with learning is in reality the hidden cost of giving institutional credit for learning that could have occurred anywhere. This gives rise to unnecessary duplication of courses that are offered in non-credit organizations (e.g. the workplace), or to the practice of granting dubious legitimacy for learning that otherwise could have been entirely self-directed, and therefore considerably less costly.

D. Context

Students were asked about the reasons they decided to enrol in their course, and why they chose the on-line version of the course. Factors such as desire to understand family issues, or the wish to improve work performance were mentioned by mature and non-degree students only. Reasons for choosing the on-line version of the course were mostly linked to personal, family and work situations. Somewhat ironically, the same factors were identified as barriers to achievement in the on-line course. The institutional context also was perceived to play a role, beginning with the fact that two versions of the course were offered by the institution, that the on-line version was higher priced but available, and that the absence of in-class meetings seem to motivate the instructors to increase student workload.

E. Media

In a previous study (Bouchard & Kalman, 1998), low computer literacy was identified as a barrier to distance learning. Here, students all had achieved reasonable competency at using computers. Some difficulties were encountered however with the consistency of access to the online environment. There were frequent system crashes and technical help was not always available. From the delivery point of view, the emphasis was placed on completing course assignments and little attention was paid to students' efforts to learn how to navigate within the system and outside. Some features of the courseware were used routinely, such as messaging and on-line exams, while others were rarely or not used (file transfer, self-corrected testing, live chat, transfer of images or animated *gif* files, etc).

This particular finding points to another area of concern for D. E. designers. There is a tendency to use technology and systems that are available, rather than those that are appropriate. This is attributable to a common management error that consists of making decisions based on past investments rather than future returns. As often happens with adult learners, they end up making their own decisions, and choosing for themselves *how* they will learn. This is the self-appropriation of the *semiotic* aspect of learning that is made possible when more than one technology is available.

V. CONCLUSION

The purpose of this study was to obtain from the learner's perspective some indication of the factors that encourage or deter from the development of self-direction in mediated learning environments. In light of the data collected, it is possible to make some recommendations that relate to the conative, algorithmic, semantic and economic dimensions of learner autonomy. Further analysis will allow us to produce a more detailed classification, but for now we will limit ourselves to a number of recommendations that are supported by our data. This information should be useful for planners who value, beyond the conformity to academic standards, the capacity for self-direction as a central goal of education.

On-line and other mediated learning environments offer much potential for supporting the development of self-directed learning skills, and can also be powerful deterrents. Realizing the potential – and reducing the deterrents – are possible if educational planners consider the importance of these two criteria when making instructional design decisions.

REFERENCES

[1] Bouchard, P. and Kalman, L. (1998). *Distance education and Learner Autonomy: Some theoretical implications*. Proceedings of the Canadian Association for the Study of Adult Education conference. University of Ottawa.

[2] Candy, Philip (1991). *Self-Direction for lifelong learning*. San Francisco: Jossey Bass.

[3] Garrison, R. (2000) Theoretical challenges for distance education in the 21st century: A shift from structural to transactional issues. *International Review of Research in Open And Distance Learning*, V (1).

[4] Holmberg, B. (1977). "Tutoring distance students". *Epistolodidactica*, 7, 4-15.

[5] Hrimech, M. and Bouchard, P. (1998). Spontaneous learning strategies in the natural setting: Learning to use computers. In Huey B. Long and Associates (eds): *Developing Paradigms for Self-Directed Learning*. Oklahoma Research Center, University of Oklahoma.

[6] Petkova Julka., Karova M., Automated system for examination tests., The Third National Conference with International Participation on e-Learning in Higher Education, 15-17 May 2009, Svishtov, Bulgaria.

[7] Keegan, Desmond (1983; 1996). *The foundations of distance education*. London : Croom Helm.

[8] Knowles, M.S. (1980). *The Modern Practice of Adult Education*. Chicago: Associated Press, Follet Publishing Company.

[9] Long, Huey B. (2002). Philosophical, Psychological and Practical Justifications for Studying Self-Directed Learning. In Huey B. Long & Associates (eds): *Self-Directed Learning: Application and Research.* Oklahoma Research Center, University of Oklahoma.

[10] Moore, M. G. et Kearsley, G. (2006). *Distance education : A systems view*. Boston: Wadsworth Publishing Company.

[11] Tough, Allen M. (1965). *The Teaching Tasks Performed by Adult Self-Teachers.* Doctoral Dissertation. Chicago University, Ill.

AUTHOR INDEX