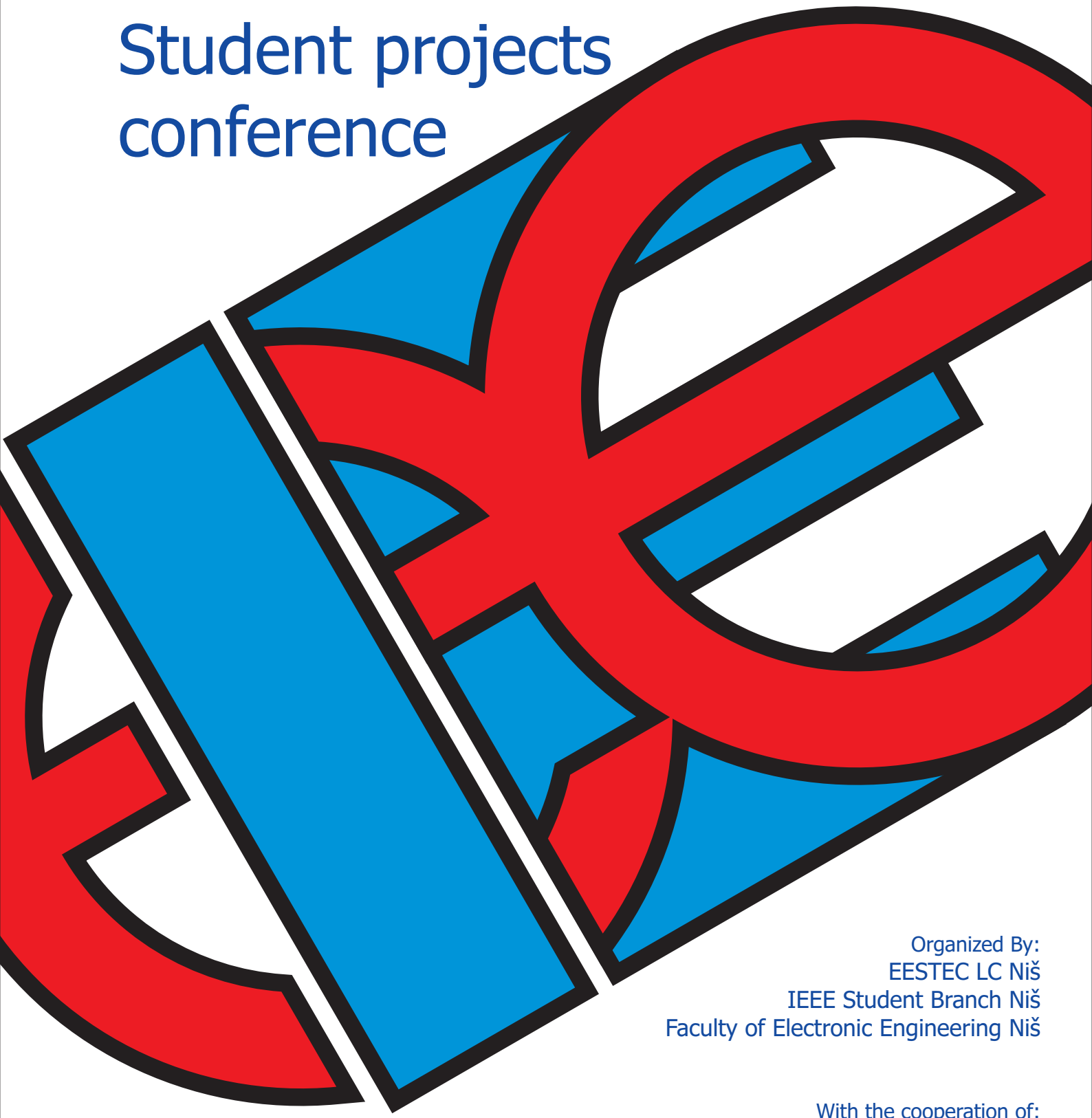
IEEESTEC

Student projects conference



Organized By:
EESTEC LC Niš
IEEE Student Branch Niš
Faculty of Electronic Engineering Niš

With the cooperation of:
IEEE Serbia and Montenegro section
IEEE Electron Devices/Solid-State Circuits Chapter
IEEE Microwave Theory and Techniques Chapter

ISBN: 978-86-85195-77-8

IEEESTEC

Student projects conference

Editors:

Prof. Dr. Stojadinović Ninoslav

Prof. Dr. Milovanović Bratislav

Prof. Dr. Marković Vera

Prof. Dr. Đorđević S. Goran

Organized By:
EESTEC LC Niš
IEEE Student Branch Niš
Faculty of Electronic Engineering Niš

With the cooperation of:
IEEE Serbia and Montenegro section
IEEE Electron Devices/Solid-State Circuits Chapter
IEEE Microwave Theory and Techniques Chapter

ISBN: 978-86-85195-77-8

Reviewers:

Andrejević-Stošić	Miona
Bjelopavlić	Darko
Dimitrijević	Aleksandar
Dimitrijević	Marko
Golubović	Snežana
Janković	Nebojša
Joković	Jugoslav
Mančić	Dragan
Manić	Ivica
Marinković	Zlatica
Milojković	Marko
Milosavljević	Aleksandar
Milošević	Nenad
Mitić	Darko
Nikolić	Goran
Pejović	Milić
Perić	Mirjana
Predić	Bratislav
Prijić	Zoran
Stanimirović	Aleksandar
Stojanović	Dragan
Tošič	Milorad
Veselić	Boban
Vladan	Mihajlović
Vračar	Ljubomir
Živaljević	Dragana

Secretary:

Danijel	Danković
Vučković	Dušan
Todorović	Darko
Simić	Vladan

Publisher:
Faculty of Electronic Engineering, Niš
P.O.Box 73, 1800 Niš
http://www.elfak.ni.ac.yu
ISBN: 978-86-85195-77-8
Circulation: 100 copies.

Publisher:

Faculty of Electronic Engineering, Niš
P.O.Box 73, 18000 Niš
http://www.elfak.ni.ac.yu

Editors:

Stojadinović Ninoslav
Milovanović Bratislav
Marković Vera
Đorđević S. Goran

Secretary:

Danijel Danković
Vučković Dušan
Todorović Darko
Simić Vladan

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

621.3(082)
004(082)
IEEESTEC Student Projects Conference (2008; Niš)
IEEESTEC Student Project Conference, [Niš, 2008] / Organized by EESTEC LC Niš [and] IEEE Student Branch Niš [and] Faculty of Electronic Engineering Niš; editors Stojadin. Ninoslav... [et a1,], - Niš : Faculty of Electronic Engineering, 2008 (Niš : N-Copy) – 100 str : ilustr. ; 25cm

Radovi na engl. i srp. Jeziku. – Tiraž 100. – Bibliografija uz svaki rad.

ISBN 978-86-85195-77-8
1. Institute of Electrical and Electronics Engineers. Student branch (Niš)
2. Electrical Engineering Students European Association (Delft), Local Comitee (Niš)
3. Elektronski fakultet (Niš)

a) Elektrotehnika - Zbornici
b) Računarstvo - Zbornici

COBISS. SR-ID 152702220

Štampa: "N-Copy", Niš
Tiraž: 100 primeraka

ISBN 978-86-85195-77-8

Table of contents

Thermovision Detection of Breast Cancer

Nikola Dojčinović

Abstract – This paper describes a technique for thermovision detection of breast cancer. Possibility of localization is discussed, with review on limitations and difficulties.

I. Thermovision Concept

Thermovision is way of non-conctact temperature measurement. Unlike other technologies of temperature measuring, final result of termovision is an image with distribution of temperature on scene or a numerical matrix [1]. Every pixel on such an image represents individual temperature detector. On base of temperature data, picture is formed suitable to human subjective temperature color impression (*red-hot, blue-cold, black-coldest*), or by level of illumination *(whit-hot, black-cold)*.

II. Thermovision in Medicine

Use of thermovision in medicine was considered even in 1950's. When possibility of computer image processing emerged in 1970's, enabling storing, precise quantization and statistic processing, application of thermovision in medicine diagnostics become reality. Because thermovision has a advantage of other ways of detection which wave a undesirable manifestation on organism, development of termovision application in medicine was very fast.

A. Temperature Changes

Application of thermovision in medicine is based on detection of temperature changes on skin of patient. If metabolic balance is disturbed, tissue temperature changes [3]. Intensive photosynthesis of proteins during inflammation will result with intensive metabolism. While metabolism is intensive, big amount of energy is being radiated, increasing the tissue temperature, and indirectly, temperature of skin. This is the mechanism for detecting metabolic inbalance.

B. Conditions of Imaging

If thermovision image is imaged as good presentation of distribution of temperature on human body, external factors must be eliminated [2]. Temperature characteristics of room in which the imaging is conducted, have influence on skin temperature, also as medicaments which result with high vascular flow of histological balance. Big influence on skin temperature is caused with cosmetics, consuming of alcohol and cigarettes. Finally, room itself must have enough proportions, in order for the fixed camera objective to give sharp image.

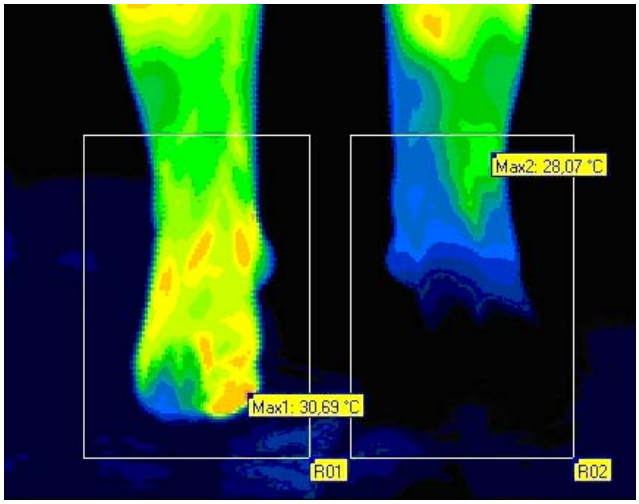


Figure 1: An example of thermovision image of feet with increased heating (left) and normal heating (right)

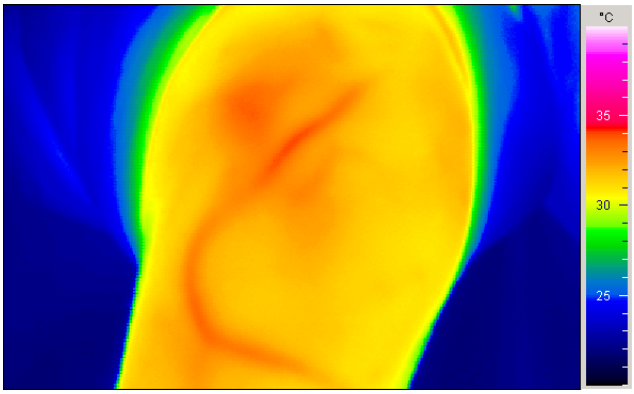


Figure 2: Increased vascular flow increases skin temperature

III. Breast Cancer

Most common breast cancer types start their development in the breast ducts (ductal carcinoma) or in milk glands (lobular carcinoma) [4]. The origin of cancer

N. Dojčinović is student at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia
E-mail: nikoladojcinovic@gmail.com

cells can be easily determined using simple microscopic inspection. Noninvasive breast cancer is a type of cancer where carcinoma cells remain on their original site not spreading themselves onto the breast tissue surrounding the ducts or the gland. The most common type of cancer from this group is in situ cancer of milk channels (Ductal Carcinoma in Situ-DCIS). With right therapy, this type of breast carcinoma can have high rate of successful treatments.

Invasive (infiltrating) breast cancer spreads the inflicted cells outside the basal membrane that lies underneath the ducts or the gland thus spreading into the surrounding tissue. Cancer cells can travel to other organs like lymph nodes.

Some of the most common cancers of this group would be:

- *Invasive Ductal Carcinoma IDC*, which covers almost 70 percent of all breast cancer cases. Cancer cells are being generated inside the duct, penetrating the duct wall and infecting the surrounding tissue. Cells can remain localized or spread (metastasize) throughout the body carried by lymph or blood.
- *Invasive lobular carcinoma ILC* is not as common as IDC. The principle of invading the healthy tissue is almost the same as aforementioned. The only difference is the origin of carcinoma cells. Patients with this type of cancer do not feel a hard lump during a standard self exam, making it harder to diagnose.

IV. Breast Cancer Detection

In contemporary breast cancer medicine the most common practice use mammography, a method using X-rays to detect the cancer. Although it is clinically proved fact that X-rays have a bad influence in the overall organism, mammography is so widely used due to its high detection rate and accuracy (with error rate less than 10^{-3}). Thermovision detection, on the other hand, does not provide as accurate results, but has a big advantage over mammography in being completely non-invasive and safe for the tissue.

In order to determine a breast cancer using this method thermograms of healthy and cancer patients must be significantly different (Figure 3). We cannot make a clear distinction between the two, but we can argue that, according to the clinical studies, there exist a certain correlation between the temperature increase and existence of a tumor [5]. At temperature differences of 3-4˚C we can say with high certainty that a carcinoma exists. Mammography scanning is not occluded, as it still presents itself as most certain diagnostic tool.

Accuracy of thermovision detection is not high. American Department of Public Health did a study in 1990 that claimed sensitivity of 61% and specificity of 74% to the thermovision method [6].

Another advantage of thermovision detection would be the possibility of early detection. Namely, in the earliest stages of breast cancer development a vascular tissue is being created which in turn increases breast outer temperature. This early symptoms cannot be declared as cancer, but give a hint to early detection of breast cancer before first pain sensations or visible deformations on the breast itself.

V. Possibilities of Cancer Localizing

As mentioned before, thermovision imagery gives the temperature density on surface of a breast. If we model this bio-thermic system we can isolate: the source of heat (tissue), insulator (healthy tissue surrounding the cancer) and projection screen (skin).

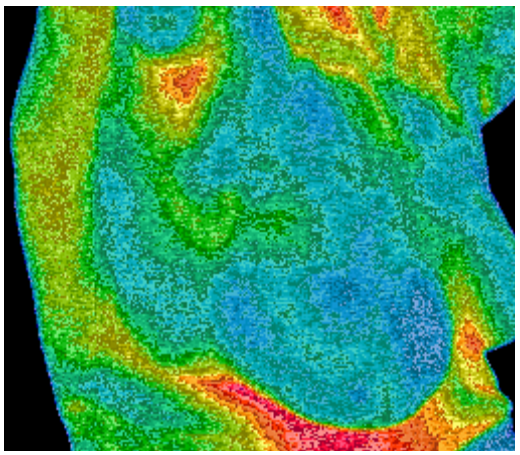


Figure 3a. Thermogram of a Healthy Breast

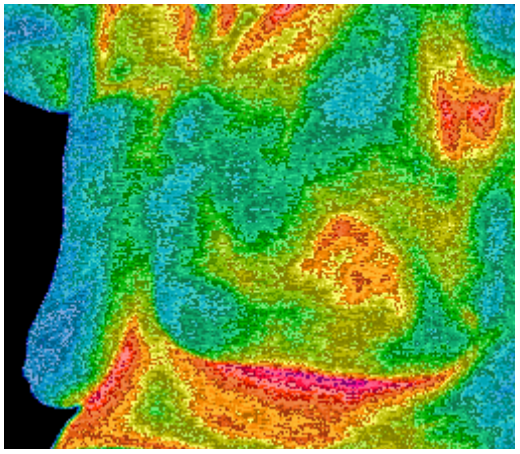


Figure 3b. Thermogram of an Infected Breast

This simplified system can be used in further theoretical modeling. Because of extreme complexity of breast tissues we must have special care in modeling this system in order to obtain any usable results.

A. Mathematical Apparatus

When decribing a system like this one, we can find a conection between its integral parts. Breast can be represented with a function of temperature attenuation f(x,y,z). If we mark the temperature of cancer with T_c disregarding its dimensions, and if we supose that temperature front is perfectly circular then skin temperature T(m,n) in (m,n) can be represented as:

$$T(m,n) = T_c - T_c S_{m,n} \quad (1)$$

where $S_{m,n}$ is attenuation. In a given direction we can calculate it as:

$$S_{m,n} = \int_d f(x,y,z)ds \quad (2)$$

by summing the attenuation function along the line d cancer point (m,n). Therefore we can write the following equation:

$$T(m,n) = T_c\left(1 - \int_d f(x,y,z)ds\right) \quad (3)$$

which connects temperature of cancer focal point with the skin temperature.

This way we reduce the problem to a single variable: line d - the distance of focal point and measurement point [7]. The temperature of cancer can be estimated by the type of cancer and degree of development.

B. Anatomic Limitations

Breast size can vary with patients and can span from 30g to more than 1kg [8]. Breast delimiters are well defined deep under the surface being demarcated by fascia pectoralis (a muscle). But, in microscopic scales, breast tissue can go deep into the muscle. This is very important, not only for the sake of our model, but for the fact that even complete mastectomy cannot fully remove all tissue.

The adult female breast consists of a series of ducts, ductules, and lobular acinar units embedded within a stroma that is composed of varying amounts of fibrous and adipose tissue. The stroma comprises the major portion of the nonlactating adult breast, and the relative proportions of fibrous tissue and adipose tissue vary with age and among individuals (Figure 4). These segments can be anatomically very bad defined and it can be hard to determine them even during the operation.

Bad segmentation makes removing a localized duct cancer possible only by performing mastectomy of the duct.

Every segment consists of a branch structure. Lobulas are draining the milk towards ducts and channels. This system opens up at the nipple. Normal lobula consists of different number of glands (between 18 and 25). The number of glands at lobula can vary significantly.

Complete anatomy of breast can vary as well. The variations can be so severe that jeopardize viability to maintain a consistent tissue model f(x,y,z).

On the other hand, if the tissue is homogeneous enough, we can perform a statistical analysis, assuming the attenuation same in all directions.

VI. CONCLUSION

Thermovision has increasing number application in medicine practice, from patient tests to diagnostics. Unfortunately, as shown in this paper, thermovision imagery does not give significant number of information to detect a breast cancer, but has its advantages.

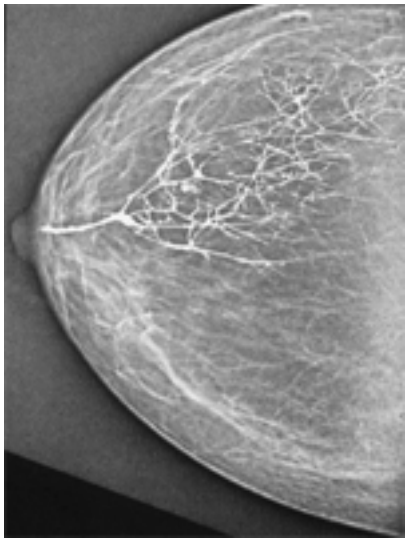


Figure 4: Ductogram of a Breast – ducts and glands

This paper also mentions the possibility of localizing the cancer based on the thermovision image. It is still too early to make any conclusions in this matter, but it opens up a very attractive question of modeling thermo conductivity of breast tissue for localization purposes.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge my indebtedness to Dr D. Mančić of Faculty of Electronic Engineering of Niš and Dr A. Petrović of Medical Faculty of Niš, for a most valuable discussion on the subject of this note.

REFERENCES

[1] MS Tomić, “Termovizija-princip rada”, Infinite, vol.2, (in Serbian),pp.34-36, sep 2005

[2] EFJ Ring and K. Amer, "The Tehnique of Infra red Imaging in Medicine", *Thermology international, vol 10/1*, pp. 7-14, sep 2005
[3] EE Stern, B. Zee, J. Sen Gupta, FW Sanders, "Thermography: Its relation to pathologic characteristics, vascularity, proliferative rate and survival of patents with invasive ductal carcinoma of the breast", *Cancer* vol77. ED-29, pp. 1324-1328, 1996.
[4] www.mayoclinic.com
[5] RN Lawson and MS Chughtai: "Breast cancer and body temperatures", *Can Med Assoc vol 88:* pp. 68-70, 1963
[6] Group of authors, "Thermography in screening for breast cancer", *Journal of Epidemiology and Community Healt vol 44,* pp. 112-113, 1990
[7] RA Crowter, DJ DeRosier and A. Klug, "The reconstruction of a tree-dimensional structure form projections and its application to electron microscopy", *Proc. Roy. Sot. London, vol ASSP-29,* pp. 237-244, 1981
[8] SE Mills: "Histology for Pathologists", *Lippincott Williams&Wilkins*, 2007

Migration from Public Switched Telephone Network to Next generation network

Nenad Denić

Abstract– Most of the telecom service providers are in the race of migrating from public switched telephone network-PSTN to Next Generation Network -NGN to launch innovative services in order to attract more and retain existing customers. Companies and government institutions (railroad companies, armed force, police) have their own networks. They have to offer new generation service to their employers to improve company efficiency. This paper describes problems with current networks, benefits of NGN different phases for migrating from PSTN to NGN.

I. INTRODUCTION

Human beings have a strong desire for better communications capabilities without the need to deal with complex technology, hence communication services should be user-centric driven [1]. In nowdays majority of service providers around the world are using PSTN for providing telecommunication services. PSTN, Internet and Cellular networks are different domains, and they have their own protocol and services. Success of Internet has lead to the deployment of huge packet-switched networks worldwide to carry voice, data and video resulting in Operational Expenditure-OPEX and Capital Expenditure-CAPEX savings. NGN is the future network for all types of communication services. The necessity of migration from PSTN to NGN is for:

- Network convergence – single network for voice, data and video
- OPEX and CAPEX savings
- New service opportunities

Migration from PSTN to NGN should be based on maximum possible reuse of existing equipment and replacement of components which are near the end-of-life.
Migration from PSTN to NGN involves:

- Replacement of TDM network elements in a phased manner
- Maximum reuse of existing resources
- Use of open and mature standards
- Convergence of access and backbone network
- Continuation of existing network capabilities and services with same or comparable QoS and security
- Interworking between different types of networks
- Addition of new services

The ideal architecture for the transition is one that provides flexibility to service providers by enabling them to deploy new services, while leveraging existing services and infrastructure as long as it makes sense.
NGN will be base for creating and carry on new multimedia applications. These applications will use all good characteristic of broadband networks and always on capability. NGN with network architecture use advantages of new technology to offer new sophisticate services.

II. NEXT GENERATION NETWORKS

NGN offers the convergence of fixed and mobile telecoms services and data networks-all based on the same architecture as the Internet.
NGN is more than just switch replacement, it is an answer to the market reality that 'fixed' voice telephony-although a major part of the telecommunications business-is no longer the only feature that the user wants from their phone service provider. Triple play services (Voice, Internet and TV) are available via Cable and xDSL already. The NGN brings mobility in to the picture and the opportunity for further bundling of high revenue services for customers. [2]
NGN support variety user services such as:

- Person-to-Person (voice, video telephony) and Person-to-Machine (gaming, distance learning) services;
- fully integrated real-time and non real-time multimedia communications (live streaming and chat);
- interaction of different services and applications (services that use location and/or presence information);
- easy user set-up of multiple services in a single session or multiple simultaneous sessions (flexibility in adapting a service: adding media components, adding parties, etc.).

III. LIMITATIONS OF TODAY'S NETWORKS

The package of services that can be offered to any single subscriber are limited in today's network. To trigger services in the network, the subscriber activates them through the access network of the service provider. When the subscriber activates a call, a single trigger is generated that can be used to initiate applications in the network which likely reside on an Service Control point-SCP. With only a single trigger, the service provider can only offer the services available on the "triggered" SCP.
Another challenge to easily selecting among application vendors is the variety of flavors of SCP access technologies including Bearer Independent Call Control-BICC, Intelligent Network Application Part-INAP and

Nenad Denić is with Serbian Armed Forces,
denic_v_n@neobee.net

Customized Applications for Mobile Enhanced Logic-CAMEL variations. Access networks are only capable of interacting with a single interface without protocol conversion. Each application essentially becomes an isolated service delivery solution within the network. This delivery approach, called a "siloed" architecture, prevents the service provider from easily maximizing services for their subscriber populations. However, one of the most significant limitations in the Intelligent Network-IN application layer is that the ideal group of services for a subscriber group are generally not on the same platform and cannot be offered to subscribers as part of a portfolio. Each time a new technology or service is introduced at the access level, existing applications must be modified and reconnected, creating operational and maintenance issues as well as bottlenecking the introduction of new services.

IV. Migration of PSTN

Today most of the voice traffic is transported over PSTN and controlled by local exchange-LE and transit or trunk exchange-TE circuit switches. The voice related signalling network is handled by the SS7 signalling network. Value Added Services are provided either by the switches or through the external IN. The Internet connectivity is provided through narrowband (PSTN or ISDN) dial-up services as well as broadband ADSL [3].
NGN migration strategy may be same for all carriers. Strategy have to obey several rules: to decrease expense of network infrastructure and maintenance, optimal usage resource of new technology, maximize usage of installed equipment, development of sophisticated services, better quality of existed services.
Different phases for migration of PSTN to NGN are given below. However, the sequence of implementation depends on the business and strategic needs. Different phases can be combined for implementation.

A. Phase I: Migration of Trunk Exchanges

First step towards NGN migration, operators can focus on scenarios to offload long-distance voice from their TDM network. Existing TE may be replaced by deploying Softswitch, Trunk Media Gateways-TMG and Signalling Gateways-SG. Some of the trunks from existing Local Exchanges-LEs may be connected to these Trunk Media Gateways and remaining may continue with the existing TDM trunk exchange. This approach ensures full protection of TDM investments, while providing the operator with a full fledged connection using packet network. Intelligent Network based Value Added Services may be accessed through the existing TDM switches. This is depicted in Fig.1.

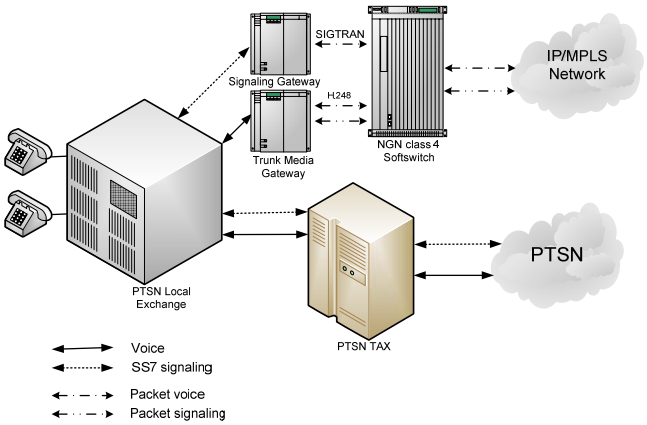


Fig. 1. Migration to NGN Class 4 Applications

In this scenario, softswitch will perform all the transit functions of analysis and routing. The softswitch will communicate with the TMG using the H.248 protocol. TMG interfaces the PSTN for the media flow from IP-PSTN to IP or vice-versa. It provides functions such as media conversion (circuit to packet, packet to circuit), echo control etc.
The signalling functions are migrated to NGN using SG functionality. SG transports the SS7 signalling over IP infrastructure using IETF SIGTRAN protocol.

B. Phase II: Migration of Local Exchanges

In this phase LEs are replaced by the Softswitch and Access Gateways-AGW with same services. This is depicted in Figure 2.

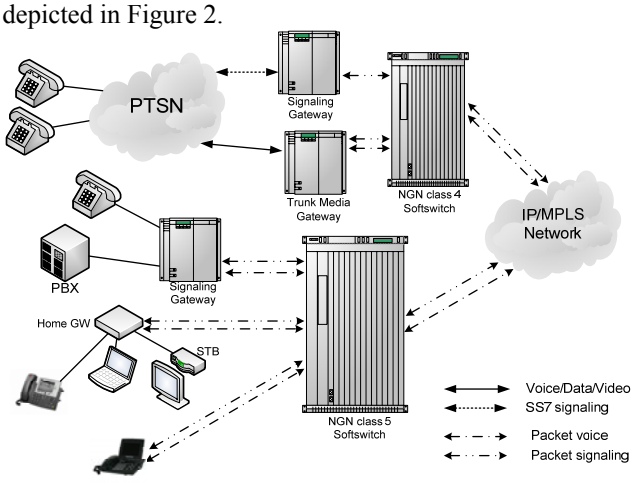


Fig. 2. Migration to NGN Class 5 Applications

Softswitch with local features will be used as a common control element for class 5 applications. Access Gateways-AG provide various types of access to the subscribers (PSTN, ISDN, V5.2, xDSL etc.) and connects them to IP core network. AGs may be configured for various class 5 applications depending on end user topology, density, service requirements, etc.

Depending upon the size of the network, a single softswitch with class 4 and class 5 applications may be planned.

V. Migration of Services and Security Aspect

While migrating from PSTN to NGN, all PSTN services with same equipment, same look and feel should be provided. Two PSTN networks connected via NGN transit network should be able to provide transparency to all bearer services.

The existing IN services are provided though SCP. The softswitch interacts with SCP through signalling Gateways, using Intelligent Network Application Protocol -INAP. In future, new IN and value-added services may be implemented using Application Servers which will be accessed by softswitch via Session Initiation Protocol.

NGN will enable the introduction of new range of data and multimedia services. These new services will require new type of terminals for IP based multimedia services. These terminals will communicate with the softswitch through SIP. Fiber to the Curb or Fiber to The Home technologies may be used for higher transmission speeds.

Once the evolution to the NGN is completed, all the IN and value-added services including number portability will be provided by Application Server (AS).

The security level in NGN should be at least same as existing in PSTN. As present PSTN/ISDN networks are migrating to NGN, new security concerns may be encountered. Therefore additional measures may be required to guarantee at least the current security level. Different security dimensions, depending on the access method, should be taken into account. Some of these are: Authentication, Non-repudiation, Data confidentiality, Communication security, Data integrity, Availability, Privacy.

VI. Conclusion

Migration from PSTN to NGN is an absolute prerequisite for all telecom service providers and for owners of corporate networks. NGN networks offers large scale deployment of innovative applications, with easy-to-use service portals.

Migration of PSTN to NGN is owner-specific and should be completed in a phased manner. Owners must begin to focus on the steps that will enable a smooth transition. Replacement have to be based on maximum reuse existing resources. Simply replacing the existing network will not generate additional revenue until the new services, which can be provided by NGN are provided to the subscribers.

REFERENCE

[1] Network Migration Strategies towards IMS, Strategiy white paper, http://www1.alcatel-lucent.com/

[2] http://www.etsi.org/website/Technologies

[3] Migration of PSTN to NGN, www.tec.gov.in

Old games on the old way – LED Formula

Aleksandar Gošić, Aleksandar Lakićević, Miljan Obradović

Abstract- Today is playing mostly reduced to sitting in front of monitor and control the game using the keyboard and mouse. Games are so complex, and the characters in them, made almost perfect to create the impression that we are in the virtual world, so the player does not have to create the image of the games and uses imagination. A game witch is presented in this paper has not been made as all today, modern game, but in the way how it was make in the last century and tricks that are then used when there was no powerful computer on which developers are creating games, but to expression came to designing appropriate modules that are controlled with primitive software, and then some whimsy solutions, that are applied in this game, are made-up.

I. INTRODUCTION

As we pointed out the modern games are made mainly for the high-resolution panel/monitors, who so faithfully represent every detail of the game where gains the impression that we are in a virtual world, while the use of imagination omitted. It loses the effect of the development of imagination and intelligence, for which they were primarily rationalize. Game, which will be described below is a modern and opposition made by the principle of the original games.

II. HARDWARE DESCRIPTION

In the next chapter will describe all the game hardware. Block diagram of game is shown on the next figure.

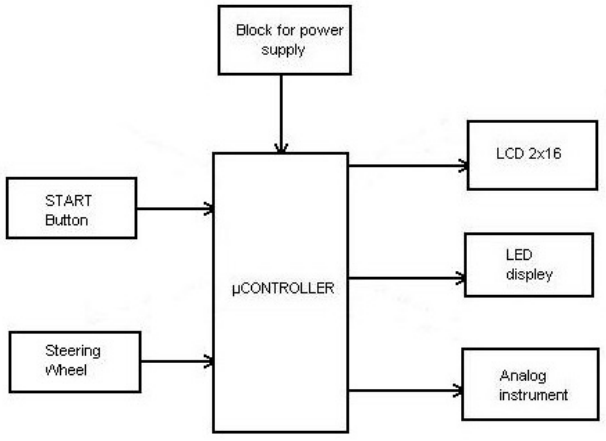


Fig. 1 Hardware block diagram

A. Gošić, A. Lakićević, M. Obradović are students at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia

E-mail: gosa46@yahoo.com, lakisa85@yahoo.com, miljan.obradovic@gmail.com

As you can see on the figure 1 is shown implemented system that includes the following blocks:

- Block that provides a stable power supply voltage for the system
- Start button for start the game
- Steering wheel/joystick for car control
- Analog instrument for the indication of gas and the remaining time to play
- LCD 2*16 display on which will be written the ranking
- LED display panel, which will be described in detail on
- PIC16F877A(1) microcontroller that the heart of the system and manage complete games

Panel, which is shown in this game is made from plywood, in which are put 64 LED (*light emitting diode*), through and are place PVC film in which are printed all car icon position. LED`s are located to form a matrix of 6 rows and 5 columns, that also represents the path by which cars are moving.

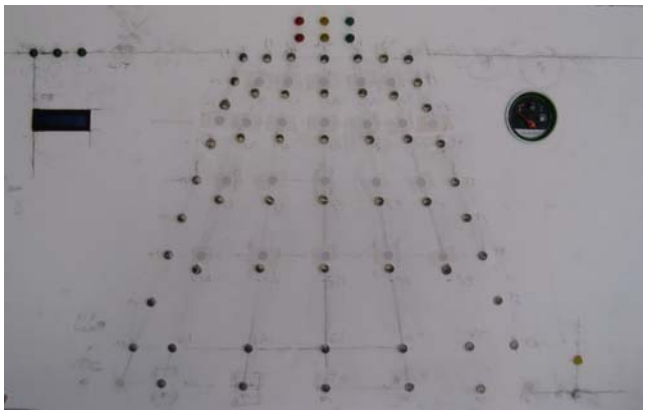


Fig. 1.Panel with 30 LED's

The entire game is based on the alternate light of LED with such timing that create impression movement car icon by track. Impression movement also endured 18 LED which are represent balusters and is located next to track. One line of 6-LED, which is on the lower part of the track, is the car icon that managed by the user/player. In the upper right corner of a panel is analog device that measures gas (time) and when the needle come to the red field the yellow LED turns on, which means that the user can enter the PIT STOP, and reload fuel, it gains a condition for the continuation of games, otherwise the lost lives. In the upper

left corner of the panels are three LED that denote their lives, which are lost when the car icon that ranges from above down overlap with players car icon, or due to a lack of fuel.

Also in the top left corner of panel is a LCD 2*16 character that writes some text, which violates little bit style games, but was necessary. At the top of panel is a scoreboard, which identifying start of the race. Management car icon is being potentiometers in which the connect to steering wheel/joystick. To play sound effects used another "small" PIC12F675(3) microcontroller, which communicates with the master microcontroller and depending of that generated the volume by the *piezo* loudspeaker. To all this worked was necessary to use 50m over the wires connecting, which corresponds to a length of olympic swimming pool.

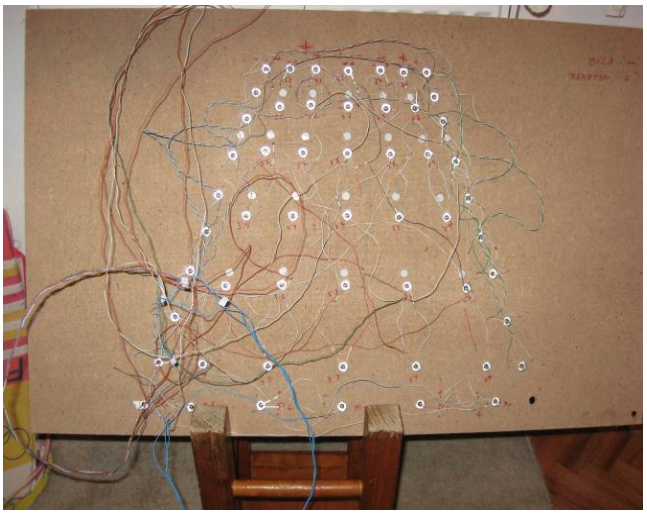


Fig. 2.Rear side of LED panel

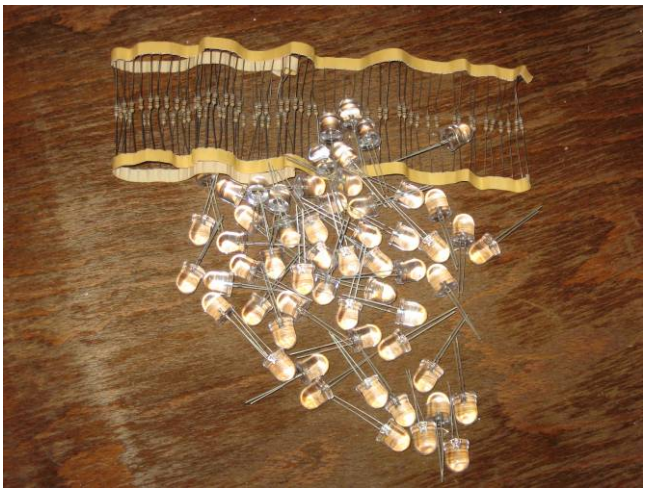


Fig. 3.LED and resistors that are put on the panel

To control LED panel is used one of the tips and tricks that is shown in the following figure where is given example of controlling 6 LED's by 3 PIN.

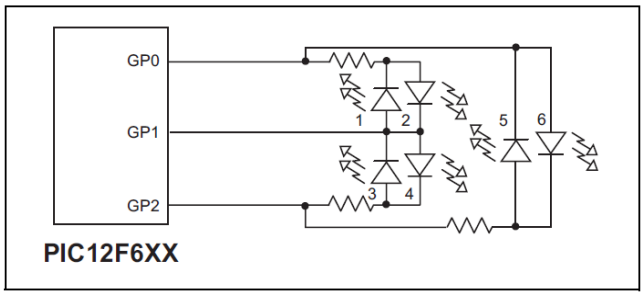


Fig. 4.Tips and tricks (2) to connect LED's

By the same principle it is possible to control many more LED's, which calculates by the following formula(2):

$$D=GP * (GP - 1) \quad (1)$$

, where D means number of possible LED's that can be controlled and GP means number of available pins.

The following figure gives an example how to set pins on the microcontroller to activate all cases LED diode.

GPx			LEDs					
0	1	2	1	2	3	4	5	6
0	0	0	0	0	0	0	0	0
0	1	Z	1	0	0	0	0	0
1	0	Z	0	1	0	0	0	0
Z	0	1	0	0	1	0	0	0
Z	1	0	0	0	0	1	0	0
0	Z	1	0	0	0	0	1	0
1	Z	0	0	0	0	0	0	1
0	0	1	0	0	1	0	1	0
0	1	0	1	0	0	1	0	0
0	1	1	1	0	0	0	1	0
1	0	0	0	1	0	0	0	1
1	0	1	0	1	1	0	0	0
1	1	0	0	0	0	1	0	1
1	1	1	0	0	0	0	0	0

Fig.5. Combinations for LED's controlling (2)

In this way, 45 LED's is controlled by 14 pins, while the other 19 LED is controlled on classical way.

Next figure shows the connection electrical scheme of car position.

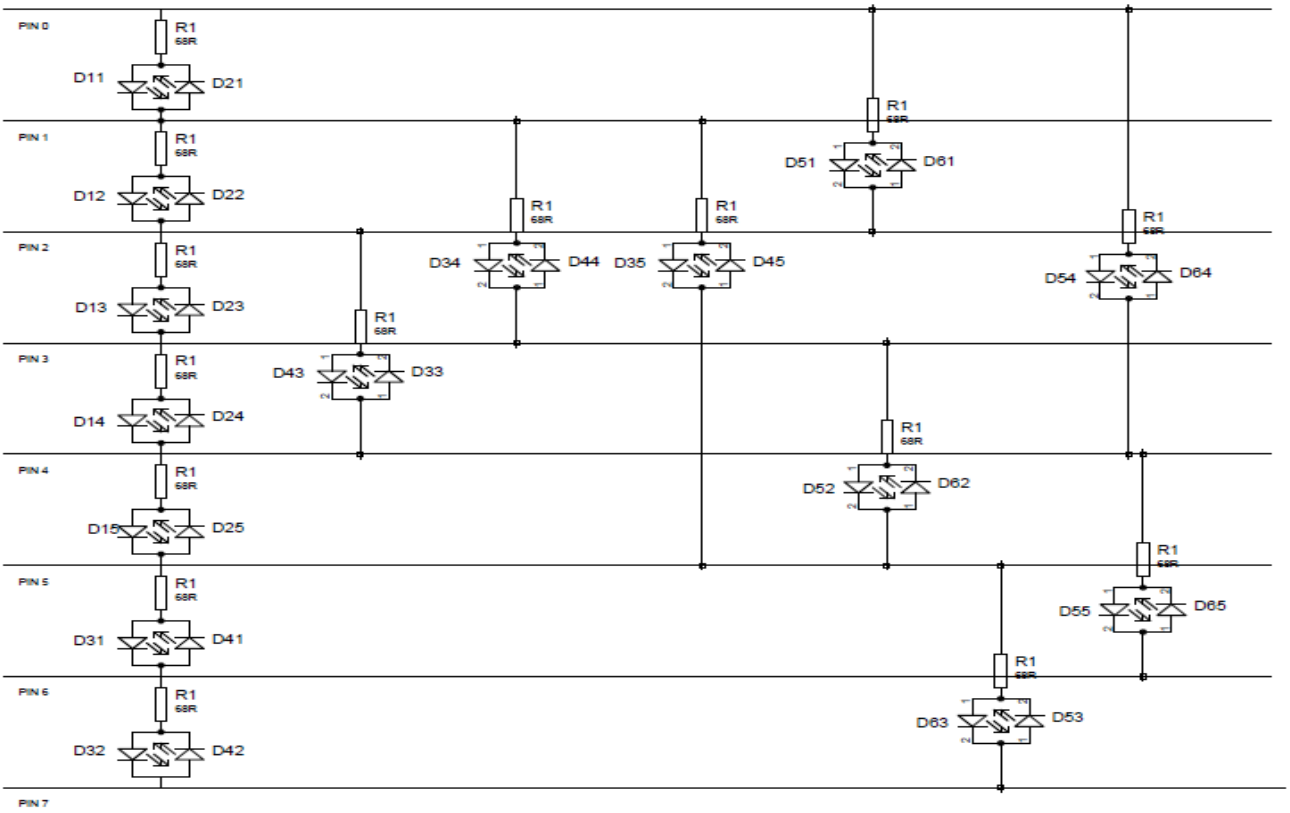


Fig.6. Electrical scheme of car position.

III. SOFTWARE DESCRIPTION

Software that is implemented in the microcontroller is developed in the software package *Proton Development Suite*(4). Global algorithm of the software is given on the figure 7. When power is connected, it is necessary to carry out initializing of registers, which is necessary for the microcontroller to do all tasks. Then come the "idle process" work where random updates LED display and waiting to start games. Then follows the testing improper position, the remaining number of life and quantity of gas for determining the next state for the continuation of game. Next the algorithm examines whether there is change of the situation steering wheel/joystick and performs the refresh of 2*16 LCD, LED display and analog instrument. This software is utmost part of algorithm because it creates a way of the games. Then the waiting timer interrupt algorithm returns to the beginning of loop. In this way has been achieved the desired effect of movement and control of the car it has been made the primary goal of projected system.

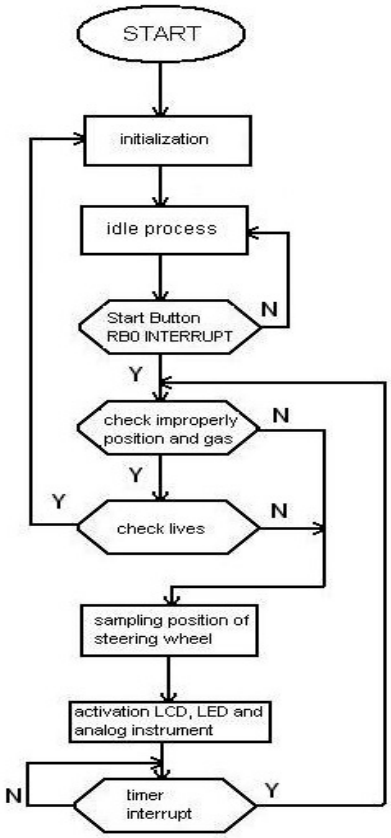


Fig.7 Algorithm for software

IV. Conclusion

The aim of this project was not only to realize one of the tricks on the visually interesting and winsome way, but to show that is not all in the FPS (frame per second), and how a 'simple' games can be a very interesting and to provide enjoyment the same as many modern games.

References

[1] PIC16F87XA Data Sheet, http://ww1.microchip.com/downloads/en/DeviceDoc/39582b.pdf

[2] Tips ‘n Tricks , Microchip, datasheet, http://ww1.microchip.com/downloads/en/DeviceDoc/40040b.pdf

[3] PIC12F675 Data Sheet http://ww1.microchip.com/downloads/en/DeviceDoc/41190E.pdf

[4]Proton Development Suite http://www.crownhill.co.uk/product.php?prod=1334

Digital video broadcasting technology and implementing DVB standards in DVB-T

N. Nikić, S. Jakovljević, M. Stanojević

Abstract - This presents overview of digital video broadcasting – terrestrial which was developed by the DVB project. The overview starts with an explanation of the wide field of application which the members of the DVB project decided to address over the year. Than it discusses the base band processing required for DVB services and looks into the specifications provided for the broadcasting over terrestrial transmitters. Further in the paper we focused more on technology used for accomplishing DVB – T. We explained channel coding and modulation used in order to analyze how an MPEG transport stream at the input of DVB – T modulator is turned into a DVB – T signal. In channel coding and modulation we explain three important parts: 1. energy dispersal and synchronization, 2. error protection and modulation, and 3. coded orthogonal frequency division multiplexing (COFDM) Then we look at the three different criteria that are used to evaluate the performance of DVB – T: the available useful data rate, carrier-to-noise (C/N) ratio required for quasi-error-free reception and the field strength required for different reception modes. In next section we deal with mobile reception that consider network planning issues antenna diversity concepts for mobile receivers, and hand over procedures. Finally, it is described that DVB – T signals can be received with mobile and portable receivers. The paper closes with a description of the newest areas of development in DVB – T.

I. INTRODUCTION IN DVB

DVB is abbreviation stands for digital video broadcasting, which is broadcast related with technologies developed by the International DVB Project. In phase 1 of its existence DVB concerted on the development of technical specifications relevant for the more traditional broadcasting of audio and video services over satellite, in cable networks and via terrestrial transmitters, in later phases DVB addressed areas which lie outside of the classical broadcast world.

Of the many fascinating areas of technology for which DVB has developed solutions the following will be described: DVB-T which is solution for terrestrial broadcasting; DVB-H, a system delivering all. sorts of content to battery-powered devices; DVB-S2 which is the next generation satellite system; DVB-IP, a solution for the carriage of broadcast content over broadband IP network.

N. Nikić, S. Jakovljević and M. Stanojević are students at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia

E-mail:maki.85@gmail.com, sloba.jakovljevic@gmail.com

II. FIELDS OF APPLICATION OF DVB TECHNOLOGIES

One of the most important rules of DVB is that the technical work is to be commercially driven by the requirements of the member organizations. In consequence, the fields of application of DVB technology developed significantly over time. So did the goals which DVB tried to achieve. One of the goal will be to enable the transmission of very high-quality HDTV images over the digital television, possibly even via terrestrial broadcasting networks. Also DVB might enable the broadcasting standard-definition television (SDTV) quality using narrow-band channels, and it will enable an increase of the number of programs offered within existing transmission channel allocations.

Expectations of DVB will be to enable a multiplication of the number of television programs which can be broadcast in one transmission channel - irrespective of whether the transmission will be over satellite, on cable networks or via terrestrial transmitters. Likewise DVB-T will offer the possibility to address receivers in all kinds of environments from the classical TV sets in the living room via portable devices in ones shirt pocket to TV receivers built into vehicles.

III. BASEBAND PROCESSING

One of the fundamental decisions which was taken during the early days of DVB was the selection of MPEG-1/2 for the source coding of audio and video and for the creation of program elementary streams, transport streams, etc. The complexity of what has been achieved is nicely reflected by the fact that integrated receiver decoders (IRDs) are classified in five dimensions. "25 Hz" or "30 Hz." depending on the nominal video frame rates; “SDTV" or "HDTV," depending on the conventional TV resolution; "With digital interface" or "baseline," depending on whether or not they are intended for use with a digital bit stream storage device such as a digital VCR; MPEG-2 video or H.264/AVC video coding formats; and audio coding formats.

IV. TRANSMISSION ON CABLE, SATELLITE, AND TERRESTRIALLY

Satellite broadcast transmission media may use DVB-S and DVB-S2. For use in cable networks was developed DVB-C, and in terrestrial transmission system we use DVB-T. These cover the full list of transmission media.

In addition, solutions have been provided for the distribution of DVB content in (satellite) master antenna TV [(S) MATV] installations, via multipoint video distribution systems (MVDS) and on microwave multipoint distribution systems (MMDS) operating in a number of frequency ranges up to 60 GHz. The encoder processing in DVB-C, DVB-S, and DVB-T is based on the same fundamental concept. The three systems differ in the form of modulation used.

V. DATA BROADCASTING

The transport layer used by the DVB broadcast systems is MPEG-2 transport stream to which adaptations have been developed which enable the transmission not only of audio and video but also of any other digitally coded information. The data container can be seen as a visual representation of the multiplex. Depending on the usable data rate of the broadcast channel, the size of the container varies. The blocks identified as program specific information (PSI) and service information (SI) are important for the housekeeping inside the data container.

Irrespective of the kind of service the digital data represents, it can be inserted into the MPEG-2 transport stream (TS). But depending on the needs of the application (for example synchronization with other services), additional provisions have to be made in order to ensure a correct transmission and to guarantee that the receiver understands the transmitted service. In consequence, four ways of transmission were defined by DVB: data piping, data streaming, the use of a data or object carousel, and multiprotocol encapsulation.

VI. INTERACTION CHANNELS

Interactive services may require varying levels of interaction between the user and the service provider or the network operator. The requirement of providing an interaction channel across the transmission network was established by the desire to enable the user to respond in some way to the interactive service and by the necessity of the service provider or network operator to listen and possibly react to that response. The user's response may take the form of some simple commands, like voting in a game show or for purchasing goods advertised in a shopping program. On the other hand, interactive services are conceivable which require that the user is able to have full Internet access at the receiver. Fig. 1 shows the generic system reference model which DVB uses for the definition of technologies for interactive services.

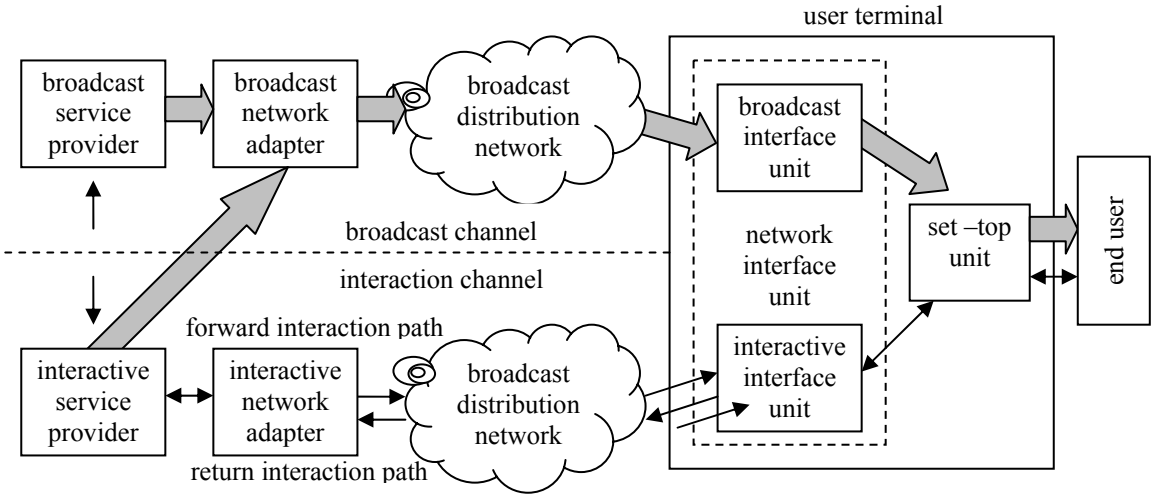


Figure 1

Adding interactivity to the DVB infrastructure requires the system to be extended by components providing communication means between the end-user and provider of the interactive service. The interactive service provider can be related to the broadcast service provider or even be the same organization. In any case, it can make use of the high bit-rate DVB broadcast channels in delivering information to the user of the interactive service at typical rates of up to 20 Mb/s per channel in terrestrial broadcast networks, and up to 38 Mb/s per channel in broadcast networks via satellite or cable.

VII. DVB-T

1. Introduction - The main objective was to support the stationary reception of terrestrial signals by means of

rooftop antennas. The support of country wide single-frequency networks (SFNs) was made a requirement. This type of network consists of neighboring transmitters which are in synchronism with each other, operate on the same frequency and transmit identical data streams. Trials demonstrated that the flexibility of DVB-T system supports networks covering not only stationary receivers indoor and mobile receivers in cars and busses.

2. Transport Stream (TS) Processing – The base band signal that is transmitted is a MPEG-2 TS. The TS is a continuous sequence of TS packets. Each packet has a length of 188 B. The first 4B contain the header of the TS packet; the following 184 B are used for the payload. The most important components of the header are the synchronization (sync) byte and the packet ID (PID).

3. Channel coding and modulation – Before the base band signal can be transmitted, it has to undergo channel coding and modulation. A forward error correction (FEC) is required which enables the receiver to correct errors that have occurred as a result of noise and other disturbances in the transmission path.

A. Energy Dispersal and Synchronization – The data at the base band interface is combined with the bit stream of a pseudorandom noise generator which is implemented by a feedback shift registar. Only the sync byte of the TS packets is left untouched in order to retain a means for synchronization.

B. Error Protection and Modulation – The outer error protection is implemented with a byte-oriented block code. The block code that is used is Reed-Solomon (255 239) code. That means that 16 correction bytes are appended to the 239 information bytes. Since the TS packet has a length of only 188 B, the first 51 B are set to zero, and are not transmitted. In this way a Reed-Solomon (204 188) code has been created. The outer interleaver that follows does not provide any additional error correction of long burst errors. For DVB-S and DVB-T, an inner error protection follows which is optimized for the correction of bit errors. A convolution encoder with basic code rate ½ is used for this purpose. The next element is inner interleavers, which purpose is to cope with the effect of frequency-selective channels which may for example result from echoes on the transmission path. In the symbol mapping, each of the individual useful carriers of the OFDM signal is separately modulated. A choice can be made between the modulation techniques QPSK, 16-QAM and 64-QAM. The DVB-T system offers the possibility of hierarchical modulation.

C. Coded Orthogonal Frequency Division Multiplexing (COFDM) is characterized by the existence of symbols which consist of a large number of carriers. The start of every COFDM symbol is preceded by a so-called guard interval. The purpose is to enhance immunity to echoes and reflections. In the DVB-T system the OFDM symbols are combined to a transmission frame. One OFDM frame contains a TPS block of 68 bits, namely, 1 initialization bit, 16 synchronization bits, 37 information bits and 14 redundancy bits for error protection. Finally we can insert the mega-frame initialization packets (MIPs) in a DVB-T TS.

4. Performance – The different criteria shall be used to evaluate the performance of DVB-T: the available useful data rate, carrier-to-noise (C/N) ratio required for quasi-error-free reception and the filed strength required for different reception modes. A dominant direct signal path is present. A transmission channel with echoes of more or less equal significance and without any direct signal path is called Rayleigh channel.

5. MOBILE RECEPTION - Compared to the traditional stationary reception for which a rooftop antenna at 10m height is assumed, mobile reception is much more difficult to realize. A loss in antenna gain of approx. 10 dB and a loss in field strength of more than 10 dB resulting from the low antenna height of about 1.5m has to be taken into account.

To achieve successful mobile DVB-T reception, a number of factors need to be considered. The receiver needs to track channel variations in time and frequency. In addition, correct channel estimation needs to be provided.

A. Network Design – For portable indoor, portable outdoor, and mobile reception, SFNs are recommended.

B. Diversity Technique for Mobile Receivers -In order to provide optimum mobile reception in an existing DVB-T network antenna diversity technique to combat the effects of signal fading can be used.

i. Selection combining (SC) - For selection combining, the subcarriers with the highest signal-to-noise ratio will be selected from the three sets of subcarriers available from the three branches. Selection combining can alternatively be executed on the bit stream leaving the symbol deinterleaver.

ii. Maximum ratio combining (MRC) – The use MRC antenna diversity lowers the required C/N ratio at low Doppler frequencies by about 7 dB.

C. Handover Techniques for Mobile Receivers - The receiver has to perform an automatic frequency change if the field strength of the actual signal decreases at the border of the coverage area. Furthermore, the interruption during this frequency change - the handover-should be as short as possible.

Before a handover can be executed the mobile receiver needs to determine its actual location as accurately as possible and it needs prepare lists including the alternative frequencies for at least the actual service.

VIII. PROSPECTS OF THE FUTURE DEVELOPMENTS IN DVB

The vision for DVB, over the next phase of its development, is to be an enabling forum for precompetitive standards setting in the "connected world" of networked digital media and applications. Broadcasting is about the creation, management, storage, delivery, and consumption of valuable content. Initially, DVB concentrated on

broadcast delivery, then on the development of interactivity-capability in receivers (MHP). The focus is now moving to the content itself - content protection and copy management, portable content formats and TV anytime/anywhere. This should include the contribution side, including professional services, as well as the consumer-focused distribution side. If valuable content is to be made readily accessible over a range of networks, it will need to be provided in a descriptive and protective package that can interact with diverse discovery, payment, and delivery systems.

IX. CONCLUSION

Over more than 14 years, DVB developed a significant number of technical solutions to commercial, scientific, and engineering problems. Over time the list of requirements developed significantly and therefore the activities of DVB reflect the changes in the industry since the early 1990s. The first DVB solutions can be considered fundamental enablers of digital broadcasting. The work ongoing in 2005 is much more devoted to finding solutions for ubiquitous access to content wherever it may be and on whatever network it may be available.

The success of DVB is the success of literally hundreds of companies and organizations and of hundreds of people working in the commercial, legal, technical, and PR departments of these companies.

REFERENCES

[1] U. Reimers, *DVB—T Family of International Standards for Digital Video Broadcasting.* New York: Springer-Verlag, 2004.

[2] Y. Wu, S. Hirakawa, U. Reimers, and J. Whitaker, "Overview of digital television development worldwide," *Proc. IEEE,* vol. 94, no. l, pp. 8-21, Jan. 2006.

[3] U. Ladebusch and C. A. Liss, "Terrestrial DVB—A broadcast technology for stationary portable and mobile use," *Proc. IEEE,* vol. 94, no. 1, pp. 183-193, Jan. 2006.

[4] G. Faria, J. A. Henriksson, E. Stare, and R Talmola, "DVB-H: Digital broadcast services to handheld devices," *Proc. IEEE,* vol. 94, no. 1, pp. 194-209, Jan. 2006.

[5] A. Morello, "DVB-S2: The second generation standard for satellite broad-band services," *Proc. IEEE,* vol. 94, no. 1, pp. 210-227, Jan. 2006.

[6] A. J. Stienstra, "Technologies for DVB services on the Internet," *Proc. IEEE,* vol. 94, no. 1, pp. 228-236, Jan. 2006.

[7] J. Piesing, "The DVB multimedia home platform (MHP) and related specifications," *Proc. IEEE,* vol. 94, no. 1, pp. 237-247, Jan. 2006.

[8] *Implementation guidelines for the use of video and audio coding in broadcasting applications based on the MPEG-2 transport stream,* ETSI Technical Specification TS 101 154.

[9] Specification for the use of video and audio coding in DVB seines delivered directly over IP, ETSI Technical Specification TS 102 005.

[101 *Specification for service information (SI) in DVB systems,* European Norm EN 300 468.

Positional Servo-System Design Using Digital Compensator

M. Mihajlović, M. Đorđević and D. Nikolić

Abstract - The paper considers positional servo-system design. The system is realized with DC motor. The desired servo-system dynamics is ensured by implementing digital compensator, which parameters are determined using pole-placement method. The designed digital controller is experimentally tested on a real servo-system.

I. INTRODUCTION

Positional servo system have very wide application in industry, in robotics and some consumer electronic appliances such as DVDs, CD-ROMs and so on. Majority of these systems are realized with DC motor, because of its very good controlling properties. In addition, DC motors are expensive and have certain shortcomings such as wear out of mechanical parts and brushes, sparking and need for maintenance.

Digital controllers have many advantages comparing to the analog ones, such as low price, flexibility and possibility for the realizations of complex control algorithms. Nowadays, most of servo-systems are controlled by digital hardware, for example [1], [2]. Modern technologies demands from the servo systems to provide good performance and accuracy in the presence of disturbances. Sliding mode control technique, which exhibit great practical robustness to parameter perturbations and external disturbances, are frequently used in servo-system design, [3], [4].

This paper considers design of positioning servo-system with DC motor. To improve position response dynamics a digital compensator is introduced. System analysis is done in discrete-time domain and the compensator parameters are tuned using pole-placement method [5].

II. MATHEMATICAL MODEL OF DC MOTOR

The well known model of a DC motor with separate excitation is given by the following equations [6]

$$\begin{aligned} u_r(t) &= R_r i_r(t) + L_r \frac{di_r(t)}{dt} + C\frac{d\theta(t)}{dt} \\ Ci_r(t) &= B\frac{d\theta(t)}{dt} + J\frac{d^2\theta(t)}{dt^2} + M_a(t) \end{aligned} \tag{1}$$

where θ is the angular position; u_r is the rotor voltage; i_r is the rotor current; M_o is the load torque; R_r is the rotor resistance; L_r is the rotor inductance; B is the friction coefficient, J is the moment of inertia and c is the motor constant. If the state coordinates are angular position $\theta(t)$, velocity $\omega(t)$ and rotor current $i_r(t)$, State space model becomes

$$\begin{aligned} \frac{d\theta(t)}{dt} &= \omega(t) \\ \frac{d\omega(t)}{dt} &= -\frac{B}{J}\omega(t) + \frac{c}{J}i_r(t) - \frac{1}{J}M_o(t) \\ \frac{di_r(t)}{dt} &= -\frac{c}{L_r}\omega(t) - \frac{R_r}{L_r}i_r(t) + \frac{k}{L_r}u(t) \end{aligned} \tag{2}$$

The control signal $u(t)$ generates the rotor voltage $u_r(t) = ku(t)$ by means of a power converter with the amplification factor k.

The electrical time constant is neglected, because it is much smaller than mechanical time constant. System order is reduced and new representation is

$$\begin{aligned} \frac{d\theta(t)}{dt} &= \omega(t) \\ \frac{d\omega(t)}{dt} &= -\frac{BR_r + c^2}{JR_r}\omega(t) + \frac{kc}{JR_r}u(t) - \frac{1}{J}M_o(t) \end{aligned} \tag{3}$$

Relation which represents system output in terms of input and disturbance is obtained as

$$\theta(s) = \frac{b}{s(s+a)}U(s) - \frac{1/J}{s(s+a)}M_o(s) \tag{4}$$

where $a = (BR_r + c^2)/JR_r$ and $b = kc/JR_r$. In case of no disturbance, (4) is reduced to

$$\theta(s) = \frac{k_m}{s(1+sT_m)}U(s) \tag{5}$$

Therefore, the resulting approximate transfer function may be written as

$$G(s) = \frac{Q(s)}{U(s)} = \frac{k_m}{s(1+sT_m)} = \frac{b}{s(s+a)} \tag{6}$$

where k_m is gain and T_m is time constant

$$T_m = 1/a\,,\ k_m = b/a \tag{7}$$

In order to control DC motor with digital controller, transfer function (6) is discretized by using Z-transform.

$$G(z) = ZL^{-1}\{G_{ho}(s)\cdot G(s)\},\quad G_{h0}(s) = \frac{1-e^{-sT}}{s} \tag{8}$$

The obtained discrete transfer function is given with

M. Mihajlović, M. Đorđević and D. Nikolić are students at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia,
E-mail: djordjevic@rocketmail.com

$$G(z)=\frac{b(-1+e^{aT}-aT+z-e^{aT}z+aTe^{aT}z)}{a^2(z-1)(ze^{aT}-1)} \tag{9}$$

III. Digital Compensator Design

Procedure of pole-placement [5] is applied in digital compensator design. Discrete transfer function (9) may be written as

$$G(z)=\frac{b_0 z+b_1}{z^2+a_1 z+a_2} \tag{10}$$

where

$$\begin{aligned} b_0 &= k_m T_m(e^{-T/T_m}-1+Te^{-T/T_m}/T_m)\\ b_1 &= k_m T_m(1-e^{-T/T_m}-Te^{-T/T_m}/T_m)\\ a_1 &= -(1+e^{-T/T_m})\\ a_2 &= e^{-T/T_m} \end{aligned} \tag{11}$$

To secure desired closed loop dynamics the following digital compensator is introduced

$$D(z)=\frac{k_p(z-z_c)}{z-p_c} \tag{12}$$

where k_p is the compensator gain, z_c and p_c are the compensator zero and pole, respectively. The considered system is shown in Fig. 1, where $k_T = kc/R_r = Jb$.

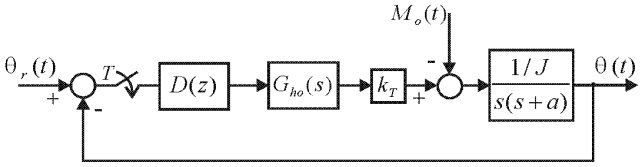


Fig. 1. Positional servo-system block scheme.

Open loop discrete transfer function is described by

$$W_p(z)=D(z)\cdot G(z)= =\frac{k_p b_0 z^2+(k_p b_1-k_p b_0 z_c)z-k_p b_1 z_c}{z^3+(a_1-P_c)z^2+(a_2-a_1P_c)z-a_2P_c} \tag{13}$$

The characteristic equation for given system yields

$$\begin{aligned} &1+W_p(z)=z^3+(a_1-P_c+K_p b_0)z^2+\\ &+(a_2-a_1P_c+K_p b_1-K_p b_0 z_c)z-a_2P_c-K_p b_1 z_c=0 \end{aligned} \tag{14}$$

The control system should provide desired closed loop dynamics, which is selected by chosen values of relative dumping factor ζ and undumped natural frequency ω_n. Therefore, closed loop dynamics is determined by the dominant complex conjugated poles z_1 and z_2

$$z_{1,2}=e^{-\xi\omega_n T}e^{\pm j\omega_n T\sqrt{1-\xi^2}} \tag{15}$$

Since the controlled system is described by the third order characteristic equation (14), the desired characteristic equation may be set as

$$\begin{aligned} &(z-z_0)(z-z_1)(z-z_2)=z^3-(z_0+z_1+z_2)z^2+\\ &+(z_0z_1+z_0z_2+z_1z_2)z-z_0z_1z_2=0 \end{aligned} \tag{16}$$

The third unknown real pole z_0 must not be dominant, i.e., it should be located closer to the origin in the z-plane than the other two.

By comparing coefficients of the actual (14) and desired characteristic equation (16), the following system of equations is obtained

$$\begin{aligned} &-P_c+z_0=-a_1-z_1-z_2-k_p b_0\\ &-a_1P_c-k_p b_0 z_c-(z_1+z_2)z_0=z_1z_2-a_2-k_p b_1\\ &-a_2P_c-k_p b_1 z_c+z_0z_1z_2=0 \end{aligned} \tag{17}$$

Solving this system for z_0, z_c, p_c gives the values of the compensator parameters and the unknown real pole z_0

$$\begin{aligned} z_c &= \frac{a_2^2+a_2(b_1k_p+z_1(a_1+b_0k_p+z_1)}{k_p(-b_1(a_1+z_1+z_2)+b_0(a_2-z_1z_2))}+\\ &+\frac{(a_1+b_0k_p)z_2+z_2^2)}{k_p(-b_1(a_1+z_1+z_2)+b_0(a_2-z_1z_2))}+\\ &+\frac{z_1z_2(a_1^2-b_1k_p+z_1z_2+a_1(b_0k_p+z_1+z_2))}{k_p(-b_1(a_1+z_1+z_2)+b_0(a_2-z_1z_2))}\\ P_c &= \frac{b_1(a_2+b_1k_p+z_1(a_1+b_0k_p+z_1))}{-a_2b_0+b_0z_1z_2+b_1(a_1+z_1+z_2)}+\\ &+\frac{(a_1+b_0k_p+z_1)(b_1+b_0z_1)z_2+(b_1+b_0z_1)z_2^2}{-a_2b_0+b_0z_1z_2+b_1(a_1+z_1+z_2)}\\ z_0 &= -\frac{-(a_2b_0-a_1b_1)k_p-(a_1+b_0k_p+z_1+z_2)}{k_p(-a_2b_0+b_0z_1z_2+b_1(a_1+z_1+z_2))}-\\ &-\frac{b_1k_p(a_2+b_1k_p-z_1z_2)}{k_p(-a_2b_0+b_0z_1z_2+b_1(a_1+z_1+z_2))} \end{aligned} \tag{18}$$

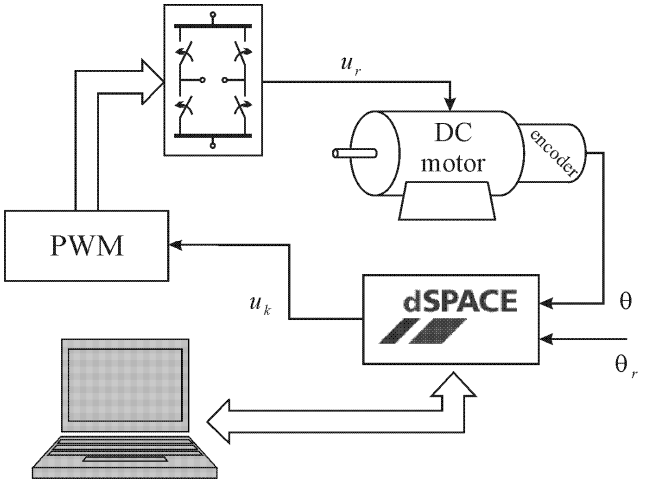


Fig. 2. Block diagram of the experimental platform.

IV. Experiment investigation

The design digital compensator has been experimentally tested on a real servo system. Experimental platform is completely described in [6] and shown in Fig. 2. Low power permanent magnet DC motor is used.

Angular position is measured by an optical incremental encoder that is mounted on the rotor shaft. The control part of the system is implemented by dSPACE DS1104 R&D Controller board, hosted by a personal computer. The board contains microprocessor Motorola MPC 8240 that is based on Power PC 603e, which operates at the frequency of 250MHz. Input-output subsystem of the board is realized by DSP microprocessor TMS320F240, operating at the frequency of 20MHz, and communication panel CLP1104 with LED indication. Sampling time is $T = 0.0004s$.

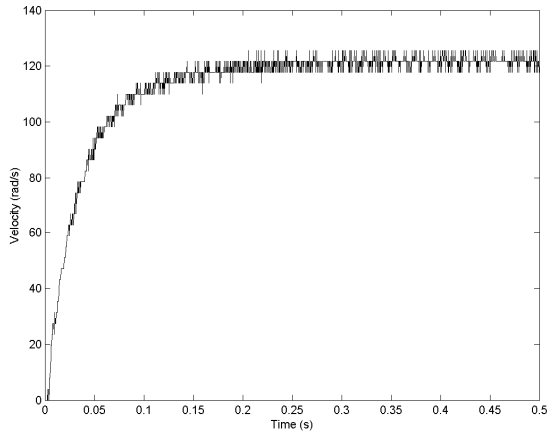


Fig. 3: Angular velocity step response.

Motor parameters have been identified using velocity step response. Transfer function with respect to angular velocity is given as

$$\frac{\Omega(s)}{U(s)} = \frac{k_m}{1+sT_m} \tag{19}$$

Identification of plant parameters is performed by standard procedure. Input signal is a step function

$$u(t) = U_0 h(t), \quad U(s) = U_0 / s \tag{20}$$

Fig. 3 shows the response for the input $U_0 = 5\text{V}$. Steady-state velocity is $\omega(\infty) = 124 \text{ rad/s}$. Using inverse Laplace transform, the response in time domain is analytically obtained according to (19) and (20) in the following form

$$\omega(t) = k_m U_0 (1 - e^{-t/T_m}) \tag{21}$$

From the steady-state, the gain k_m may be obtained as

$$k_m = \frac{\omega(\infty)}{U_0} \tag{22}$$

that gives $k_m = 24.8$. Replacing (22) in (21) yields

$$\omega(T_m) = \omega(\infty)(1 - e^{-1}) = 0.63\omega(\infty) \tag{23}$$

According to (23), the estimated value of the time constant is $T_m = 0.0379$.

For given values of $\zeta = 0.707$ and $\omega_n = 25$ rad/s, $z_{1,2} = 0.99293 \pm j0.007022$ is obtained using (15). Compensator gain is selected to $k_p = 10$. Substituting these values into (18), the compensator parameters as well as unknown real pole are obtained as: $p_c = 0.724584$, $z_c = 0.974834$ and $z_0 = 0.727702$.

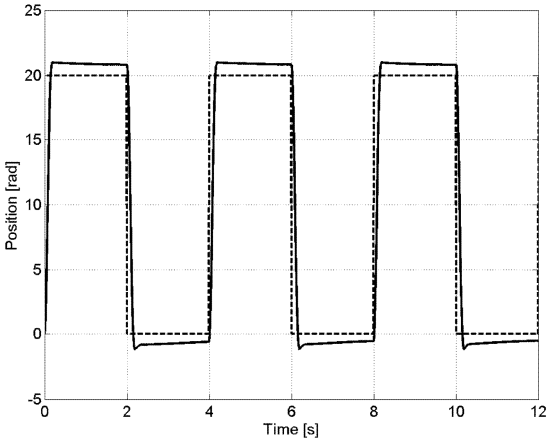


Fig. 4: Position response

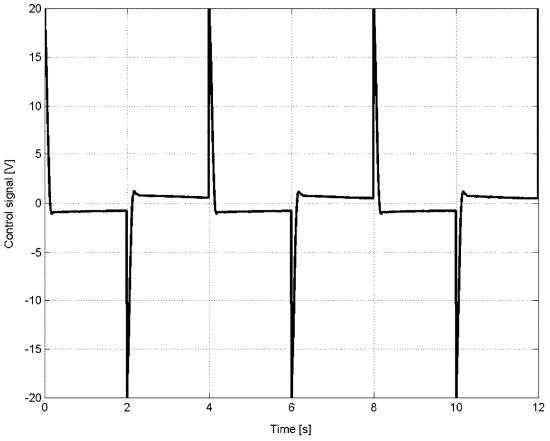


Fig. 5: Control signal.

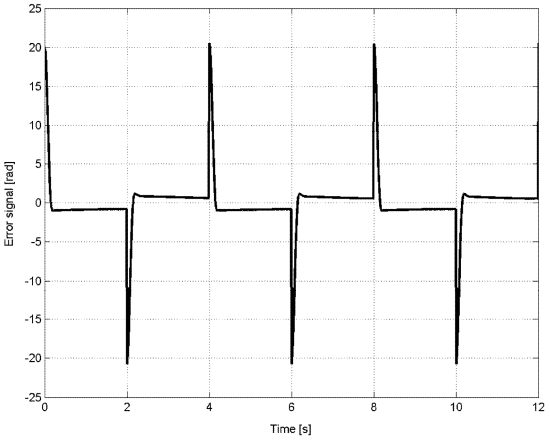


Fig. 6: Positioning error.

Position response is shown in Fig.4. Referent position is depicted by dashed line and servo system output is given by solid line. The designed system is of type one, since the plant has the integrating property. Theoretically, system

output should track ideally step position references. A certain error and deviation may be noticed in Fig. 4. Dry friction, unmodeled dynamics and nonlinearities present in the real system are the main factors for such nonideal behavior, which may be noticed only in experimental study.

Control signal is shown in Fig. 5. Fig. 6 gives the positioning error. Error signal rapidly converges to zero value but it never reaches it because of the presence of the above mentioned imperfections.

V. CONCLUSION

The paper presents the design of positional servo system, by using digital compensator. Analysis of the presented system is done in discrete-time domain, and the compensator parameters are selected using pole placement method. Experimental results show certain differences with respect to the expected accuracy, which are caused by an un-modeled dynamics, nonlinearities and especially by dry friction. However, the desired positioning dynamics is attained.

REFERENCES

[1] Kiyoshi Onishi, Masato Nakao, Kouhei Ohnishi and Kunio Miyachi "Microprocessor-Controlled DC Motor for Load-Insesitive Position Servo System", IEEE transactions on industrial electronics. Vol. IE-34. 1, February 1987.

[2] Milic R. Stojic "Design of the Microprocessor-Based Digital System for DC Motor Speed Control", IEEE transactions on industrial electronics. Vol. IE-31. no. 3. August 1984

[3] Fu-Juay Chang, Hsiang-Ju Liao and Shyang Chang "Position Control of DC Motors via Variable Structure Systems Control: A Chattering Allevation Approach", IEEE Transcations on industrial electronics, vol. 37, No. 6, December 1990

[4] Ali J. Koshkuoei and Keith J. Burnham "Control of DC Motors Using Proportional Integral sliding mode", Control Theory and Applications Centre, Coventry University, Coventry CV1 5FB, UK

[5] Milica Naumovic "Control System Design", Faculty of electronic engineering and WUS Austria, Nis, 2005, (in Serbian)

[6] Boban Veselic, Doctoral Dissertation "Application of Digital Sliding Mode in Sinthesys of Robust for Coordinated Tracking of Complex Trajectories", Faculty of Electronic Engineering, Nis, 2005 (in Serbian)

A Microcontroller Based System for Testing Fuel Injectors

B. Manojlović

Abstract - The fuel injectors are very sophisticated and important parts of a motor vehicle and need special attention during the maintenance. Being expensive to buy, it is necessary to repair them as many times as possible. Since their work can not be inspected while they are running inside the engine, testing and cleaning of injectors requires special equipment. That equipment is also very expensive and not many workshops possess this kind of tool. Therefore, there is a plenty of space on the market to develop and sell variety of tools for testing fuel injectors which cost less then the existing ones and perform equally or even better.

This product is meant for the mid-sized workshops which perform universal maintenance of motor vehicles and are limited in funds. Testing and cleaning of fuel injectors is not their primary duty, so they don't invest vast sums of money in the injectors' testing equipment. But since the overall performance of the engine greatly depends on the proper work of fuel injectors, all kinds of workshops need fully functional testing equipment. Solution for them is the equipment that has all functions as the expensive heavy-duty equipment, but costs several times less and is intended for less frequent use.

I. Introduction

The fuel injectors are very sophisticated electromagnetic valves that precisely control the amount of fuel injected into the air/fuel mixture of the engine. In the multi-point injection engines there are as many injectors as there are cylinders. There is one injector for each cylinder and it is situated in the intake port just before the intake valve. By pressing the acceleration pedal, driver defines the amount of air to be surged into the engine. Then, it is up to the fuel injectors to inject the adequate amount of fuel for the air/fuel mixture to be appropriate. Air/fuel ratio in the mixture must be kept constant at any time.

The injectors work in a way that in each cycle of the engine they inject adequate amount of fuel. They accomplish this by being open for appropriate amount of time in each cycle. Generally, they are driven by electrical pulses. Period of these pulses is proportional to the revolutions per minute of the engine, and the duration of the pulses is proportional to the amount of fuel injected. There is an electronic control unit (ECU) in the engine which collects data from several different probes and calculates the duration of the pulses. This electronic unit also generates these pulses.

Manufacturer of the injectors defines the flow for each model (the amount of fuel injected in the amount of time at defined pressure). ECU uses this data when calculating the pulse length. But during the time of exploitation, flow of the injectors changes. Fuel is never pure. It contains water and dry matter. The scale from water and dry matter settle at the nozzle which disperses fuel thus causing the reduction of flow. The only way to alleviate this is to take out the injectors from time to time, measure their flow, clean them if necessary and put them back into the operation.

II. Problems with Faulty Injectors

The scale and dry matter from fuel reduce the performance of the injectors in three ways: flow becomes inadequately low; flow differs from cylinder to cylinder; and jet of fuel loses proper shape. There are also mechanical problems with the injectors that can occur.

A. Inadequate flow

Impurities from the fuel settled on the nozzle of the injector cause the reduction of flow. Electronic system attached to the engine is capable of dealing with this kind of problem but only to one extent. ECU drives injectors in a feedback. One of the parameters ECU uses into account when calculating pulse duration is the structure of exhaust fuses. There is a probe that detects the amount of each gas in the exhaust gases. By looking at the exhaust gases one can tell whether the air/fuel mixture was too weak or too strong. When ECU, through the probe, detects that there was not enough fuel in the air/fuel mixture, ECU then through the feedback increases the duration of pulses in each cycle. But this works only to a certain degree of flow reduction. When flow gets greatly reduced problems occur first at high revolutions of the engine. At high revolutions of the engine, the period of each cycle is very short and there may not be enough time to inject the enough fuel. Therefore, maximum power of the engine is reduced. There is also one more problem related to this. When the engine is at lower revolutions and the driver steps hard on the acceleration pedal, huge amount of air runs into the engine. Due to the reduction of flow, injectors might not be able to inject enough fuel to make adequate air/fuel mixture. This causes a car to accelerate at lower rates.

B. Manojlović is a graduate engineer, School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia,

E-mail: manojlovic.b@sezampro.yu

All in all, the reduced flow of injectors causes a car to lose it's acceleration and maximum power. Luckily, this is something that an average driver can detect and go to the mechanic for further inspection.

B. Unequal Flow Among Cylinders

Although all injectors in an engine come from the same manufacturer, are the same model and are bought at the same time, they are not 100% equal. And during the period of exploitation their flow reduces differently. Most cars have a probe that measures joint exhaust fumes from all cylinders. That probe gives the information of whether the air/fuel mixture was weak or strong, but that is the average information for all cylinders. Apart from the very expensive ones, cars do not monitor air/fuel ratio for each cylinder. ECU generates the same pulses for all injectors. If they don't have the same flow, they will inject different amounts of fuel into combustion chambers. So, some cylinders will have a too weak mixture, and some too strong. On average, mixture will be adequate, but almost none of the cylinders will gather maximum power from combustion.

Maximum power from a combustion is gathered only if the air/fuel ratio is 14.7:1 in favour of air. Only then all the fuel is burned while all the oxygen is used. Deviation from this in some cylinders will cause overall degradation in engine performance. Unlike the previous, most cars can not compensate for this problem.

C. Improper shape of the stream

The scale and the dry matter deposited on the nozzle of the injector not only reduce the flow, but change the shape of the stream. The stream should be cone shaped, misty and almost invisible (Fig. 1.). Drops produced by the nozzle should be microscopic and stream should be as wide as possible. That is the best way for the fuel to completely mix with the air. During the time, stream becomes narrow and drops become larger. This way, not all fuel mixes with air. Excess fuel that is not burned simply leaves the combustion chamber with the exhaust fumes. ECU calculates that the mixture is weak and pours more fuel which produces more unusable excess fuel. This way engine increases fuel consumption while downgrading performance.

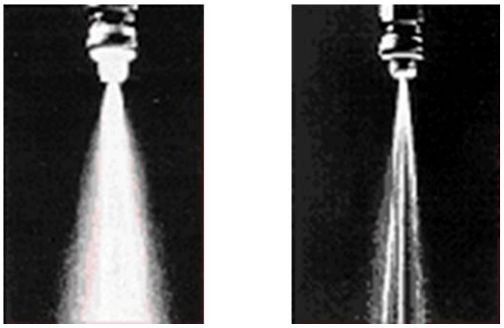


Fig. 1. Shape of the injector's stream: good (left) and bad (right)

D. Mechanical problems with the injectors

The injector is an electro-mechanical component. Therefore it is susceptible to mechanical failures. It is possible that with time, the injector begins to open or close slower. Or it becomes unable to fully open or fully close. This can be worsened if the injector is faulty in a way that in some cases opens and closes well, and some times not. All these result in unpredictable amount of injected fuel. Usually only one injector from the set gets faulty this way, however it influences work of the whole injection system.

III. System for Testing Fuel Injectors

The fuel injectors are situated in the intake port of an engine. Nozzle and the stream are not visible while the engine is running. Therefore, it is not possible to directly examine the work of injectors while they are in the engine. Only by monitoring some other parameters of the engine one can suspect that the injectors may be cause of some problems. The only way to be sure whether injectors are working well or not is to take them out of the engine, simulate their work on the working table and observe their operation. This is what we call testing of the fuel injectors.

Testing of fuel injectors requires special equipment. That equipment should be able to supply injectors with the pressurized liquid and to drive them with the same pulses like the ECU in the engine does. The liquid used for testing is petroleum. Petroleum has similar density as petrol but is less volatile. Apart from being able to simulate work of the injectors, testing equipment should also be capable of measuring the injectors' performance.

System described in this paper (Fig. 2.) has all the abilities mentioned above and is able to perform several different test and measurements with the injectors.

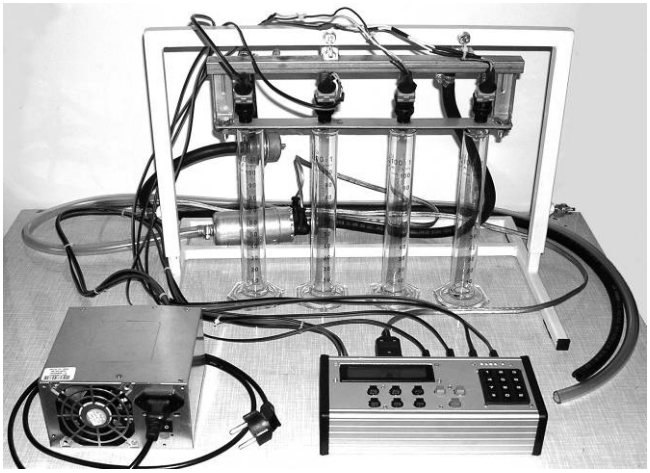


Fig. 2. The Microcontroller Based System for Testing Fuel Injectors

A. Mechanical components of the system

This system consists of the following parts: metal frame, box with electronics (Fig. 3.), pump, hoses, gauge glasses and power supply.

Metal frame is height adjustable to accommodate various types of injectors. It can be easily set and fixed at any height by two screws situated sideways. Metal frame holds the fuel rail in which the injectors are put. The fuel rail is connected to the pump and supplies injectors with petroleum. The frame is also designed to accommodate four gauge glasses that collect injected petroleum and measure the amount of injected fuel.

Box with electronics is the main part of the system. These electronics drive the injectors and power the pump. The box is equipped with a keyboard and an LCD which are used to enter testing parameters.

Pump and hoses form the hydraulic system that supplies injectors with petroleum at appropriate pressure. Petroleum is held in a container from where it is suctioned by the pump. The pump pressurizes the petroleum to 5-6 bar. Petroleum under that pressure enters the rail. At the other side of the rail there is a control valve set at 3 bar. Excess liquid is passed through the valve and directed by another hose back into the container. The valve ensures that the pressure in the rail (and in the injectors) is kept constantly at 3 bar whether the injectors are opened or closed and independent of their flow. All hoses used are pressure-resistant.

Petroleum that passes through the injector is injected into the gauge glasses. There is one gauge glass per injector and they allow a mechanic to measure the flow of each injector. Maximum capacity of a gauge glass is 100 ml which is more than enough for a test that lasts 60 seconds.

Since the pump and the injectors operate at high pressure, they require a lot of power. The system is equipped with the power supply which is plugged into the 220V network and supplies electronics, pump and injectors with adequate DC voltage.

B. The use of the system

The first step when planning to test the injectors is to take them out of the engine and put them into the rail on the frame. The rail has the ability to additionally tighten the injectors from below in order to prevent the from being blown downwards by the pressure of the petroleum. Once the injectors are positioned, the gauge glasses should be put under the injectors and the frame should be set at the right height.

The electrical part of the system is designed in a way that all the cables are equipped with the connectors so parts can be easily detached. There are four cables that supply injectors with electrical pulses, and they should be connected one to the injectors and others to the box with electronics. The same should be done with the pump.

Once everything is connected, the power can be switched on (there is an on/off switch on the power box). The power box is attached to the 220V network and supplies the electronics with 12V DC voltage. The pump and the injectors are electrically supplied from the electronics box. This way the operator can control the pump and the injectors via the keyboard.

The purpose of this system is not to merely supply the injectors with petroleum and current, but to perform several different tests under various conditions. The system can perform three different tests.

The first test is the pulse test. This test is meant to simulate the ECU from the engine. The electronics generate the exact pulses as the ECU does, but here the

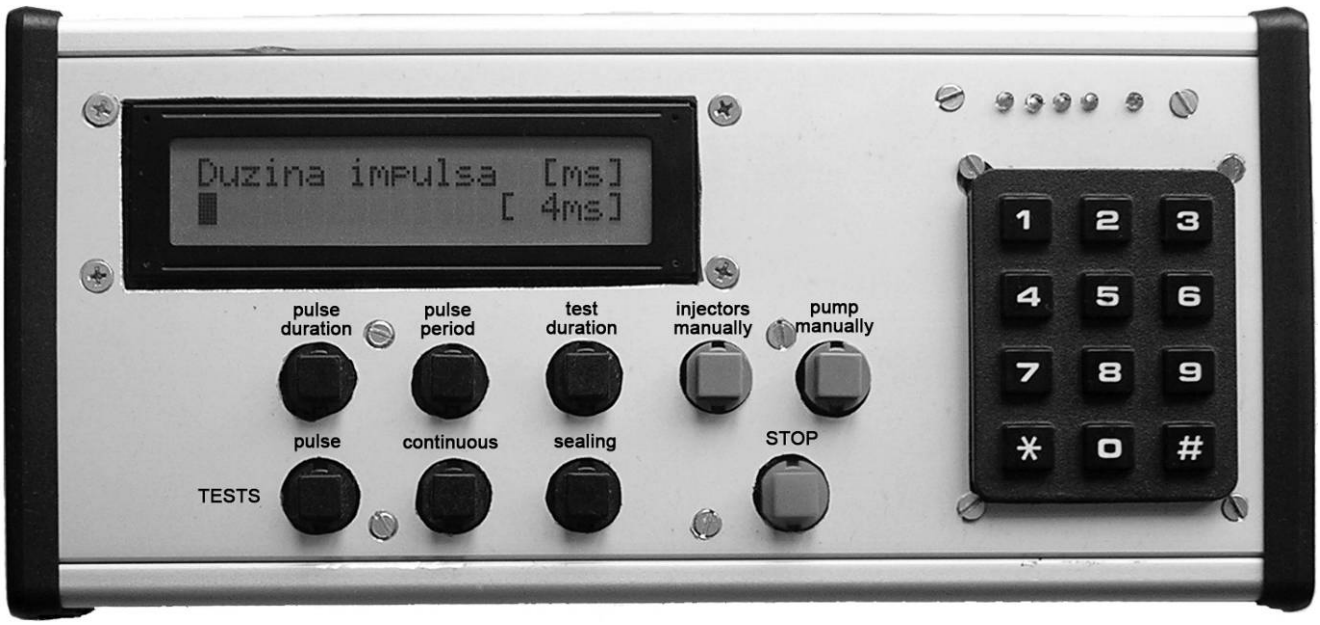


Fig. 3. The box with electronics

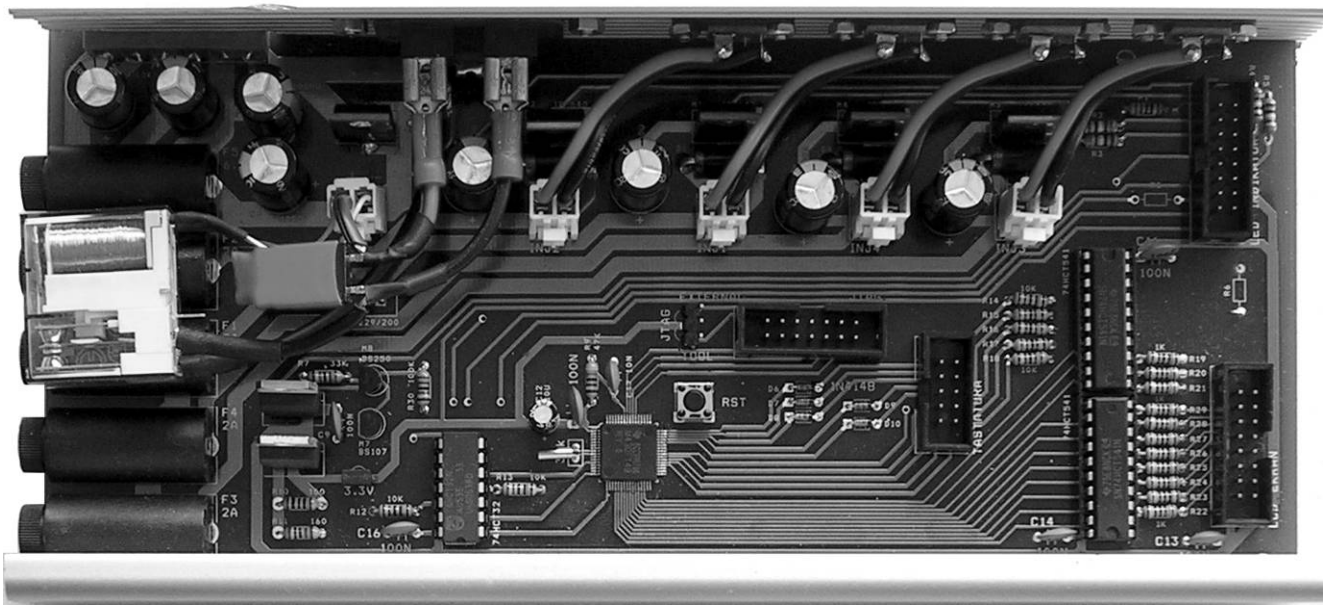


Fig. 4. The printed circuit board

operator chooses the pulse duration and the pulse period. The pulse duration and period are set using the keyboard before the test is started. The purpose of this test is to measure the flow of each injector and to compare whether the flow is equal among the injectors.

The second test is the continuous test. This test keeps the injectors and the pump running all the time and it is meant to visually observe the shape of the stream (the shape of the stream is much better visible when it flows continuously than when in pulses).

The third test is the sealing test which keeps the pump running while the injectors are shut. Good injectors should not drip the liquid while they are shut under pressure.

Apart from the mentioned parameters, testing time can be set for all three tests. Testing time is usually around 60 seconds.

When the tests are finished the best thing should be to compare the test results with the new injectors so the mechanic can say whether the injectors need repair or not. So, the workshop should test the new examples of injectors for the models they repair and save those data for later comparison. If the tests show that the injectors are dirty, they can be cleaned in the UV chamber (not included in this system) and tested again. If the cleaning proved successful, the injectors are put back into the operation. Otherwise, the injectors that can not be cleaned should be replaced with the new ones.

IV. The Electronics

The central part of the electronic subsystem is the microcontroller. The microcontroller generates the pulses for the injectors, switches the pump circuit, measures all necessary times, drives the LCD and scans the keyboard. The whole electronics is implemented on a printed circuit board (Fig. 4.) sized 220x100 mm and accommodated in the appropriate box. PCB is double layered with 35µm copper layers. The board accommodates the microcontroller with the surrounding components, the fusses (one for each injector and one for the pump), the transistors that amplify the pulses, the relay switch for the pump and flat cable connectors for LCD and keyboard (both mounted on the case cover). The board is supplied with 12V DC and there are components that transform this voltage into 5V and 3.3V DC required by digital circuits.

V. Conclusion

There are similar products on the market but cost several times more then this one. The saving is achieved by avoiding expensive sophisticated mechanical components which makes this product affordable to an average automobile workshop. This product should contribute the testing and cleaning of fuel injectors to become a common practice and to be performed on a regular basis and not only when the problem occurs.

References

[1] M. Prokin, *"Mikroprocesorska elektronika"*, Akademska misao, Belgrade, 2003.
[2] L. Kraus, *"Programski jezik C sa rešenim zadacima"*, Akademska misao, Belgrade, 2001.
[3] D. Živković, M. Popović, *"Impulsna i digitalna elektronika"*, Akademska misao, Belgrade, 2000.
[4] *"MSP430x1xx FamilyUser's Guide"*, Texas Insruments, 2006.

Lock-free Read Set Structure for Software Transactional Memory

Ivana Teodor, Nikola Nikolić, Dušan Stanković and Miroslav Trajanović

Abstract – Transactional Memory is a new promising technology for development of parallel applications. Idea is that a block of instructions is executed as a transaction: atomically (all or nothing) and isolated (there is no memory interleaving with other transactions). There are several approaches in research and development of Transactional Memory: Software Transactional Memory - where everything is implemented on software level in compiler and runtime system, Hardware Transactional Memory -where everything is implemented in hardware without any software support and Hybrid Transactional Memory - where some parts are implemented in software and some in hardware.

In this paper we will present the new approach in development of Software Transactional Memory (STM). We will introduce several techniques which will significantly increase the performance of previous STM systems. We will introduce the time stamp based metadata structures, lock-free read set structure, two level read and write set structures and pipelined commit. We have implemented those techniques and obtained promising results. Our results showed that those techniques could be implemented in existing STM systems, which would increase their performance.

I. INTRODUCTION

Computer systems with a single processor have almost reached their limits. In the last decade, we saw incredible improvements in CPU design and performance. In the beginning of the nineties, there were CPUs with one million transistors (e.g. Intel 486) and in 2005. there were CPUs with one billion transistors (e.g. Intel Itanium), which is an improvement of 1000x. CPUs with mode transistors have mode hardware units and are much mode efficient, but now they are running into problems which will prevent their future improvements. For example, more transistors spend much more electricity and they heat the die more. Today's CPUs dissipate the same amount of energy on a square millimeter as the burner while cooking coffee. Very soon we will reach the point when it will be impossible to cool down the single CPU. Other problem is that if we increase the clock frequency two times, CPU speed will be increased 1.5 times and the dissipation will be increased 4 times. There are many other problems in electronic and electromagnetic which are also preventing future improvements. For all of those reasons, researchers are trying to put smaller CPUs on a single chip instead of one big. Idea is that every 18 months we double the number of CPUs (cores) on a chip, instead doubling the number of transistors in a single core. This is a trend in today's research and we even have commercial laptops with 2 or 4 cores which can be bought on the market.

Although it is feasible to have several cores on the single die, programming of multicore systems is still very difficult. Present programming models are extremely difficult to use for the majority of mainstream programmers. This is the biggest problem why multicores are not in the wider use, although their hardware potential is obvious. Simply, there are very few programmers who can exploit the full potential of multicores. If we want to exploit potential we have in multicore systems, we need to come up with new programming models. Current programming models for parallel and concurrent programming are based on locks and some of the problems of those models will be presented in following section.

A. Problem with programming models based on locks

Locks are a very well known concept. In order to have safe concurrent memory access, before access to a shared memory location, a lock needs to be acquired. This is a very simple concept ,but it introduces many problems.

If we want a lock to guard a larger memory area, then we are talking about coarse grain locking. This is easy to use, but performances are very poor, because it reduces the concurrency and the gain from multicore systems is minor. On the other side, if lock guards a smaller memory area, we are talking about fine grain locking. Performances of software written using fine grain locking strategy are very well, performances are good, but this programming model is extremely difficult to use. Some of the problems which are introduced with fine grain locks are deadlocks, live locks, priority inversion, starvation,

I. Teodor and D. Stanković are with the Department of IT System Management, Faculty of Information Technology, Trg Republike 3, 11000 Belgrade, Serbia, E-mail: ivana.teodor.259@fit.edu.yu

N. Nikolić is with the Department of Computer Science, Faculty of Electrical engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia, E-mail: nikolanixi@gmail.com

M. Trajanović is professor at the Faculty of Information technology Trg Republike 3, 11000 Belgrade, Serbia, E-mail: miroslav.trajanovic@fit.edu.yu

problem with debugging (use case scenarios cannot be reproduced) and come understanding and maintenance.

Deadlock is the most serious problem. We can easily have a deadlock in the system if we don't acquire locks in order. This is very difficult to achieve especially in large systems, when programmer doesn't know all implementation details of the library functions which he or she is using. This limits code composability and reuse which is mandatory feature in modern programming models. Those problems are very serious. They slowdown the system performance, they enlarge the development time and reduce the programmers productivity.

Transactional Memory is a promising technology which should tackle the problem of writing the concurrent applications. It should remove the burden of writing the parallel programs using locks. Concept of memory transaction is similar with concept of transaction in database systems. Difference is that transactions in DB systems usually last a couple of milliseconds, while memory transactions last a couple of micro seconds. Although the concept is the same, since memory latencies are different ,as well as other characteristics, different algorithms and approaches need to be applied.

Transactional Memory can be implemented on different levels, therefore, there are several kinds of Transactional Memory: Software Transactional Memory (STM) – where memory transactions are done on compiler/runtime level (to hardware support), Hardware Transactional Memory (HTM) – where memory transactions are done in completely in hardware, without any software support, and Hybrid approach which should integrate the best parts from STM and HTM.

In this paper we will describe one novel approach of software transactional memory implementation. The main characteristic of this approach is that it will accelerate small transactions, which are the most frequent , and it will accelerate the cleanup of data structures (metadata) which are needed for management of memory transactions. We will also describe the design of the lock-free data structures which are used for implementation of the read and write set. This design is the most crucial one, since those structures are stored on the memory access critical path, and they have direct impact on the performance.

Contributions of this paper are:

- We introduce the concept of Transaction Time stamp for the fast metadata cleanup;
- We introduce the lock-free read set for the faster concurrent access to that structure;
- We introduce the two-level read and write sets, for acceleration of small transactions, which are the most frequent;
- We introduce the statically allocated metadata structures for management of all transactions as a single array. This is good because the number of concurrent transactions is rather limited (by the number of cores), so in majority of cases we don't need dynamically allocated metadata structures;
- Pipelined commit operation for master commit.

In the Section 2 we will present the existing solutions. In section 3 we will describe details of the solution developed by Milovanović et al. [15][17] since we will base our research on that. In section 4 we will present our improvement and in section 5 we will present the results we achieved. In section 6 we will conclude our current work.

II. Existing Solutions

There are several research implementations of software transactional memory: Transactional Locking 2 (TL2) developed by Dice et al. [19], Rochester STM [20], Intel C++ STM Compiler [21] and Nebelung developed by Milovanović et al. [15][17].

TL2 and Rochester STM are based on versioning. They use 32bit words as "version-lock" for protection of transactional data. Version lock is consisted of 1 bit as information if memory is locked or unlocked and other 31 bits as the version. They statically pre-allocate a number of version locks (by default 1M) and on memory access they work with those pre-allocated lock. Memory location is typically assigned to its version lock using a hash function. When memory is read the first time, its version is stored in some temporary location. When the transaction is to be committed, commit algorithm checks if the version of the data in memory is the same as it was read in the beginning. If that is the case, then the transaction is committed and all versions of committed data are incremented, and if it is not the case, then the transaction is aborted. This algorithm is very elegant and very fast, but the main disadvantage of this approach is that transactional space is very limited and it is not likely that transactions will be as small in real world applications. This fact limits the potential of this approach.

Intel C++ STM Compiler is a very good compiler which supports transactional memory. This is the first professional prototype compiler which can be downloaded from Intel's website and is used for writing real word TM applications. But, of course, its code is closed so we couldn't do our research on fore mentioned compiler.

III. Experience with Nebelung 1.0

We evaluated the work of Milovanović et al. and their STM library implementation Nebelung 1.0 [15][17]. The main assumption in their library was that there can be an arbitrary number of concurrent transactions and that transactions' read and write sets can be of arbitrary large size. For that reason, all metadata structures for management of transactions were allocated dynamically, which was the main source of the overhead (heavy use of malloc and free). Since those nested transactions are implemented using flattening technique, it turned out that the number of transactions is at most equal to number of threads. In case of high performance applications, number of threads is equal to number of cores, which took us to

the conclusion that the number of maximal parallel transactions is reasonably small, 128 or much smaller. On the other hand, in applications we executed ,we realized that the read and write sets of the transactions are very small. In case of transactional insert operation in B+tree, where B+tree contained 2MB of useful data, average read set was just 300B and average write set was just 60B! This took us to two conclusions - that read and the write sets should have two levels: first level should be statically allocated and used as the fast path for the small transactions, and the second level which should be dynamically allocated and used in case the first level is overflowed.

IV. Transactional Metadata Structures

We realized that the maximal number of concurrent transactions is reasonably small (equal to the number of cores), as we have already mentioned in the previous section. This took us to the conclusion that all transactions' metadata structures should be statically allocated as an array with N elements, where N is the number of cores. In this way, each thread automatically knows which metadata structure to use, according to its own ID. Transaction metadata structure is presented in Figure 1. Each transaction has its own ID and the status. In case that the transaction is committed, only the status should be changed to TS_EMPTY, and there is no need for any other operations in order to handle the metadata structures.

Read and write sets are separate structures, because they have the different properties. The main difference is that read set can be accessed from multiple threads, and that's why it has to be thread safe, but write set is used just from the single thread and doesn't have to be thread safe. We separated them and designed each of them according to the specific needs and responsibilities.

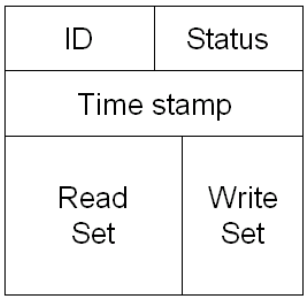


Fig. 1. Transaction metadata structure. Transaction metadata structure is consisted of the transaction ID, status, time stamp and read and write sets.

We have also introduced the concept of transaction time stamp. We have realized that cleanup of metadata structures should be fast, no matter if transaction is aborted or committed. In both cases transaction finished the "old" execution, and therefore metadata structures need to be cleaned. The time stamp of the transaction should be understood as the number of the iterations in which the transaction is executed. All data in the read and write set will also be marked with exactly the same time stamp as the transaction time stamp. In case of aborting of the transaction, the time stamp should be just increased and immediately it will invalidate all metadata.

IV. Read Set Structure

Read set is a structure which should keep the information about the variables (e.g. addresses) which were read during the transaction. The only thing it has to keep is the address of the read data. Read set structure is presented in the Figure 2.

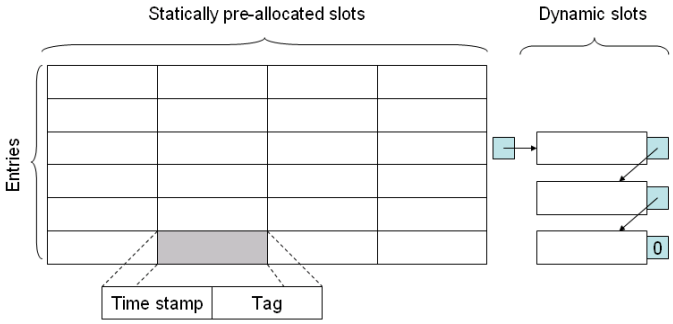


Fig. 2. Read Set structure. Read set structure has two levels. The first is statically allocated and should be capable to handle "small" transactions. The second level is dynamic and should be used in case of overflow of the first level. Each data which is stored in the read set has the time stamp which is the same as the time stamp of the running transaction.

Read set is basically organized as the hash table. For each address, which needs to be stored, is calculated the entry number and the TAG value. Each entry has a several statically allocated slots. If there is free slot in the static part it should be used. If there is no free slot in the static part, the new slot should be allocated in the dynamic part. Each entry has its own dynamic part. Slots in the static and the dynamic part are exactly the same.

Slots in the dynamic part shouldn't be released on the transaction commit or abort, because it is likely that the memory footprint of the transaction in the next iteration will be very similar to the previous one and the dynamic entries will be needed again.

V. Lock-free design of Read Set

We analyzed the design of our system and the access to read set structure, and we realized that it can be accessed from at most two threads at the time, where one is the reader and the other is the writer. Only the transaction that owns the read set can modify it (one writer). Also, the transaction which performs the commit operation will access the read sets of other transactions in order to check the conflict detection (reader). Since that only one transaction can commit at the time it means that

we can have at most one external reader of the reads set. Those two conclusions are the key of our design. Read set structure will be designed for one writer and one reader.

Since, in the read set in each slot we have two values: the time stamp and the tag, it would be sufficient if the reader and the writer access those values in the opposite order. It means that the writer writes the first the tag and then the time stamp, and the reader first checks the time stamp and then the value. In this scenario, reader will always read the consistent data.

We also need to prove that this lock free implementation will not affect the overall transactional mechanism. Figure 3 presents the pseudo code of the transactional read and commit operations. Transactional read will always check first if the data is in the local buffer (e.g. write set) (R1); if data is not in the local buffer, the address will be stored in the read set (R2) and on the end, data will be read directly from the memory (R3). Commit operation will be executed in the following way: data will be committed unconditionally to memory first (C1) and then it will be checked if some other transaction should be aborted because of the conflict (C2).

```
void* stm_read(Transaction* t, void*
address){
  R1. if (data is in write set)
       return the address in the write set

  R2. put the address in read set

  R3. (indirectly means read from the
memory)
  return address
}

void commit(transaction* t){
  Lock(commit_lock)
  // One transaction can commit at the time

  C1. write local buffer to memory

  C2. check conflicts
  Unlock(commit_lock)
}
```

Fig. 3. Transactional read and commit operation. Pseudo codes for the transactional read and commit operations.

Operations R2 and C2 can occur at the same time in the different threads and to access the same read set structure. From the read set point of view R2 is the writer and C2 is the reader. Thanks to the underneath implementation, it is not a problem to execute R2 and C2 at the same time. We cannot guaranty which operation will be executed "logically" first, so we need to be sure that in both scenarios the result will be correct. We need to be sure that every conflict is either detected or the reader is reading the updated value. Let's analyze the case where the operation R2 wants to insert address A in the read set RS1, and operation C2 wants to check at the same time if the address A is in the same read set RS1. If operation R2 is executed before operation C2, then conflict will be detected, since the address A will be inserted in the read set before the conflict check. If operation C2 is executed before the operation R2, that means that the logical order of operations is C1, C2, R2, and R3. Since the operation C1 comes before the operation R3, then the transaction which executed read will read the fresh copy of the data, no matter if conflict is detected or not, because of the address A.

VI. Lock-free design of Read Set

In our implementation, write set structure is the transaction local storage or the buffer for speculative memory writes (in the literature referred as *the lazy versioning*). This storage will be written in memory on the transaction commit. Only the transaction owner can access this structure, so it doesn't have to be thread safe.

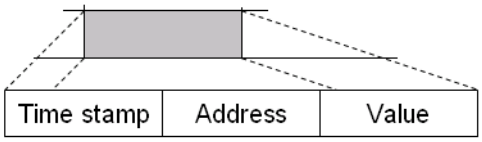


Fig. 4. Write set structure. Write set has the same structure as the read set except that slots now contain the time stamp, the real virtual address for the data and the data it self.

In applications we executed, the write set is usually very small. For that reason we will design the write set to be able to handle the small transactions in the first level.

VII. Pipelined commit

At the beginning, we implemented the "atomic" commit. Commit operation was consisted of flushing of the write set to memory and conflict detection with other transactions. The whole commit operation was executed atomically. We realized that in that case 50% of the time is spent on the commit operation or in waiting to commit. That's why we decided to parallelize the commit operation.

Approach we took was to execute the commit operation in a pipeline fashion. We designed the two-phase and the three-phase commit operation. Figure 5 illustrates the commit operation phases.

In both cases of two and three phase commit we have the commit queue. Whenever a transaction wasn't to commit, it put itself in the commit queue and waited its turn to commit. In our current implementation transactions will commit in FIFO order in the pipelined fashion.

In case of two phase commit, in the first phase we execute the flush of the write set of the first transaction from the commit queue and check for the conflict with the next transaction in the commit queue. In the second phase, we check the conflicts with all other transactions in the

system. Expected gain is to execute the flush of the write set and the conflict checking of two successive transactions in parallel.

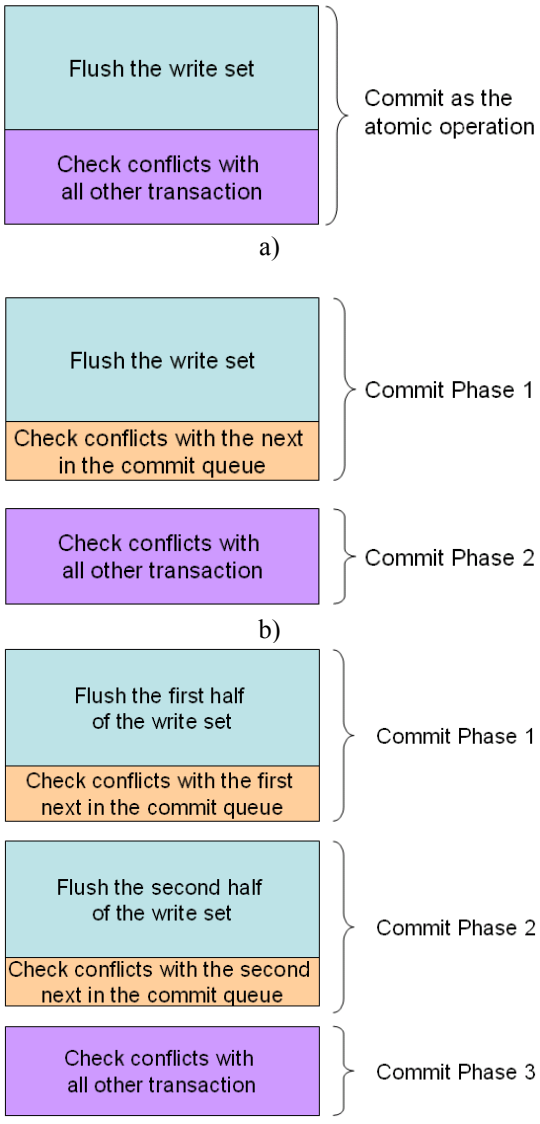


Fig. 5. Commit operation. a) atomic commit, b) two-phase commit and c) three-phase commit. Cases a) and b) are self explanatory. In case c) we used the property of the write set that address A will always go to row f(A), so the first and second half of each two write sets are disjoint and we can flush them in parallel.

In our experiments, we noticed that flush of the write set is two times longer then the conflict detection. We want to exploit that property in case of a three-phase commit operation. We also exploited the property of our write set structure (hash table in essence) that address A is always stored in the same row - f(A). That means that the first half and the second half of the write set of any two transactions will be disjoint. So, we split the flush to memory in parts. In the first phase, we flush the first half of the write set and check for conflicts with the first next transaction from the commit queue. In the second phase, we flush the second half of the write set to memory and check for conflicts with the second next transaction in the commit queue. In the third phase, we check for conflicts with all other transactions. Expected gain in this case is: 1) two transactions will commit in the same time and 2) conflict detection operation of one transaction will be executed in parallel with the flush to memory of two other transactions.

Similar to those two approaches, we can invent any other multi-phase commit operation. In theory, from the core/CPU point of view, it would be optimal if we had N+1 phase commit, where N is the number of cores. That is because we should divide the write set in N parts, so each core can commit one part in parallel and one phase for the conflict detection. Now, we realized that memory becomes the problem/bottleneck. It is true that multiple cores can commit (flush write sets) in parallel, but the question is if the memory can absorb all those writes. Also, conflict detection is not so fast, and maybe conflict detection should be also split in more phases, but still, conflict detection accesses to memory and the question is if memory can absorb so many accesses.

A. Write Set Reduction

When developed the commit queue, we realized that in fact write set are stored in the queue and that they will be committed in order. It can happed that address A should be written from both WSi and WSj where $i < j$. If any other transaction T access the same location A, when it tries to commit it will be stored on position k, where $j < k$. That means that transaction T should see the version of A, which should be committed from WSj. That means that whenever we have the intersection of write sets in the commit queue ($WSi \cap WSj \neq \emptyset$), we can flush just $WSi \setminus WSj$ and WSj.

Idea: If we see the intersection of WSi and RSj we can bypass the information and just update the write set WSj, instead of aborting the transaction. This can be helpful in case of locations which frequently cause conflicts, like shared counters.

X. Evaluation

We evaluated our implementation with several benchmarks: B+tree benchmark, Genome and Kmeans form STMAP benchmark suite, Gauss-Seidel application and AMMP application. We compared our solution against solution developed by Milovanović et. al and we used the same benchmarks as they did. Our solution in general introduces the speedup of arround 30% to 40%. We executed our tests on an 8 core 3.2GHz Intel(R) Xeon™ CPU machine with 8MB of cache

A. B+tree application

In B+tree application we tested the performance of our system, and compared it against Nebelung 1.0, when it operates on a complex data structure with huge amount of data (~100MB). We executed 80 transactions per thread where each transaction atomically takes the data from one B+tree structure, performs compute-intensive non conflicting operations for an average of 50µs and stores the result in the other B+tree.

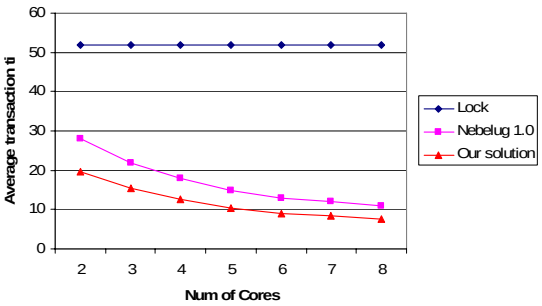


Fig. 6. Results for the B+tree application. Test compares the average transaction execution time (total execution time / number of transactions) of the Nebelung Runtime system and of our system and the appropriate coarse-grain lock based system (total time / number of iterations). One transaction corresponds to the one iteration in coarse grain approach. Results show performance on a par with a fine-grain lock based approach even though the programming model is as easy as a coarse-grain lock based approach. Our system introduces a speedup of around 30%.

Figure 6 compares the average transaction time using coarse grain lock-based approach, Nebelung TM system and our system. TM is better because B+tree structures are filled with a large amount of data so the probability of the conflict is very low (read sets intersects but write sets and read sets don't) so transactions are really executed in parallel without too many aborts. This is obvious for example in case with 2 cores, TM approach is two times faster then the coarse grain lock-based approach. We also see that our approach introduces around 30% of speedup.

In case of the B+tree data structure, performance of the coarse-grain and fine-grain lock based solution is almost the same. That is because when we insert the new value in a B+tree, there is a possibility that even root node can be changed so we need to lock the root also and that is exactly what we are doing in case of coarse-grain solution. So with this application we showed that our system and Transactional Memory in general can give better performance in cases when fine-grain solution doesn't exist or is to difficult to be implemented (e.g. two phase insert in B+tree or similar).

B. Stamp – Genome and Kmeans applications

We also successfully executed two benchmarks from STAMP benchmark. We had 28% speedup over Nebelung. Original Nebelung executed those applications for around 15 and 30 seconds. We executed those applications for 10.1 and 21.6 seconds.

C. Gauss-Seidel application

We tested our system against Gauss-Seidel application, which implements Gauss-Seidel finite difference method for solving a linear system of equations [18] using Transactional Memory and retry. For each element of the matrix we created a separate thread, which will iteratively calculate the value of that element. Element $m^T_{i,j}$ (value of the element at the position *(i, j)* in the iteration *T*) should be calculated using the expression:

$$m^T_{i,j} = f(m^T_{i-1,j}, m^{T-1}_{i+1,j}, m^T_{i,j-1}, m^{T-1}_{i,j+1})$$

```
#pragma omp transaction
{
            x1=0;
            if (r-1>=0){
                    while   (m[r-1][c].step   !=
t)
                            #pragma omp retry
                    ;
                    x1=m[r-1][c].value;
   }
                …
}
```

Fig. 7. Sample of the code from Gauss-Seidel application. In order to calculate element at position *(i, j)* in the iteration *T*, we need to wait for its adjacent elements to be in the proper iteration. That is where retry can perfectly help.

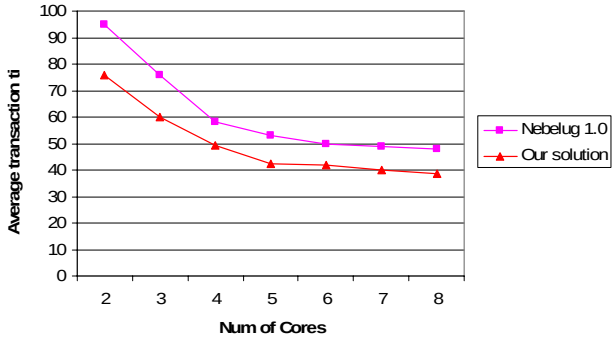


Fig. 8. Results achieved from Gauss-Seidel application. Graph presents the average transaction time (total execution time / number of transactions) on a different matrix size. Number of threads is equals to the square of the matrix dimension and each thread executes one transaction at the time.

As we can see from the above expression, in order to calculate the element at position *(i, j)* at iteration *T*, its adjacent elements needs to be in the proper iteration (data dependency). That is where we used the retry to create a nice parallel application, without any problem with

synchronization or conditional variables or locks. Figure 7 presents the part of the code of the Gauss-Seidel application and figure 8 presents the obtained results.

Again we had around 25% speedup over original Nebelung algorithm. Results obtained from Gauss-Seidel application shows that systems scales very well until the square matrix dimension 6, because number of threads raises with the square of the matrix dimension. In that moment number of threads/transactions is 36 and the number of available cores is 8 so effect of trashing begin to be dominant. But with the larger number of cores we expect that our application will continue to scale even for the bigger matrices.

D. AMMP application

AMMP is application from SPEC Benchmark 2000. We executed this application successfully. Our execution time is 10.05 minutes, while original Nebelung executes it for 14 minutes, so we have speedup of around 33%.

XII. CONCLUSION

Since we obtained very good results while evaluating our idea we truly believe that our ideas can be good starting point for development of efficient real world STM system. With our idea we are able to execute small transactions as fast as possible and for sure small transactions will be the most frequent. Our two levels and fast cleanup system will execute small transactions with minimal possible overhead. Our second level will handle big transactions and allow then to be executed as well. They will be executed slower but since they are not so frequent, it will not slow down the overall system in general. For all of those reasons we think that our solution has good potential and deserves future research and development.

REFERENCES

[1] J. Larus and R. Rajwar, “Transactional Memory”, Morgan Claypool, 2006.

[2] OpenMP Architecture Review Board, OpenMP Application Program Interface, May 2005.

[3] T. Harris, M. Plesko, A. Shinnar and D. Tarditi, “Optimizing Memory Transactions”, PLDI ’06: ACM SIGPLAN 2006 Conference on Programming Language Design and Implementation, June 2006.

[4] E. Allen, D. Chase, V. Luchangco, J.-W. Maessen, S. Ryu, G.L.Steele Jr. and S. Tobin-Hochstadt. The Fortress Language Specification. Sun Microsystems, 2005.

[5] P. Charles, C. Grothoff, V. Saraswat, C. Donawa, A. Kielstra, K. Ebcioglu, C. von Praun and V. Sarkar. X10: an Object-oriented approach to non-uniform Cluster Computing. In Proceedings of the 20th Annual ACM SIGPLAN Conference on Object-oriented Programming Systems Languages and Applications (OOPSLA), pages 519-538, New York USA, 2005.

[6] Cray. Chapel Specification. February 2005.

[7] N. Shavit and D. Touitou, "Software Transactional Memory", Proceedings of the 14th Annual ACM Symposium on Principles of Distributed Computing, 1995, pp. 204-213.

[8] M. Herlihy and J. Eliot B. Moss, “Transactional Memory: Architectural Support for Lock-Free Data Structures”, In Proc. of the 20th Int’l Symp. on Computer Architecture (ISCA’93), pp. 289-300, May 1993.

[9] P. Damron, A. Fedorova, Y. Lev, V. Luchangco, M. Moir and D. Nussbaum, "Hybid Transactional Memory", Proceedings of the Twelfth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), October 2006.

[10] S. Kumar, M. Chu, C. J. Hughes, P. Kundu, A. Nguyen, “Hybrid Transactional Memory”, In the Proceedings of ACM Symp. on Principles and Practice of Parallel Programming , March 2006.

[11] B. Saha, A. Adl-Tabatabai, Q. Jacobson. “Architectural Support for Software Transactional Memory”, 39th International Symposium on Microarchitecture (MICRO), 2006.

[12] A. Shriraman, V. J. Marathe, S. Dwarkadas, M. L. Scott, D. Eisenstat, C. Heriot, W. N. Scherer III, and M. F. Spear. “Hardware Acceleration of Software Transactional Memory”, TRANSACT 2006.

[13] J. Balart, A. Duran, M. Gonzàlez, X. Martorell, E. Ayguadé and J. Labarta. “Nanos Mercurium: A Research Compiler for OpenMP”. European Workshop on OpenMP (EWOMP'04). Pp. 103-109. Stockholm, Sweden. October 2004

[14] X. Martorell, E. Ayguadé, N. Navarro, J. Corbalan, M. Gonzalez and J. Labarta, “Thread Fork/join Techniques for Multi-level Parallelism Exploitation in NUMA Multiprocessors”. 13th International Conference on Supercomputing (ICS’99), Rhodes (Greece). June 1999

[15] M. Milovanović, O. S. Unsal, A. Cristal, S. Stipić, F. Zyulkyarov and M. Valero, “Compile time support for using Transactional Memory in C/C++ applications”, 11th Annual Workshop on the Interaction between Compilers and Computer Architecture INTERACT-11, Phoenix, Arizona, February 2007.

[16] W. Baek, C.-C. Minh, M. Trautmann, C. Kozyrakis and K. Olukotun, “The OpenTM Transactional Application Programming Interface”. In *Proc. 16th International Conference on Parallel Architectures and Compilation Techniques (PACT’07)*. Romania, September 2007.

[17] M. Milovanović, R. Ferrer, O. Unsal, A. Cristal, E. Ayguadé, J. Labarta and M. Valero, “Transactional Memory and OpenMP”. In *the Proceedings of the Intl. Workshop on OpenMP*, Beijing/China, June 2007.

[18] Gauss-Seidel Finite Difference Method for solving the linear system of equations. Online material available at: http://www-unix.mcs.anl.gov/dbpp/text/node17.html#SECTION02330000000000000000, August 27, 2007.

[19] D. Dice, O. Shalev and N. Shavit, “Transactional Locking II”. Online material available at: http://research.sun.com/scalable/pubs/DISC2006.pdf

[20] Rochester Software Transactional Memory. Online material available at: http://www.cs.rochester.edu/research/synchronization/rstm/

[21] Intel C++ STM Compiler. Online material available at: http://softwarecommunity.intel.com/articles/eng/1460.htm

Sudoku solver: Proposed solution and algorithms overview

M. Jevremović

Abstract – This paper proposes algorithm for rapid solving of Sudoku and other problems of the same class. Algorithm includes both heuristic-based logic for straight-forward reasoning and advanced state-space search in case of ambiguities. Solution is presented with respect to some other algorithms which can be utilized for solving Sudoku.

I. INTRODUCTION

Like many other game problems, Sudoku is a perfect testing ground for different kinds of algorithms. Unlike magic square, which is an arithmetical problem, Sudoku can be extended to any given set of elements and associated rules of relationship among them, for instance, Sudoku can be described as an "exact cover" problem. This fact increases usefulness of Sudoku solving algorithms on wider assortment of problems.

Some comparison is given among different algorithms which can be used for solving Sudoku. Also, for further comparison, results of the proposed solution are given for a few difficulty classes of puzzles.

II. ALGORITHMS

First, something should be said about complexity of Sudoku. It has been established that there are exactly 6,670,903,752,021,072,936,960 final Sudoku grids (calculated by Bertram Felgenhauer and Frazer Jarvis). If we count in some symmetries such as rotation, reflection and relabeling, we end up with 5,472,730,538 final grids (shown by Ed Russell and Frazer Jarvis).

There are three classes of algorithms for solving Sudoku. Those are brute-force solvers, rapid-style solvers and human solving techniques. Brute-force algorithms use one of uninformed search strategies, often with included database of final grids. These solutions are easy to program as no knowledge of problem solving logic is required, but generally require a lot of memory and time to compute the solution. Rapid-style solvers use one of many developed metaheuristic algorithms, such as hill-climbing, simulated annealing, genetic algorithm and other. Human-style solvers use more or less the same logic humans use while solving Sudoku, adopted for computer realization. More on Sudoku algorithmics can be found at [3].

M. Jevremović is with the Department of Signals and Systems, Faculty of Electronic Engineering, University of Belgrade, bul. Kralja Aleksandra 73, 11120 Belgrade, Serbia, E-mail: marko_jevremovic@yahoo.com

A. Human logic approach

As it is a big part of the proposed solution, more should be said on human strategies for solving Sudoku [5]. It can be represented as a combination of three processes: scanning, marking up and analyzing.

Scanning is performed throughout the solution. It consists of two techniques: cross-hatching and counting. The scanning of rows to identify which line in a region may contain a certain numeral by a process of elimination is called cross-hatching. It is repeated with columns, and needs to be performed systematically for all the digits 1-9. Counting represents the basic rule of Sudoku - in every row it contains all the digits 1-9, the same for columns and boxes. It may be possible to reduce cell's options to a single entry by eliminating all the numbers in a row, column and box that contain observed cell.

Marking-up is a method that helps people through analysis. While performing scanning, one may find that the observed cell's options have been reduced to 2 or 3 possible digits. In that case, those digits are noted in the cell as subscripts or in dot notation [2].

There are two approaches in the analysis: candidate elimination and "what-if" technique. When there is the same group of 2 or 3 digits marked-up in 2 or 3 cells, respectively, in the same scope (row, column, box) it is said that those cells are matched. Although there is an ambiguity in deciding single value for those cells, it also means that placing those numbers anywhere else within same scope leads to a conflict. If matched cells belong to 2 scopes it helps reducing options in some cells, possibly to a single value.

When logic can no longer contribute to further conclusions, only thing that can be done is to choose one cell, preferably with its options reduced to 2, chose one of possible numbers and see where it leads us. The chosen number is either correct and leads to the solution, or it is a wrong number and leads to a conflict. In that case, the solution process is taken one step back and another number is chosen. This "what-if" process is called backtracking or "Ariadne's thread".

B. Ariadne's thread

This approach got its name from an ancient Greek myth of the Labyrinth. In the story, young hero Theseus, finds his way out of the labyrinth using thread which King Minos' daughter Ariadne gave him. Every time Theseus

found himself in a dead-end, all he had to do was to go back up the thread to the first crossway and choose another path, until he found exit.

At the first glance, it is obvious that this approach is not the perfect one for Sudoku. Labyrinths typically have only small number of paths that lead to exit (goal), while Sudoku always has many paths that lead to goal. In a "what-if" phase before choosing one of the possible digits in cell, we should, obviously, choose a cell. Cells are usually chosen from a set of cells that contain 2 options. If we choose a cell with coordinates (x1,y1) and its value v1, that can lead to a dead-end. Then we should go back one step and choose value v2, which leads to the solution, but imagine that we chose different cell (x2,y2) and its value v1. It leads to a dead-end. Then we choose value v2 and it leads too, to a dead-end. Then we have to choose another cell (x3,y3) and repeat the process. The fact that we need to search through cells to find the one that has the possibility to lead us to the solution is a huge setback in extremely difficult cases that require many consecutive guesses.

III. Proposed solution

In the proposed solution the problem is defined a little differently than when humans do when solving Sudoku. Humans fill blank cells either by marking possible values or entering single values; here each cell contains set of digits. Cells with starting values contain singleton sets whose only element is cell's value, other cells contain sets of all nine digits. This way all the steps that humans employ (except "what-if" phase) are translated to simple operations on sets.

The first phase of operations is to choose one cell, and if it contains single value remove it from all the other sets in that row, column and box. This is repeated until the whole grid is scanned. Next phase is to choose one cell that contains more than one element and calculate the difference between that set and union of all the remaining sets in the scope. If resulting set contains a single value, that value is written in selected cell, else there is no certain conclusion and no change on selected cell's set is made. This is repeated for every scope (row, column, box) separately, and is repeated until the whole grid is scanned. These two steps are repeated until no new single value is obtained.

The first phase is equal to counting and marking-up, and the second phase is equal to cross-hatching and candidate elimination. Example of cross-hatching is given in figure 1.

In the figure 1. (left) each dot in the cell represents one digit, based on its position (as shown in fig. 1 /right/). Green dots represent digits that belong in the cell. Using red pencil seven cells are cross-hatched for digit 5, one contains the single value (6), so the only possible position for digit 5 is the seventh cell in that box. Now, let's look at the dot notation that represents set values. The only cell that contains a 5 is the seventh cell. If we make union of all the remaining sets (including singleton set in eighth cell), we will end up with set that contains digits {1,2,3,4,6,7,8,9}. Set in the seventh cell contains digits {1,3,4,5,7}. The difference between these sets is {1,3,4,5,7}/{1,2,3,4,6,7,8,9}={5} singleton set. Same operation works with candidate elimination, because cross-hatching is actually a special case of candidate elimination.

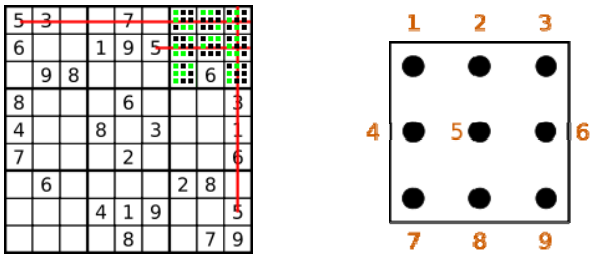


Fig. 1. Phase 2 example using dot notation (left); dot notation (right)

When no new single value can be obtained using logic, "what-if" step must be employed. Problems with the "Ariadne's thread" algorithm are already studied in the previous section. Utilizing backtracking in this approach is rather straight-forward; we choose one cell with two options, send its content to stack, choose one value (v1) and set it as the cell's single value and then call recursively the function with this algorithm using grid modified in this way. If algorithm finds a conflict, it goes one step back and value v2 is set as the cell's value, repeating recursive call. If algorithm finds yet another conflict, it goes one step back, resets the cell's content using value from stack, and then chooses a different cell, repeating the whole process until it finds the solution. This is extremely inefficient approach in cases where there are many consecutive recursions.

Modification proposed here is to choose only one cell on each level of recursion. It is already stated that this would lead to incompleteness of algorithm, but that can be easily corrected by widening search space for the next level of recursion. Before passing modified grid to the next level, all the cells containing more than one value are expanded to a full set of all 9 digits. Algorithm logic will deal with most of this expanded state-space in one pass, but this way if cell (x1,y1) is chosen before recursive call, and its value v1 leads to a conflict, then value v2 **must** lead to the solution, and there will be no need at all for choosing a different cell. This is true for the first level. If, on deeper levels, both values, v1 and v2, lead to a conflict, that means that value chosen in the previous step is wrong, not that we have to choose a different cell on the same level. Algorithm becomes complete even if we choose only one cell per level of recursion, thus saving time.

Conflict is easily detected. Before recursive call, algorithm checks for empty sets. If there is at least one empty set in the grid, there is a conflict. If program ends

and some cells are blank (which correspond to an empty set), it means that no solution exists for given starting numbers.

Both algorithms, backtracking and modified backtracking are utilized in the program, for comparison. For most of the puzzles, even very hard ones, they find solution in about the same time. The difference is evident in extremely hard puzzles [4] such is the one in the figure 2. Modified algorithm has found the solution in 17 seconds, which is about average time for this example. Backtracking algorithm worked for more than 4 hours, on several attempts, without finding the solution. On the forum [4], it is stated that the best time for solving this example using hill-climbing type solver is "some hours". In other examples the proposed algorithm found solution for even less time, while backtracking spent hours solving some of them (all times mentioned are mean times on at least 4 trials for backtracking and at least 10 trials for modified algorithm).

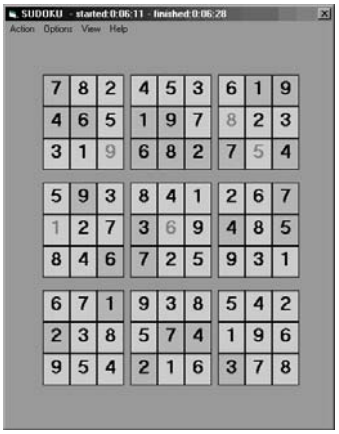


Fig. 2. finished puzzle (starting values are in dark gray boxes, guessed values are light gray)

IV. CONCLUSION

From above noted it is evident that proposed algorithm is superior to algorithms with backtracking, and even some metaheuristic algorithms. Examples have been graded based on mean solving time of both algorithms. Choosing cell before recursive call is randomized. Other methods exist for grading puzzles, such as average number of guesses in solution.

The proposed algorithm is an attempt to create optimal way of solving Sudoku class problems. As it is written for this problem specifically, it should be faster than metaheuristic algorithms, and it has already proven to be faster than human-based logic algorithms.

REFERENCES

[1] S. Russel and P. Norvig, "Artificial Intelligence: A modern approach", second edition, Pearson Education, Inc., 2003.
[2] http://en.wikipedia.org/wiki/Sudoku
[3] http://en.wikipedia.org/wiki/Algorithmics_of_sudoku
[4] Sudoku Programmers Forum Index -> Solving sudoku
[5] M. Mepham, "Solving Sudoku", Crosswords Ltd. 2005.

ECG Holter: Basic steps in development

M. Jevremović, S. Marković

Abstract - This paper depicts basic steps in the development of an ECG holter. PDA device is chosen as a platform for holter's hardware and MATLAB is utilized to develop the software. Software testing was conducted on data sets from MIT/BIH database.

I. INTRODUCTION

The idea behind the project is to approach the vast problem of developing a new ECG holter in its basic form, disregarding end-user interface and other commercial norms. The focus is set on creating a working prototype which can later become the starting point in developing a stand-alone ECG holter.

The project has been split into two separate problems which were later solved individually. One part of the problem is connecting hardware into one device whose purpose is to collect ECG data from subject. It also has to be small and portable. The main part of hardware is a PDA device with an A/D card. It is small enough, powered by batteries and capable of storing relatively large amount of data.

The other problem is to develop reliable software for ECG data processing which can be utilized in ECG holter. It should be fast and calculate several most important parameters for machine decision making and for human doctors.

II. HARDWARE

The hardware consists of PDA Acer n30, A/D converter NI USB-6008 and ECG amplifier. Amplifier is made with two separate power sources; one is for amplifier circuit and the other for exit circuit for galvanic separation. Amplifier alone is built with integrated circuit INA 114P, with input via RC bridge. That way we have both input HP filter that does not pass DC component and input resistance is large enough not to strain the signal source.

A/D converter is configured to accept signal between channel 1 and ground. Single-ended acquisition of data has proved to have much better signal to noise ratio than differential acquisition (input is connected between two channels).

M. Jevremović and S. Marković are with the Department of Signals and Systems, Faculty of Electronic Engineering, University of Belgrade, bul. Kralja Aleksandra 73, 11120 Belgrade, Serbia, E-mail: marko_jevremovic@yahoo.com

Typical output range of amplifier is -5 V to 5 V. A/D card is configured to input range of -10 V to 10 V. This way a compromise is made between signal resolution and possibility of having larger oscillations on the amplifier output.

Operating system of PDA device is unreliable and has often demanded complete reset when program for data acquisition froze. Device processor is powerful enough to maintain data acquisition and signal processing in real-time only if data is not displayed at the same time. Displaying data requires a significant amount of memory and processor time.

III. ACQUISITION SOFTWARE

For purpose of developing this software LABVIEW 8 package was used. The program is converted to PDA executable format. Two versions of the program were compiled, one with sample frequency of 200 Hz, and another with a 500 Hz sample rate. Continuous data acquisition, saving data and real-time displaying simultaneously is possible only with 200 Hz sample rate. In the other version of the program, type of acquisition is set to finite buffer. Downside of finite buffer in comparison with continuous buffer is limited working time, but if memory of PDA device is small, this can be a very elegant trick that can save PDA from memory overflow. Another glitch is observed while using finite buffer technique – base line of signal is shifted for constant amount. The reason for this behavior is left unknown, simply because solving it would bring no difference to the next step, data processing.

IV. DATA PROCESSING SOFTWARE

The central peace of data processing software is the function written in MATLAB that enables upgrade to a program that can process signal both in real-time and offline. In the case of real-time processing some latency is inherent, and depends on the size of buffered signal sent to the main function. The function calculates heart rate, exact position of R peaks, and position of the beginning and the end of QT intervals.

The position of R peaks is calculated first, by filtering signal and processing its first derivate. Sometimes, QRS wave is too small for function to detect it, so after the initial detection all RR intervals are checked for inconsistencies. If one of the detected RR intervals is outside statistical expectation centered on mean heart rate,

it is further analyzed. The function can then detect those small QRS waves and mark them for expert to analyze them later. Example is given in figure 2.

The location of the beginning of QT wave is estimated based on the location of R peak. The next step is to filter the signal with MA filter. Filtered signal will have QT wave that looks more like a sinusoid (signal is shaped in the form shown in figure 1b). Signal is then windowed to the period where the end of QT wave is expected (central half of RR interval). The end of QT wave is then placed in the location of the largest negative derivate of the windowed signal. The function makes mistakes when confronted with pathological signals, but that can be easily detected, and marked, from inconsistent statistics of calculated parameters.

The function can also make a mistake in case of a "dropped" T wave. Example is shown in the figure 4. Because of different topology of QT wave, the function can make slight error in deciding position of the end of QT wave.

The function is tested on 20 different ECG datasets from MIT/BIH database, among which are normal datasets, signals with different levels of noise, signals with base-line wander and arrhythmias.

Heart rate, RR intervals statistic, QT length statistic and detection of pathological QRS waves are all parameters that can be further used in the machine decision making process. The main function has also an option of drawing heart variability (figure 6), which represents averaged RR intervals centered on R peak. This is the only information that is useful strictly to specialist doctors, and is included to help in quick clinical analysis of acquired ECG data.

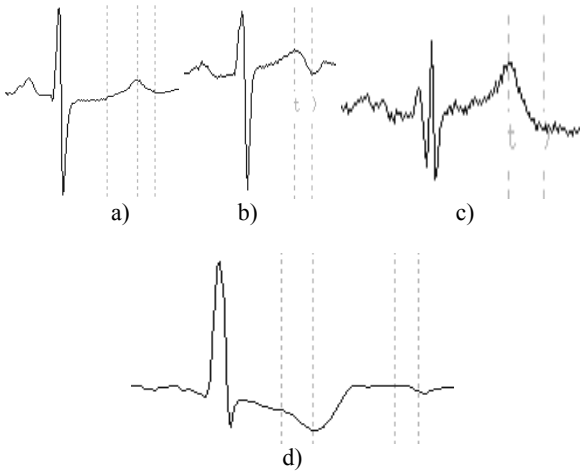


Fig.1. Four types of RR intervals

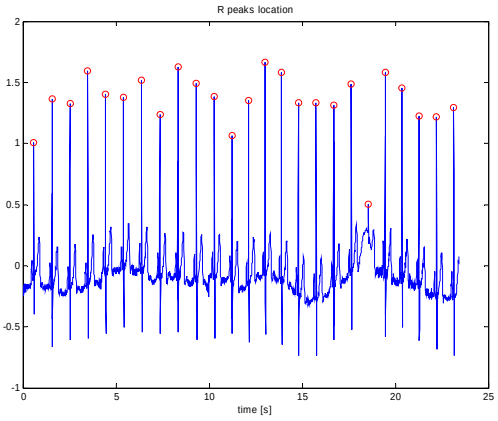


Fig. 2. Example of irregular single R peak

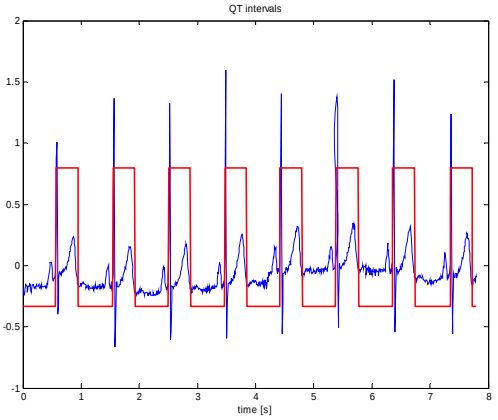


Fig. 3. Example of well defined QT intervals

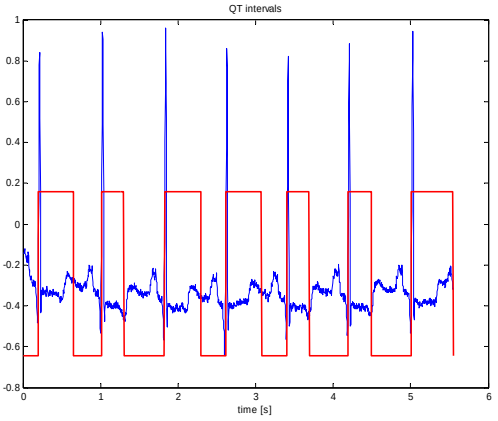


Fig. 4. Example of error made while defining QT intervals

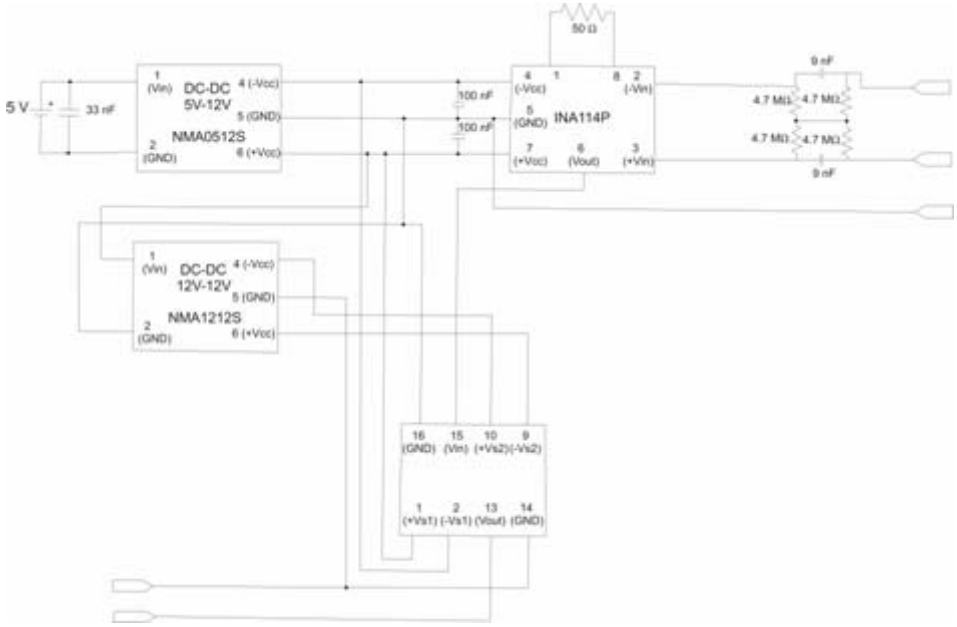


Fig. 5. Input amplifier circuit

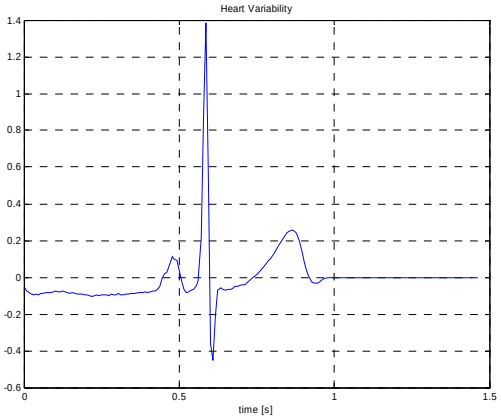


Fig. 6. Heart Variability

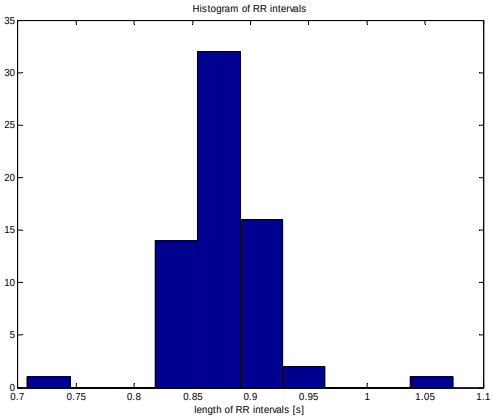


Fig. 8. Histogram of RR intervals

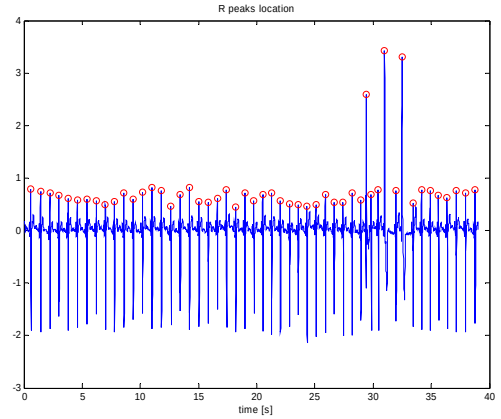


Fig. 7. Example of short-time irregular heart beats

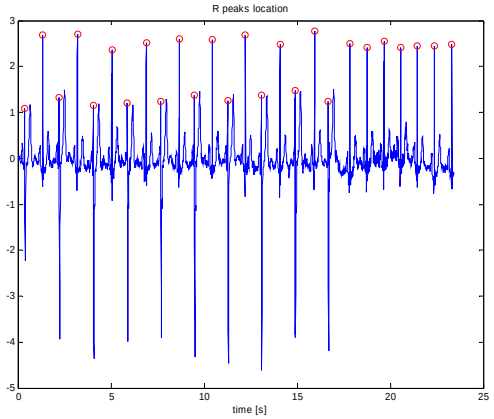


Fig. 9. Example of period of irregular heart beats

V. CONCLUSION

The hardware was successfully connected and, with the exception of slightly unreliable operating system of PDA, worked satisfyingly well. For integrating software for real-time analysis a PDA device with more processing power is needed.

There are several examples of results of processing software. It has been written with minimum filters needed (only MA filters were used), so it can easily be implemented in any programming language and any processor.

There are many ways this project can be upgraded. Neural network can be constructed, for instance, using statistics provided by signal processing function as input, which could detect an irregular ECG and save it for clinical analysis.

It is the authors' hope that this write-up has successfully presented the first few basic steps in the development of an ECG holter, and has pointed to some of the problems that can be encountered during such project.

REFERENCES

[1] Emil Jovanov, Pedro Galabert, Reza Adhami, Bryan Wheelock and Robert Adams, "RealTime Holter Monitoring of Biomedical Signals", DSP Technology and Education Conference DSPS'99, 1999.

[2] R. Jané, A. Blasi, J. García and P. Laguna, "Evaluation of an Automatic Threshold Based Detector of Waveform Limits in Holter ECG with the QT database", Computers in Cardiology, 1997.

[3] PhysioTour PhysioBank.htm, tutorial on MIT's database of ECG and other electrophysiological signals

[4] D. Popovic, *Acquisition of electrophysiological signals*, lectures, 2007.

Дигитална Стеганографија

М. Милошевић, С. Јанковић

Abstract – In this manuscript is presented the basics of steganographic method of embedding text and digital photography in another digital photography using the method of least significant bit. Firstly, text was encrypted using asymmetric key cryptography. The least significant bit of every pixel of the "innocent" image was used to embed text inside digital photography. Four least significant bits of every pixel of the "innocent" image were used to embed image inside another image, so it was possible to hide secret image inside the image of the same size. The results in this study indicate that these methods could be successfully used for hiding secret data inside digital images. It would degraded the "innocent" image, but changes could be noticed only if is possessed the original of "innocent" image.

Садржај – У овом раду изложени су основни принципи стеганографске методе сакривања текста и слике у дигиталну слику коришћењем најмање значајног бита. Текст се прво шифрује методом криптографије асиметричног кључа. За сакривање шифрованог текста користи се метод бита најмање тежине који се примењује на сваки пиксел фасадне слике. За сакривање слике, користе се четири последња бита сваког пиксела фасадне слике, тако да је могуће сакрити тајну слику унутар фасадне слике исте величине. Резултати показују да се ове методе успешно могу користити за сакривање информација унутар дигиталних слика. Сакривањем долази до деградације фасадног фајла, али је промена приметна само ако се поседује оригинал.

I. Увод

Сакривање поверљивих информација одувек је било у интересу људи који их поседују. Временом, технике сакривања су се развијале, а са појавом рачунара и дигитализацијом информација доживеле су праву експанзију. Два основна правца развоја били су стеганографија и шифровање (криптографија).

Стеганографија (грч. steganós – покривен; graphía – писање) је наука о сакривању порука унутар других тако да је само постојање тајне информације неприметно [1]. Стеганографске методе прикривања су коришћење невидљивог мастила, читање сваког другог реда у тексту, читање последњег слова у свакој речи, а савременије методе између осталог подразумевају смештање порука у дигиталне фотографије или звучне фајлове и уметање слика у видео фајлове. Стеганографијом се сакрива постојање поруке, али самим откривањем где је порука сакривена открива се и њено значење.

Криптографија је наука о састављању кодираних или заштићених података тако да они могу бити разумљиви само ономе ко поседује шифру.

M. Milošević, S. Janković are students of Faculty of Electrical Engineering, University of Belgrade, bul Kralja Aleksandra 73, 11000 Belgrade, Serbia,
E-mail: milana_milosevic@yahoo.com

Заснива се на постојању алгоритма за шифровање и кључа (шифре) која се користи за 'закључавање' или 'откључавање' датог података. Најпознатије криптографске методе су анаграм, моноалфабетско шифровање, Цезарова шифра, и криптографија симетричног и антисиметричног кључа [2].

Свака, па и поверљива, информација има свој рок трајања, односно време у току којег се може (зло)употребити. Када информација није више актуелна ни њена тајност није од значаја, па је заштиту информација потребно обезбедити у току неког одређеног и коначног временског периода. Зато није од пресудног значаја да сакривене информације буде немогуће открити, већ је битно да се оне не могу открити довољно дуго времена. Један од начина да се постигне потпунија и дуготрајнија заштита података је комбиновање стеганографије са криптографијом, односно сакривање шифроване информације.

У овом раду изложени су основни принципи стеганографске методе сакривања текста и слике у дигиталну слику коришћењем најмање значајног бита и криптографије антисиметричног јавног кључа које имају велику комерцијалну примену у сфери интернет пословања: приликом интернет куповине, за проверу аутентичности садржаја нађеног на интернету, за прављење бележака о коришћењу одређених сервиса, провери дигиталног потписа, итд. Стеганографија се такође примењује и код ласерских штампача у боји и водознацима, за спречавање фалсификовања новца.

II. Алгоритам Заштите Података

A. Стеганографија

Стеганографска метода прикривања информација у дигиталној слици заснива се на измени дигиталног записа фасадне слике на тај начин да је промена што мање приметна. Простор за овакве модификације постоји услед ограничења људског вида, али и несавршености у фасадним фотографијама које су последица несавршености фотографске опреме. Деградација фотографије може бити последица недостатака сочива и других оптичких елемената, квалитета фото-сензора и шума који уносе електронске компоненте у дигиталним фотоапаратима [3].

Дигитална слика је бинарни фајл који садржи бинарну репрезентацију боје или интензитета светлости за сваки елемент слике (пиксел) који она обухвата. Слике могу бити приказане као комбинација нијанси сиве боје (*grayscale*) или могу бити у боји. Ако је слика приказана у нијансама сиве боје, за чување информације о сваком пикселу користи се 8 битова, односно 1 бајт. Бројне вредности за нијансу сиве боје за сваки пиксел су у опсегу од 0-255 где 0 представља црну, а 255 белу боју. Ако је слика у боји, онда се

користи 24-битна шема боја. Сваки пиксел је репрезентован са 24 бита (3 бајта), при чему сваки бајт представља вредност интензитета једне од три основне рачунарске боје: црвене, зелене или плаве (RGB). Бројне вредности интензитета сваке од боја су у опсегу 0-255. Што је бројна вредност већа, то је та боја интензивнија. Бела боја се кодира са (R, G, B) = (255, 255, 255), а црна са (R, G, B) = (0, 0, 0).

Постоји више поступака за стеганографско сакривање информација. Најприступачнији је коришћење методе најмање значајног бита (*Least Significant Bit*, LSB). Суштина ове методе је да се модификацијом последњег бита (бита најмање тежине) сваког пиксела може сакрити било која дигитална информација унутар носеће (фасадне) слике. Пошто је за кодирање ASCII карактера потребно 8 битова (опсег вредности ASCII карактера је баш 0-255), груписањем осам по осам пиксела и коришћењем LSB-а сваког од њих може се кодирати текст. Променом вредности LSB-а, мења се и оригиналан запис слике. Међутим, та промена је еквивалентна промени од највише 1, од могућих 256 нијанси, што је неприметно за људско око, а самим тим и занемарљиво.

Ако се текст сакрива унутар црно-беле слике, могуће је користити 8 пиксела за сакривање једног карактера. Ако се ради са сликом у боји, један пиксел се може искористити за сакривање три бита (један бит по боји), тако да је потребно само 3 пиксела за сакривање једног карактера. Промене које настају на фасадној слици у боји су мање приметне, јер се интензитет боје мења за једну вредност од могућих 256^3 (све нијансе свих боја).

За сакривање слике у слици могу се користити разне методе, али је најприступачнија модификована метода LSB. Ако се за сакривање информације са тајне слике користе последња четири бита (уместо последњег бита у оригиналној имплементацији), могуће је сакрити тајну слику унутар фасадне слике исте величине. Поступак се састоји у томе што се четири водећа бита (највеће тежине) сваког пиксела тајне слике умећу уместо четири последња бита (најмање тежине) одговарајућег пиксела фасадне слике. Оправданост коришћења ове методе је у томе што се овако вредност интензитета неког пиксела може променити за највише 16 (од могућих 256) нијанси, што у зависности од избора фасадне и тајне слике не мора бити очигледна промена. Тајна слика се реконструише тако што се из фасадне слике екстрахују четири последња бита, која сада представљају бите највеће тежине и допуне са четири произвољна бита на местима мање тежине. Очигледно је да тајна слика трпи деградацију у односу на оригиналну тајну слику, услед трајног губитка 4 бита мање тежине, али и упркос деградацији садржај може бити видљив.

За поређење оригиналне и новодобијене фасадне слике користи се коефицијент кроскорелације:

$$C = \frac{\sum_x \sum_y [A(x,y) - \overline{A}] \cdot [B(x,y) - \overline{B}]}{\sqrt{\sum_x \sum_y [A(x,y) - \overline{A}]^2} \sqrt{\sum_x \sum_y [B(x,y) - \overline{B}]^2}} \qquad (1)$$

где је: (x, y) позиција пиксела на слици, $A(x, y)$ и $B(x, y)$ су интензитети боје пиксела на позицији (x, y) за прву и другу слику, респективно, а $\overline{A}$ и $\overline{B}$ средње вредности интензитета боје пиксела на првој и другој слици респективно. Вредности коефицијента кроскорелације су у опсегу од 0 до 1, при чему вредност блиска јединици означава малу промену квалитета слике. Коефицијент кроскорелације одређује се за сваки канал појединачно, па се за *црно-беле* слике добија само један коефицијент, док се за RGB слике добијају три коефицијента.

Б. Криптографија анитсиметричног кључа

Дуго је била примењивана криптографија симетричног кључа, односно шифровање и дешифровање уз помоћ истог кључа. Недостатак ове методе је што су обе стране морале да имају кључ, што је отварало проблем тајности размене кључа. Данас комерцијално најзаступљенији модел шифровања је асиметрично шифровање, односно шифровање путем јавног кључа. Свако може шифровати јавним кључем, али се дешифровање може извести једино уз помоћ тајног кључа.

Један од најпознатијих алгоритама за шифровање путем јавног кључа је RSA алгоритам (име по ауторима: R.L. Rivest, A. Shamir, and L.M. Adleman) . У основи овог алгоритма леже достигнућа теорије бројева [1]. Поступак RSA алгоритма је следећи:

1. бирају се два проста броја p и q
2. формирају се бројеви $n = p \cdot q$ и $z = (p–1)\cdot(q–1)$
3. бира се број m такав да је узајамно прост са z; тражи се одговарајући број d такав да је $m \cdot d$ узајамно просто са z
4. пар (n, m) је јавни кључ и служи за шифровање
5. пар (n, d) је тајни кључ и служи за дешифровање
6. шифрује се операцијом $x^m \equiv y \pmod{n}$
7. дешифрује се операцијом $y^d \equiv x \pmod{n}$.

при чему је $a \equiv b \pmod{c}$ ако бројеви a и b дају исти остатак при дељењу са c.

Нека треба шифровати број $x = 88$, а нека су елементи кључева следећи: $p = 17$, $q = 11$, $n = 187$, $z = 160$, $d = 23$, $m = 7$. Применом алгоритма шифровања на број x, $88^7 \equiv 11 \pmod{187}$, добија се $y = 11$. Обрнути процес, дешифровање броја y, даје $11^{23} \equiv 88 \pmod{187}$, односно $x = 88$.

Осим за шифровање, овај алгоритам се може употребити за проверу аутентичности података. Тако, на пример, дигитални потпис заправо представља тајни кључ *d1*, особе која шаље поруку, при чему је познат њен јавни кључ (*n1*,*m1*). Прималац поруке лако може проверити да ли послати потпис задовољава поменуте услове за јавне и тајне кључеве RSA алгоритма. На овај начин прималац не само да се уверава о пореклу поруке већ и о томе да ли је порука пресретана или можда оштећена у путу.

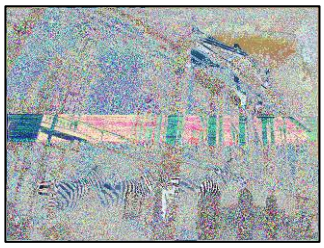
III. Резултати И Дискусија

На бази алгоритама за стеганографско сакривање, описаних у претходном поглављу, формиран је код у програмском пакету Matlab. Програм је тестиран за сакривање текста у црно белој фасадној слици и у једном каналу фасадне слике у боји, као и за различите комбинације сакривања дигиталне слике: (1) сакривање црно-беле слике у црно-белој фасадној слици исте величине, (2) сакривање RGB слике у RGB фасадној слици исте величине и (3) сакривање смањене слике (360 × 270 пиксела) у једном каналу фасадне слике (640 × 480 пиксела).

Сл1.- Оргинална фасадна слика

Сл.2.- Тајна слика

Сл.3.-Фасадна слика са тајном сликом



Сл.4. - Разлика између Сл.1. и Сл.3.

На слици 1 приказана је фасадна слика величине 640 × 480, а на слици 2 тајна слика исте величине [4]. Текст који се сакрива је шифровани текст овог рада. Вредности коефицијента кроскорелације за све наведене тест комбинације дате су у табели 1.

ТАБЕЛА I.
КОЕФИЦИЈЕНТИ КРОСКОРЕЛАЦИЈЕ ЗА ОРИГИНАЛНУ И СТЕГАНОГРАФСКИ МОДИФИКОВАНУ ФАСАДНУ СЛИКУ

црно-бела фасадна слика	коефицијент кроскорелације C		
текст	0.999977		
црно-бела слика	0.995837		
RGB фасадна слика	C(R)	C(G)	C(B)
текст	0.999975	1.000000	1.000000
RGB слика исте величине као фасадна слика	0.994926	0.996154	0.997343
смањена RGB слика у једном каналу RGB фасадне слике	0.997623	1.000000	1.000000

Коефицијент кроскорелације за убацивање текста у црно-белу дигиталну фотографију је 0.999977 што представља деградацију слике мању од 0.03‰. Када је фасадна слика у боји, при убацивању текста у један канал слике, до деградације долази само у том каналу, а вредности су сличне као и када је за фасадну слику употребљена црно-бела фотографија. Овако висока вредност је очекивана с обзиром да се за сакривање текста користи само последњи бит сваког пиксела фасадне слике. Када се за фасадну слику користи фотографија у боји и то само један канал овај коефицијент је још ближи јединици. Како је у случају RGB фасадне слике коефицијент корелације дефинисан само за један канал, са аспекта алгоритма заправо не постоји разлика у деградацији фасадне слике била она црно-бела или у боји. Разлика у коефицијентима кроскорелације последица је статистичке расподеле битова најмање тежине за конкретну фасадну слику.

Када се у црно-белу фотографију убацује слика истог типа коефицијент је такође веома близак јединици, а деградација је мања од 4,2‰. Када се у фасадну слику у боји убацује фасадна слика у боји исте величине као и фасадна слика вредности коефицијента кроскорелације и даље остају блиске јединици, мада су нешто мање од предходних, а деградација је око 5‰, што је и очекивано јер се сада користе последња четири бита сваког канала. Када се умањена слика сакрива у само једном каналу деградација постоји само у том каналу, а вредност коефицијента кроскорелације је слична као и у осталим случајевима. Важно је приметити да коефицијент кроскорелације зависи и од тајне и од фасадне слике, односно његове вредност се израчунава за сваки пар слика.

За фасадну слику изабрана је слика пејзажа (сл.1) [4]. На овој слици постоје и униформне површине (небо) и "шарене" површине (трава) и захваљујући томе боље ће се уочити разлике у деградацији слике на појединим површинама. На униформној површини новонастале стеганографски модификоване фасадне слике (сл.3.) примећују се контуре тајне слике (сл.2.), док на "шареној" површини, за људско око разлике готово да нема. Такође, могуће је и графички приказати апсолутну разлику између интензитета сваког пиксела оригинала и новонастале фасадне слике. Пошто је максимална разлика 16 вредности, онда су све вредности помножене са 16 да би се добила слика на којој се разазнају разлике (сл.4).

При сакривању слике у другу слику на било који од описаних начина, поред деградације фасадне слике долази и до деградације тајне слике јер се неповратно губе последња четири бита сваког пискела. При сакривању текста то није случај, тако да је екстраховани текст идентичан полазном. Вредности коефицијента кроскорелације за оригиналну тајну слику и тајну слику након екстракције из фасадне (или носеће) слике дати су у табели 2.

ТАБЕЛА II.

КОЕФИЦИЈЕНТИ КРОСКОРЕЛАЦИЈЕ ЗА ОРИГИНАЛНУ ТАЈНУ СЛИКУ И ТАЈНУ СЛИКУ НАКОН ЕКСТРАКЦИЈЕ

	коефицијент кроскорелације *C*		
црно-бела тајна слика из црно-беле фасадне	0.995837		
	C(R)	*C*(G)	*C*(B)
RGB тајна слика исте величине као фасадна слика	0.994926	0.996154	0.997343

Предности стеганографске методе најмање значајног бита за сакривање текста и слика су релативно мало коришћење рачунарских ресурса, брзо извршавање и једноставна имплементација. Процес сакривања и откривања текста траје само неколико секунди, док је неколико десетина секунди потребно за слике величине 640 × 480 пиксела.

Највећа мана стеганографске методе најмање значајног бита је што мора да се води рачуна при избору фасадне слике, јер од ње зависи колико ће промене настале сакривањем информација бити видљиве. На деловима фасадне слике који су униформно обојени промене у битовима мале тежине постају видљиве. Зато је препоручљиво да се за фасадну слику узме слика са што више шара, различитих боја и контура и што мање униформно обојених површина, да би се избегла видљивост промена. Примена стеганографских метода је могућа само за формате код којих не долази до губитка информација (lossless формат), односно компресије при чувању фајла, тако да је најпогодније да слике буду у BMP формату.

Као додатна метода заштите текста који је убациван у слику коришћен је RSA алгоритам. За имплементацију RSA алгоритма генерисан је програм у програмском језику C++ који користи елементе кључева $m = 47$, $d = 23$, $p = 11$, $q = 19$, $n = 209$, и примењује се на шифровање тесктуалне датотеке. Примера ради, ако је текст који треба шифровати: „26tyd94sz1g69de6zst54dyh*td*6g“, након извршавања програма добија се низ карактера „Ž±_:f~Ž_±–mŽ:±ť\□±\Ž~`“,

Разоткривање шифре код RSA алгоритма, односно откривање тајног кључа, своди се на факторисање броја *n*, што за довољно велике вредности *n* представља дуготрајан процес који зависи од процесорске снаге рачунара који се користи. Модификације приказаног алгоритма као основу за шифровање могу користити произвољан број битова, биран са унапред дефинисаном учестаношћу или помоћу неког компликованијег алгоритма избора, што додатно отежава процес дешифровања. Модифкација овог програма може се употребљавати за шифровање датотека произвољног типа [5].

IV. ЗАКЉУЧАК

У овом раду изложени су основни принципи стеганографске методе сакривања текста и слике у дигиталну слику коришћењем најмање значајног бита. Текст се прво шифрује методом криптографије антисиметричног кључа. За сакривање шифрованог текста користи се метод бита најмање тежине који се примењује на сваки пиксел фасадне слике. За сакривање слике, користе се четири последња бита сваког пиксела фасадне слике, тако да је могуће сакрити тајну слику унутар фасадне слике исте величине. Коефицијенти кроскорелације за оригиналну фасадну слику и стеганографски модификовану фасадну слику, као и за оригиналну тајну слику и тајну слику након екстракције из фасадне слике, су блиски јединици, што значи да се за одговарајуће изабрану фасадну слику, стеганографском методом најмање значајног бита може постићи задовољавајуће мала деградација и фасадне и тајне слике, што овај алгоритам чини поузданим. Резултати показују да се стеганографија најмање значајног бита у комбинацији са криптографијом антисиметричног кључа може успешно користити за поуздано сакривање информација унутар дигиталних слика.

ЛИТЕРАТУРА

[1] Интернет, http://www.garykessler.net/library/steganography.html

[2] З. Огњановић и Н. Крцавац, *Увод у теоријско рачунарство*, Београд – Крагујевац, 2004.

[3] F.Galand and G..Kabatinsky, "Steganographi via Covering Codes," Proc. ISIT'03, Yokohama, Japan, July 2003.

[4] Интернет, http://www.cs.vu.nl/~ast/books/mos2/zebras.html

[5] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding", IEEE Trans. on Information Theory, Vol. 49. No.3, pp.563-593, March 2003.

Design and Realization of Variable Reactance Reflection Phase Shifter

V. Rinčić, N. Stojanović

Abstract – Design and practical realization of circuit that is common used in telecommunication for phase shifting are describe in this text. Principle operation, design and simulation of variable reactance reflection phase shifter are explained. At the end, after practical realization, measurement and comparing of derived and assumed results are accomplished.

I. INTRODUCTION

A phase shifter [1] is a two-port network with the provision that the phase difference between the output and the input signals may be controlled by a control signal – DC bias. Phase shifters can be analog or digital. Phase shifters are digital when the differential phase shift can be changed by only few discrete values, such as $180°$, $90°$, $45°$, $22.5°$ and $11.25°$. In analog phase shifters, the differential phase shift can be varied in a continuous manner by a corresponding continuous variation of control signal. Digital phase shifters find extensive applications in phase-array antenna systems. Phase control of the signals fed to the various elements of the array allows the direction of the radiated beam to be scanned electronically.

There are two methods for designing digital phase shifters at microwave frequencies. One is to use the properties of ferromagnetic materials for obtaining switchable phase shift. The other important design for digital phase shifters uses semiconductor devices. Phase shifters using semiconductor devices can be either of the reflection type or the transmission type. In reflection type phase shifters, the basic design unit is a one-port network, and it is the phase shift of the reflected signal that is changed by the control signal. This basic one-port phase shifters can be converted into useful two-port components either by using a circulator or a hybrid. Because of the ease integration, the hybrid type phase shifters are more common.

II. VARIABLE REACTANCE REFLECTION PHASE SHIFTER

The most common used type of reflection phase shifters are variable reactance reflection phase shifters [2]. Circuit uses a 90-degree hybrid and variable reactance (varactors). The schematic of the variable reactance reflection phase shifter is shown in Fig. 1.

V. Rinčić, N. Stojanović are with the Department of Telecommunications, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: pantla@verat.net, stoj_nik@yahoo.com.

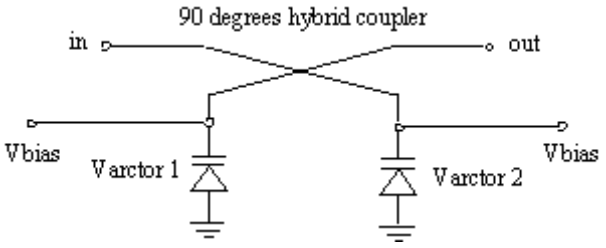


Fig. 1. Variable reactance reflection phase shifter

As hybrid coupler, the quadrature single-box branchline coupler is used here (Fig. 2. note: all 4 arms are $\lambda/4$ in length). Signal from port 1 is equally distributed to port 2 and port 3, while port 4 is isolated. The phase difference between signals at port 2 and 3 is 90 degrees (see Fig. 2) [3].

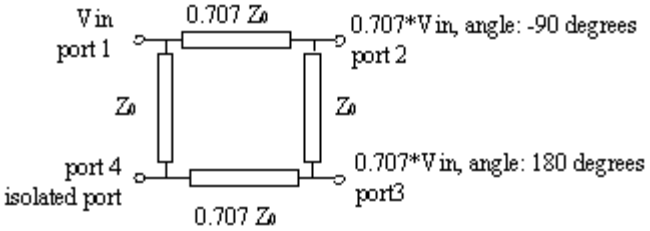


Fig. 2. Quadrature single-box branchline coupler

Ports 2 and 3 are loaded with varactors. Signal from port 1 (in), which is equally distributed to port 2 and port 3, is reflected from these reactances, and travels to port 4 (out), with phase shift.

There are two states in operating of phase shifter. In first state, when the DC bias is low, there is one value of phase shift, and when DC bias changes its value, the capacitance of varactor is changed and there is different phase shift. Difference between these two shifts represents overall phase shift.

Equation (1) shows phase shift when the DC bias has value V1, where C_{V1} is capacitance of varactor at DC bias V1. Equation (2) shows phase shift when the DC bias has value V2, and equation (3) shows overall phase shift.

$$\varphi_{V1} = 2 \cdot \arctan(\omega C_{V1} Z_0) \quad (1)$$

$$\varphi_{V1} = 2 \cdot \arctan(\omega C_{V1} Z_0) \quad (2)$$

$$\varphi_o = \varphi_{V1} - \varphi_{V2} \quad (3)$$

III. CIRCUIT DESIGN AND SIMULATION

A. Coupler design and optimization in ADS software package

Circuit design starts with hybrid design [3]. Here, the ADS software package (product of Agilent Technologies), i.e. his accessory *LineCalc* is used. Dimensions of microstrip coupler are calculated, based on: substrate thickness h = 1.5mm, substrate relative dielectric const ε = 4.34, metallization thickness t = 0.015mm, central frequency f = 1GHz, characteristic impedance $Z_0 = 50\Omega$ and effective electric length 90˚. Calculated values for line length and line width are as follows: in case where $Z_C = Z_0 = 50\Omega$ - W = 2.85723mm, L = 41.3073mm, and in case where $Z_C = 0.707 \cdot Z_0 = 0.707 \cdot 50\Omega = 35.35\Omega$ - W=4.88455mm, L = 40.2818mm.

Designed hybrid coupler is then simulated in ADS software. Fig. 3 and Fig. 4 show results of this simulation. Following parameters are shown: S_{11} (reflection coefficient), S_{21} and S_{31} (transmission coefficients), and S_{41} (insulation coefficient) changing with frequency.

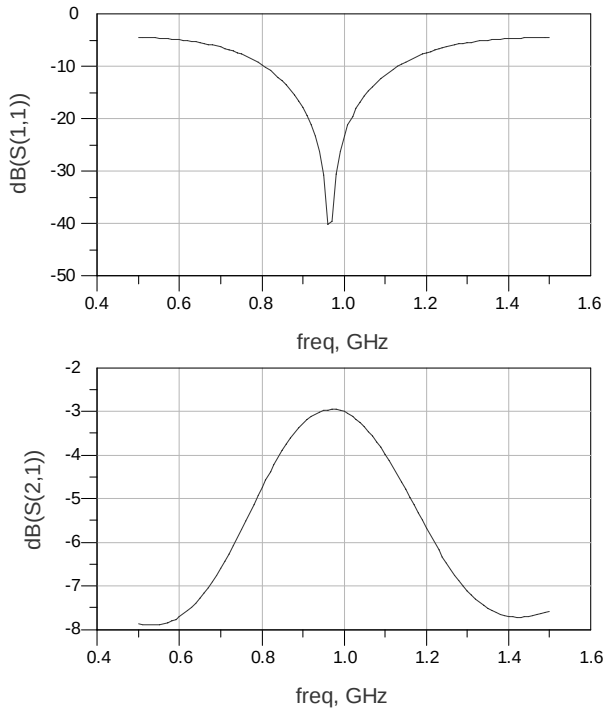


Fig. 3. Parameters S_{11} and S_{21} at f = 1GHz.

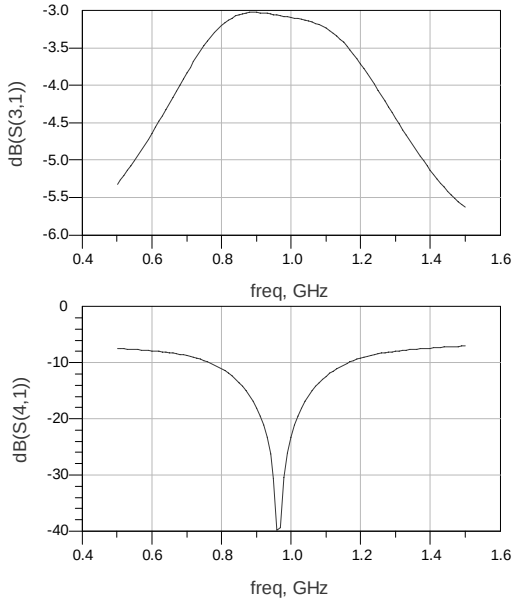


Fig. 4. Parameters S_{31} and S_{41} at f = 1GHz.

Figures show that there are no minimums of parameters S_{11} and S_{41} at frequency f = 1GHz, and maximums of S_{21} and S_{31} parameters. This means that aim of coupler usage is not achieved. Entirely diagram is shifted to left, toward lower frequencies. Thereby parameters optimization is needed. Also, line optimization is done here. The goal of optimization is that parameters S_{11} and S_{41} achieve min values at desired frequency f = 1 GHz and that parameters S_{21} and S_{31} have maximum (should not be lower than -3dB). After optimization and after updating values characteristics of coupler have shape shown in Fig. 5 and Fig. 6. Figures show that optimal coupler characteristics at desired frequency are achieved.

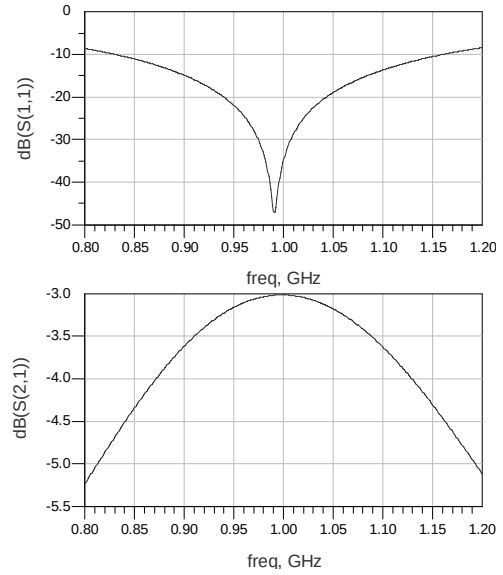


Fig. 5. Parameters S_{11} and S_{21} at f = 1GHz after optimization

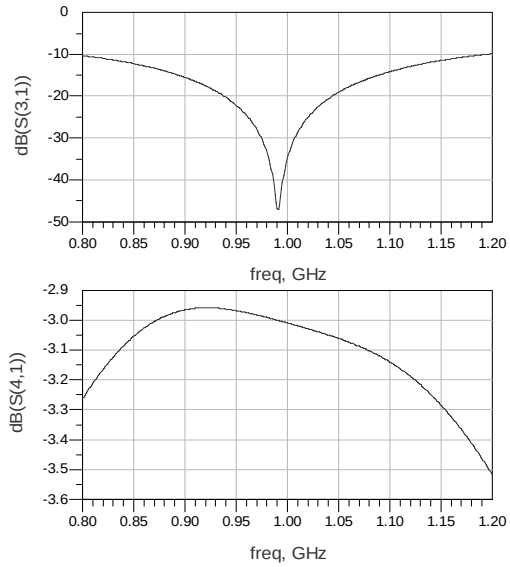


Fig. 6. Parameters S_{31} and S_{41} at f = 1GHz after optimization

B. Device simulation in ADS

When hybrid coupler was designed and optimized it is needed to simulate device work, in order to see what phase shifts are made by changing of polarization voltage of varactors. Varactors can be changed with capacitors which have values of capacitance that correspond to voltage values. It can be seen from varactor datasheet. For example, there is phase of S_{21} parameter for capacitance of 5 pF of varactor (polarization voltage 10V) shown in Fig. 7. It should be pointed out that port 4 of hybrid coupler is now denoted as port 2 – output port.

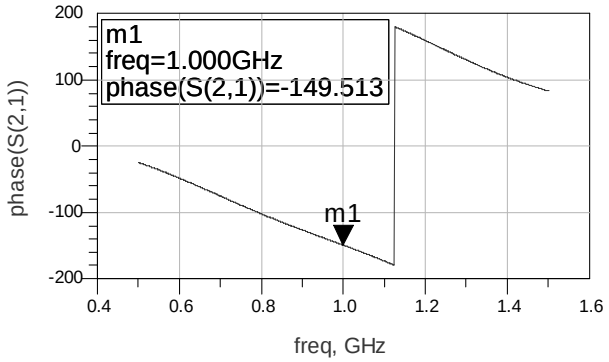


Fig. 7. Phase of output signal for Vbias = 10V

IV. Device making and performance measuring

A. Layout design

On the basis of good simulation results, layout design has started. First of all, it is necessary to create input/output ports in ADS, and so called gaps for varactors. Gaps are created on the basis of dimensions supplied from datasheet. Additionally, in Fig. 8 two extra microstrip lines after varactors can be noticed. These lines have to be large enough for drilling holes. Varactors are soldered through these holes to metallization on the other side (ground). Layout of phase shifter is shown in Fig. 8.

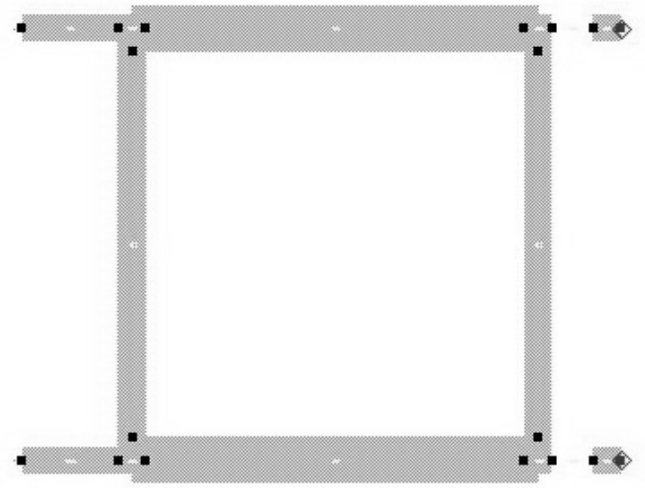


Fig. 8. Phase shifter layout

Layout is made on metallized substrate where varactors and connectors are situated on. Varactors BB105A and female F connectors are used for this project. Final shape of this device is given in Fig.9.

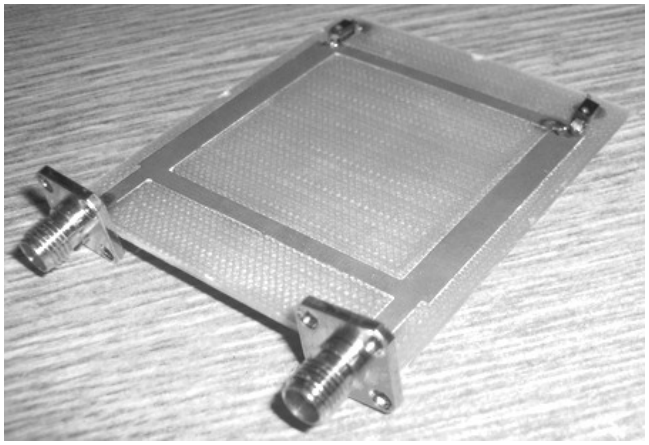


Fig. 9. Final shape of phase shifter.

B. Measuring device performance and results

Performance of phase shifter is measured by spectrum analyzer manufactured by *Hewlett-Packard* company. Dependence of phase shift from supply voltage is measured. It is needed to measure phase component of parameter S_{21} for different varactor polarization voltage

values. Fig.10 shows the way that phase shift is changing, by changing DC voltage.

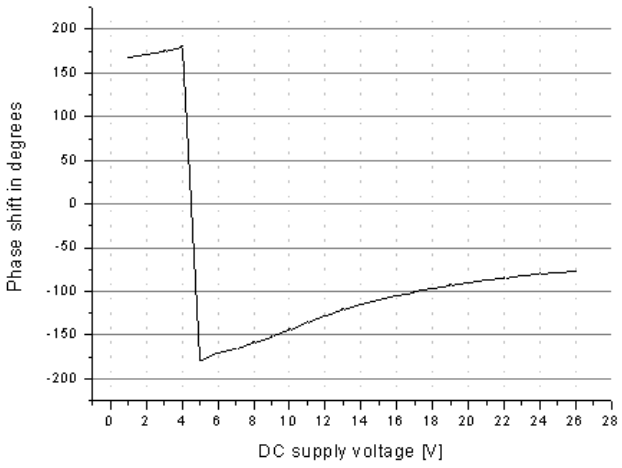


Fig. 10. Measured dependence between phase shift and DC supply voltage.

V. Conclusion

In Table 1 are shown both simulated and measured values for phase shifts for various DC voltage values (for various capacitance values of varactor). On the basis of these results it can be concluded that realization of this device was successful and that noticed differences in results are acceptable.

TABLE 1
SIMULATED VALUES COMPARED AGAINST MEASURED VALUES

DC Voltage [V]	Capacitance [pF]	Phase Shift [°] (ADS)	Phase Shift [°] (Measured)
2	12	169.5	171
4	9	179	180
6	7.3	-169	-171
8	6	-159	-159
10	5	-149	-144
12	4	-190	-128
14	3.4	-114	-115
16	3.1	-106	-105
18	2.8	-97.5	-97
20	2.3	-93.5	-90
22	2.2	-84.5	-85
24	2.1	-83	-80
26	2	-78.5	-77

References

[1] I. Bahl, P. Bhartia, *Microwave solid state circuit design 2nd edition*, Wiley-Interscience, Wiley&Sons Inc., 2003.

[2] "Phase Shifter Design Tutorial" http://www.rfic.co.uk

[3] David M. Pozar, *Microwave engineering*, John Wiley&Sons, Inc, 1998.

Prepoznavanje slika korišćenjem neuronskih mreža i *backpropagation* algoritma

M. Bežulj

Apstrakt - Neuronska mreža za prepoznavanje slika realizovana je u programskom paketu MATLAB korišćenjem backpropagation algoritma. Mreža je uspešno testirana na digitalnim slikama ljudskih lica.

I. Uvod

Veštačke neuronske mreže funkcionišu po uzoru na ljudski mozak i očekuje se da će biti jedna od ključnih koraka u daljem razvoju veštačke inteligencije. One se uspešno mogu primenjivati na mnogim zadacima koji su ljudima jednostavni i prirodni, ali se na računarima veoma teško implementiraju klasičnim metodama programiranja. Najvažnija prednost neuronskih mreža jeste to što se one obučavaju na ograničenom skupu primera. Rešavanje problema neuronskim mrežama ne zahteva poznavanje složenih matematičkih i logičkih funkcija koje povezuju ulazne i izlazne podatke, kao što je to potrebno kod algoritamskog programiranja, gde moramo precizno definisati funkcije prenosa sistema.

Da bi se mreža obučila da ispravno aproksimira određenu funkciju, dovoljno je na odgovarajući način mreži predstaviti problem i obučiti je na reprezentativnom skupu podataka (u daljem tekstu - trening skup). Mreža će, ukoliko je to moguće, pronaći odgovarajuće rešenje, tj. funkciju koja povezuje ulazne i izlazne podatke prilikom procesa obučavanja. Važno je naglasiti da neuronske mreže i klasično programiranje nisu tehnike koje se međusobno isključuju, već naprotiv one se međusobno dopunjuju.

II. Arhitektura Neuronkih Mreža

Ponašanje neuronske mreže zavisi od težinskih koeficijenata između neurona (perceptrona), modela samih neurona, odnosno njihove aktivacione funkcije i od arhitekture mreže.

Aktivaciona funkcija *f* moze imati različite oblike, najčešće se koriste linearna, odskočna, tangens hiperbolička i sigmoidna (*logistička*) funkcija:

$$f(x) = \frac{1}{1+e^{-kx}}$$

Razlog što je odabrana sigmoidna funkcija leži u tome što se njen izvod, koji nam je potreban prilikom obučavanja mreže, najjednostavnije računa i pritom daje odlične rezultate.

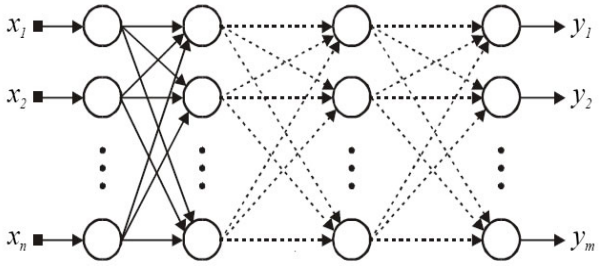
Arhitekturu neuronskih mreža čini specifično povezivanje neurona. Razlikujemo dve klase neuronskih mreža: mreže sa povratnim vezama(*rekurentne*) i mreže bez povratnih veza (*nerekuretne* engl. *feed forward*).

Najpoznatiji predstavnik rekurentih mreža su *potpuno povezane neuronske mreže.* Za njih je karakterisično da se izlazi svakog neurona povezuju na ulaze svih neurona u mreži, uključujuži i sopstveni neuron. Još jedan od karakterističnih predstavnika rekurentnih mreža su *ćelijske neuronske mreže*.

Najznačajniji predstavnik feed forward mreža je *slojevita neuronska mreža*, koja je korišćena u ovom radu. Kod slojevitih mreža (*Slika 1*) neuroni su grupisani u slojeve. Karakteristično za ove mreže je to što se podaci kroz mrežu kreću samo u jednom smeru. Treba naglasiti da je izlaz svakog neurona iz jednog sloja povezan sa svim neuronima u sledećem sloju.

Osnovna podela slojevitih mreža, vrši se prema broju slojeva u mreži, pa razlikujemo jednoslojne i višeslojne. Slojevi višeslojnih mreža dele se, opet, na: ulazni sloj, skriveni sloj i izlazni sloj. Ulazni sloj je prvi, on prima ulazne podatke. Izlazni sloj je poslednji sloj, prikazuje izlaze mreže. Skriveni sloj, služi za internu obradu informacija i obuhvata sve slojeve između ulaznog i izlaznog.

Važno je reći da broj slojeva, u opštem slučaju, ne određuje performanse mreže. Broj neurona u ulaznom sloju je ekvivalentan broju ulaznih podataka, odnosno u ovom radu broju piksela slike koja se obrađuje. Broj neurona u izlaznom sloju jednak je broju izlaza mreže odnosno broju grupa različitih slika iz trening seta na kojima se mreža uči.



Slika 1. Šematski prikaz slojevite neuronske mreže. x - ulaz, n - broj neurona u ulaznom sloju, y - izlaz, m - broj neurona u izlaznom sloju

Najpraktičniji načiin obučavanja mreže je promena težinskih koeficijenata, baš zato što je nakon projektovanja i hardverske realizacije mreže teško promeniti organizaciju ili model neurona. Obučavanje neuronskih mreža moguće je podeliti na dve osnovne kategorije: nadgledano (*supervised*) i nenadgledano (*unsupervised*)

Kod nadgledanog obučavanja mreži se prosleđuju ulaz (pobuda) i očekivani izlaz, zatim se upoređuju stvarni i očekivani odziv i razlika se prosleđuje

proceduri za obučavanje, koja na osnovu razlike menja težinske koeficijente mreže. Najbolji primer ovakve obuke je *backpropagation* algoritam.

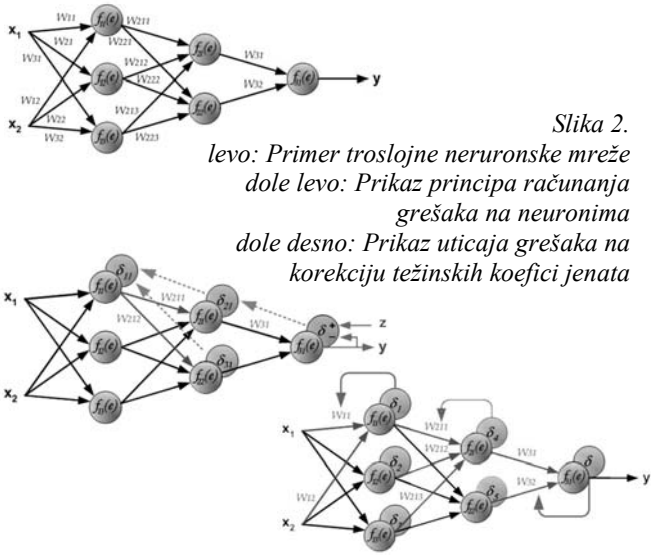
Kod nenadgledanog obučavanja ne postoji informacija o željenom odzivu, od mreže i procedure obučavanja se zahteva da uoče pravilnosti u ulaznim podacima i da definišu konkretan odziv. Dakle procedura učenja menja težinske koeficijente isključivo na osnovu trenutnog ulaza i izlaza.

III. Princip Forward propagation-a i treniranje višeslojne neuronske mreže Backpropagation algoritmom

Forward propagation algoritam podrazumeva prosleđivanje podataka kroz mrežu i izračunavanje izlaza mreže. Zasniva se na klasičnom radu neurona koji su slojevito povezani. Neuroni svoje ulaze sumiraju i propuštaju kroz aktivacionu funkciju čime se definiše izlaz. Svaki neuron svoj izlaz prosleđuje na sve neurone u sledećem sloju. Backpropagation algoritam, kao i svi metodi nadgledanog obučavanja, svodi se na minimiziranje grešaka između dobijenog i željenog odziva mreže.

Koerekcija greške počinje u izlaznom sloju pošto je mreži predstavljen ulazni podatak i završen forward propagation. Izlaz svakog neurona izlaznog sloja se upoređuje sa željenom vrednošću izlaza, a zatim se na osnovu toga za svaki neuron izračunava vredost greške. Nakon toga računa se kreška predhodnih slojeva sve do ulaznog. Kada se završi proces računanja grešaka. Ispravljaju se težinski koeficijenti veza

Primer**:** Troslojna arhitektura, mreža ima 3 neurona u ulaznom sloju, 2 u skrivenom sloju i 1 neuron u izlaznom (*slika 2*). Dva ulazna podatka dolaze na mrežu i prosleđuju se na sva tri ulazna neurona.



Slika 2.
levo: Primer troslojne neruronske mreže
dole levo: Prikaz principa računanja grešaka na neuronima
dole desno: Prikaz uticaja grešaka na korekciju težinskih koefici jenata

Nakon što se propusti ulazni podatak kroz mrežu, izlaz *y* se poredi sa željenim izlazom iz trening seta *z*. Dobijena vrednost δ predstavalja grešku na poslednjem neuronu. δ Zatim se izračunava greška skrivenog sloja za svaki neuron. Greška prvog neurona u skrivenom sloju definiše se kao proizvod koeficijenta koji povezuje izlazni neuron sa ovim neuronom i greške na izlaznom neuronu. $\delta_{21} = \delta * w_{31}$, gde je δ_{21} greška prvog neurona u skrivenom sloju, δ greška izlaznog neurona, w_{31} su težinski koeficijenti koji povezuju ova dva neurona. Analogno tome greška drugog neurona je $\delta_{22} = \delta * w_{32}$.

Greške neurona ulaznog sloja definišu se kao sume proizvoda težinskog koeficijenta koji povezuje neuron ulaznog sloja sa neuronima u skrivenom sloju i greške na odgovarajućem neuronu skrivenog sloja. $\delta_{11} = \delta_{21} * w_{211} + \delta_{22} * w_{212}$, gde je δ_{11} greška prvog neurona ulaznog sloja, δ_{21} greška prvog neurona u skrivenom sloju, δ_{22} greška drugog neurona u skrivenom sloju, w_{211} težinski koeficijent između prvog neurona ulaznog sloja i prvog neurona skrivenog sloja, w_{212} težinski koeficijent između prvog neurona ulaznog sloja i drugog neurona skrivenog sloja (slika 2. dole levo) analogno tome:

greška za drugi neuron $\delta_{22} = b_{31} * w_{221} + \delta_{32} * w_{222}$

greška za treći neuron $\delta_{23} = b_{31} * w_{231} + \delta_{33} * w_{232}$.

Zatim sledi korekcija koeficijenata veza, na sledeći način (slika 2. dole desno) $w_{11}' = w_{11} + n * \delta_{11} * x_1 * f_{11}'$, gde je w_{11}' novi težinski koeficijent između prvog ulaznog podatka i prvog neurona u ulaznom sloju, w_{11} stari težinski koeficijent, δ_{11} greška prvog neurona prvom sloju, x_1 prvi ulazni podatak, f_{11}' izvod aktivacione funkcije prvog neurona ulaznog sloja, pri istim ulaznim podacima i obavezno pre promena koeficijenata veza, n realan broj iz opsega od 0 do 1, koji predstavlja brzinu učenja, analogno tome:

$$w_{12}' = w_{12} + n * \delta_{12} * x_1 * f_{12}'$$
$$w_{13}' = w_{13} + n * \delta_{13} * x_1 * f_{13}'$$
$$w_{21}' = w_{21} + n * \delta_{11} * x_2 * f_{11}'$$
$$w_{22}' = w_{22} + n * \delta_{12} * x_2 * f_{12}'$$
$$w_{23}' = w_{23} + n * \delta_{13} * x_2 * f_{13}'$$

Za skriveni i izlazni sloj koristi se isti princip, ali bez uticaja ulaznih podataka *x*

$$w_{211}' = w_{211} + n * b_{21} * f_{21}'$$
$$w_{212}' = w_{212} + n * b_{22} * f_{22}'$$
$$w_{221}' = w_{221} + n * b_{21} * f_{21}'$$
$$w_{222}' = w_{222} + n * b_{22} * f_{22}'$$
$$w_{231}' = w_{231} + n * b_{21} * f_{21}'$$
$$w_{232}' = w_{232} + n * b_{22} * f_{22}'$$

$$w_{31}' = w_{31} + n * b_3 * f_{31}'$$
$$w_{32}' = w_{32} + n * b_3 * f_{31}'$$

Koeficijent n definiše brzinu učenja mreže. Postižu se bolji rezultati kada se n menja prilikom procesa

obučavanja. Postoji nekoliko metoda kako se to najefikasnije radi. Jedna od metoda, sa porastom broja iteracija smanjuje parametar n. Komplikovanija metoda počinje od male vrednosti parametra, naglo ga povećava kada greška mreže dostigne prosečan nivo, a zatim ga smanjuje do krajnjeg stadijuma učenja. Mala početna vrednost parametra n omogućava bolje definisanje znaka koeficijenta mreže.

Proces učenja neuronske mreže prestaje onog trenutka kada ukupna greška mreže, za sve podatke iz trening seta, ne dostigne zadatu minimalnu vrednost.

IV. Softverska implementacija

Algoritam se sastoji iz nekoliko osnovnih delova. Datoteka *loadpict.m* učitava sve podatke (konkretno slike) i filtrira ih tako da sve budu iste rezolucije i u istoj paleti boja, crno-beloj. Datoteka *koef.m* (prilog 1) generiše mrežu sa proizvoljnim sinaptičkim težinama veza. Pokreće se samo prilikom prve inicijalizacije mreže. Težine veza su u opsegu od -0.5 do +0.5, i nalaze se u matricama *w1*,*w2* i *w3*. *Backprop.m* (prilog 4) je deo programa koji obuhvata proces treniranja mreže.

Program koji je napisan u MATLABu potpuno je fleksibilan i može se jednostavno izmeniti i prilagoditi nekom drugom slučaju. U fajlu *parametri.m* (prilog 2) čuvaju se osnovni podaci o konstrukciji mreže.

Iz priloženog fajla možemo zaključiti da je u ulaznom sloju broj neurona jednak broju piksela slike, odnosno da se svaki piksel slike prosleđuje na zaseban neuron. Ovo svakako nije najoptimalnije rešenje ulaznog sloja, pre svega ono koristi maksimalni broj neurona, međutim ukoliko system radi na ovakvom primeru, ne postoji razlog da kada se uvede izdvajanje lica iz slike klasičnim programiranjem sistem ne radi, ali to je sledeći korak. Empiriski je određeno, uzimajući u obzir vreme učenja i rezultate da u srednjem sloju ima 2/3 neurona u odnosu na ulazni sloj.

Datoteka *calc.m* (prilog 3) predstavlja funkciju koja izvršava *Forward Propagation*. Ona prosleđuje trenutni ulaz *x*, koeficijente mreže matrice *w1,w2,w3* i broj neurona u svakom sloju. Kada se izračuna izlaz svakog neurona u ulaznom sloju, matrica *izl1*, svaki od njih svoj izlaz prosleđuje u aktivacionu funkciju, datoteka tresh.m. Nakon toga, na sličan način računa se i izlaz drugog sloja, *matrica izl2* i konačni izlaz mreže, matrica *izl3* odnosno *izl*.

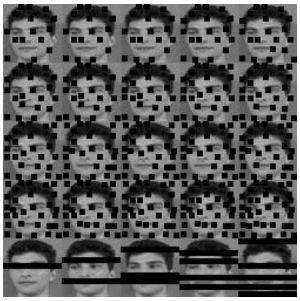
Nakon inicijalizacije programa (učitavanja parametara, trening seta i generisanja koeficijenata mreže) definiše se koeficijent učenja mreže i zatim sledi proces obučavanja mreže, datoteka backprop.m. Obučavanje počinje generisanjem idealnog izlaza mreže za svaku sliku iz trening seta, koji se kasnije koristi za upoređivanja i računanje ukupne greške sistema. Idealni izlaz definiše se tako da redni broj neurona označava grupu istih slika i samo on daje na izlazu 1 kada je na ulazu slika iz njegove grupe.

Izlaz mreže organizovan je na sledeći način: svaka različita slika zauzima jedan neuron na izlazu i kada mreža prepozna određenu sliku, izlaz teži ka tome da samo izlazni neuron koji odgovara grupi u kojoj je ta slika (što je ujedno i redni broj tog neurona) ima na izlazu ima 1, a ostali neuroni 0. Nakon definisanja idealnog izlaza, sve slike iz trening seta prolaze kroz mrežu i računa se početna ukupna greška sistema. Kada je definisana početna greška sistema počnje proces treniranja. Taj proces traje sve dok ukupna greška sistema ne dosigne unapred određenu minimalnu vrednost. Minimalna vrednost je definisana promenljivom *err_limit,* kao broj slika u trening setu pomnožen sa 0.0005. Sastavni deo bloka za obučavanje je i blok za testiranje mreže. Međutim, ako bi se mreža testirala pri svakoj iteraciji obučavanja, vreme obučavanja bi bilo duže, a efikasnost mreže bi ostala ista. Broj iteracija na koliko se vrši testiranje, nalazi se u promenljivoj *err_rec*.

Da u slučaju greške proces treniranja mreže ne bi trajao beskonačno dugo, uveden je brojač iteracija koji zaustavlja algoritam ako broj iteracija pređe preko unapred zadate vrednosti promenljivom *it_limit*. Koeficijent brzine učenja mreže je definisan tako da se menja prema ukupnoj grešci sistema. Početna vrednost iznosi 0.75. Kako se ukupna greška sistema učenjem smanjuje, implementirano je da se koeficijent brzine učenja smanjuje kako se ukupna greska smanjuje. Na primer kada ukupna greška dostigne 200 puta veću vrednost err_limita, greska se smanjuje na 0.5, a kada dostigne 20 puta veću vrednost err_limita smanjuje se na 0.3. Jedna iteracija backpropagation algoritma sadrži uzimanje slučajne slike iz trening seta, propagaciju slike unapred kroz mrežu, poređenje dobijenog sa željenim izlazom i ispravljanje koeficijenata.

V. Verifikacija i rezultati

Za prvo testiranje uzeta je slika iz grupe 4 i postepeno zatamnjivana. Na slici 3 nalazi se 25 različito zatamnjenih slika. Na prvoj slici zatamnjeno je 15% slike i na svakoj sledećoj zatamnjen je 1% više (jedan crni kvadratić više) i to pravilo važi do dvadesete slike. Dvadesetprva slika zacrnjena je za 10% i svaka sledeća je zacrnjena za 10% više.



Slika 3.Prvi test primer

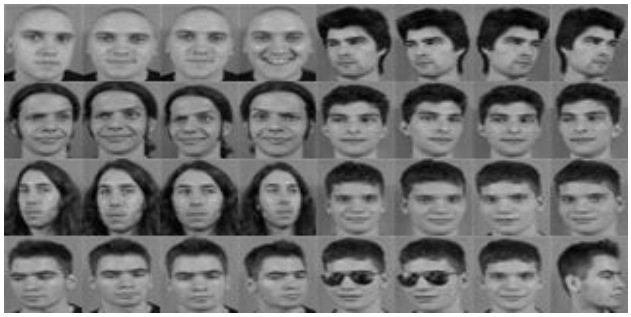
Sl 1.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl 2.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl 3.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl 4.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl 5.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl 6.	0.0000	0.0000	0.0000	0.9998	0.0000	0.0000	0.0000
Sl 7.	0.0000	0.0000	0.0000	0.9998	0.0000	0.0000	0.0000
Sl 8.	0.0000	0.0000	0.0000	0.9998	0.0000	0.0000	0.0000
Sl 9.	0.0000	0.0000	0.0000	0.9995	0.0000	0.0000	0.0000

Sl 10.	0.0000	0.0000	0.0000	0.9997	0.0001	0.0000	0.0000
Sl 11.	0.0000	0.0002	0.0000	0.9999	0.0000	0.0000	0.0000
Sl 12.	0.0000	0.0003	0.0000	0.9997	0.0000	0.0000	0.0000
Sl 13.	0.0000	0.0003	0.0000	0.9996	0.0000	0.0000	0.0000
Sl 14.	0.0000	0.0005	0.0000	0.9982	0.0000	0.0000	0.0000
Sl 15.	0.0000	0.0004	0.0000	0.9940	0.0000	0.0000	0.0000
Sl 16.	0.0000	0.0008	0.0000	0.8809	0.0000	0.0000	0.0000
Sl 17.	0.0000	0.0029	0.0000	0.8719	0.0000	0.0000	0.0000
Sl 18.	0.0000	0.0026	0.0000	0.8615	0.0000	0.0000	0.0000
Sl 19.	0.0000	0.0103	0.0000	0.6323	0.0000	0.0000	0.0000
Sl 20.	0.0000	0.0298	0.0000	0.3593	0.0000	0.0000	0.0000

Iz izlaznih podataka može se zaključiti da mreža daje 99.9% tačnosti za za šum do 30%. Ako pogledamo petnaestu sliku na slici 4, možemo reći da ni čovek ne može sa 100% da utvrdi ko je na slici. Zadnjih 5 slika daje bolje rezultate čak i pri šumu od 40% zato što je šum ravnomerno raspoređen u određenoj oblasti, odnosno postoje veći delovi slike koji su bez šuma.

Sl21.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl22.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl23.	0.0000	0.0000	0.0000	0.9995	0.0000	0.0000	0.0000
Sl24.	0.0000	0.0002	0.0000	0.7456	0.0000	0.0000	0.0000
Sl25.	0.0000	0.0000	0.0000	0.0077	0.0000	0.0000	0.0000

Na slici 4, prvih dvadeset osam slika predstavlja trening set neuronske mreže, dok su poslednje četiri slike samo neki od primera kojima je mreža testirana



Slika 4. Trening set i neke od slika za testiranje mreže

Za testiranje je korišćen podprogram, datoteka testallout.m, koja za priloženu sliku ispisuje sve izlaze mreže. Nakon završenog obučavanja mreže izlaz slika iz trening seta izgleda ovako:

Sl1.	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sl2.	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sl3.	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sl4.	0.9992	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sl5.	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sl6.	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sl7.	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sl8.	0.0000	0.9997	0.0000	0.0000	0.0000	0.0000	0.0000
Sl9.	0.0000	0.0000	0.9992	0.0000	0.0000	0.0000	0.0000
Sl10.	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Sl11.	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Sl12.	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Sl13.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl14.	0.0000	0.0024	0.0000	1.0000	0.0000	0.0000	0.0000
Sl15.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl16.	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Sl17.	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Sl18.	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Sl19.	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Sl20.	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Sl21.	0.0000	0.0005	0.0000	0.0000	0.0000	1.0000	0.0060
Sl22.	0.0000	0.0001	0.0000	0.0000	0.0000	1.0000	0.0000
Sl23.	0.0000	0.0012	0.0000	0.0000	0.0000	1.0000	0.0000
Sl24.	0.0000	0.0001	0.0000	0.0000	0.0000	1.0000	0.0000
Sl25.	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Sl26.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Sl27.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Sl28.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Može se primetiti da postoje odstupanja idealnog izlaza. Ukupna greška ove neuronske mreže iznosi 0.012393, što je u proseku 0.443×10^{-3} i daje zadovoljavajuće rezultate. U navedenom slučaju bilo je potrebno 320 iteracija treniranja. Po izvršenim testiranjima potrebno je od 250 do 500 iteracija da bi ukupnu greška, čiji je maksimum definisan kao broj slika u trening setu pomnožen sa 0.5×10^{-4}, dostigla nivo ispod dozvoljenog maksimuma. Ako bi se ukupna greška povećala, broj potrebnih iteracija bi se znatno smanjio, na račun mogućnosti prepoznavanja mreže. Nakon optimizacije backpropagation algoritma, vreme potrebno za obuku mreže na datom primeru se svodi na opseg od 2 do 5 minuta, dok je za forward propagation potrebno manje od jedne sekunde. Iz ovih rezultata se zaključuje da je obučena mreža u stanju da obrađuje podatke u realnom vremenu.

Najveća ušteda vremena optimizacije jeste ideja da mreža pre početka učenja generiše srednju sliku slika iz trening seta (u daljem tekstu - maska) , zatim se od svake slike iz trening seta oduzima maska. Maskom se postiže da mreža uči samo razlike u licima trening seta, što dovodi do znatno bržeg obučavanja. Bez ideje o formiranju maske sistema i modifikacije trening seta bilo je potrebno više od 20000 iteracija za obučavanje mreže do istog nivoa greške, odnosno obučavanje je trajalo približno 40 minuta.

Poslednje četiri slike sa slike 3 predstavljaju karakteristične test primere mreže, jer se znatno razlikuju od slika iz trening seta. Izlaz mreže za prvu od četri slike je:

0.0000 0.0067 0.0001 0.0000 0.0001 0.9956 0.0627

Što znači da mreža sa sigurnosću od 99,56% prepoznaje da to jeste slika iz grupe 6. Izlaz za drugu sliku je

0.0000 0.0247 0.0012 0.0000 0.0002 0.9253 0.0241

Pored toga što je na toj slici lice još više okrenuto u levu stranu. Mreža sa sigurnošću od 92,53% svrstava sliku u grupu 6. Treća slika je još jedan primer koga nema u trening setu, lice je okrenuto direktno ka posmatraču, mreža daje izlaz

0.0000 0.0666 0.0000 0.0000 0.0000 0.9983 0.2673.

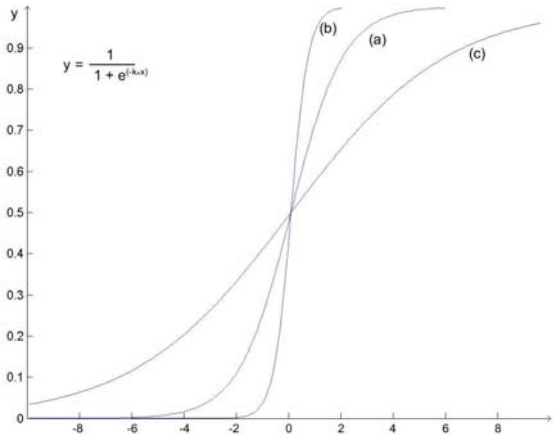
Primećuje se da je najveći procenat

prepoznavanja dodeljen grupi 6, mada je grupa 7 dobila 26.73 %. Pretpostavlja se da je ovakav rezultat posledica nedostatka frontalnog izgleda slike u trening setu, koji se pri maloj rezoluciji zaista poklapa sa frontalnim pogledom lica iz grupe 7. Četvrta slika iz test skupa je nešto komplikovaniji primer, kada se ona dovede u mrežu na izlazu se dobija:

0.0010 0.0000 0.0000 0.0001 0.0000 0.0000 0.4820

Mreža sada sa samo 48.20% tvrdi da to jeste slika 7, međutim u odnosu na ostale izlaze ovaj se znatno izdvaja. Verovatno kada bi u trening setu bilo neko drugo lice u sličnom položaju mreža bi i na izlaznom neuronu te slike dala veći procenat prepoznavanja.

Jedna od metoda izoštravanja rezultata mreže je povećavanje parametra *k* u threshold funkciji. Na slici 5 kriva *a* je izgled aktivacione funkcije kada je $k = 1$, i ovaj oblik je korišćen za treniranje mreže i testiranje. Smanjivanjem parametra *k* izgled funkcije sve više teži ka linearnom (kriva *c*). Taj oblik se nije pokazao koristan u ovom radu. Međutim povećanje parametra k za posledicu ima da sigmoidna funkcija sve više liči na odskočnu. To ima za posledicu da mreža ako su rezultati veći 55% daje sigurnije izlaze, a manje rezultate smanjuje. Tako da, sa dobro projektovanim trening setom (koji obuhvata sve sve varijante objekta koji se očekuje) povećanje parametra *k* nakon procesa obučavanja poboljšava rezultate mreže. Obučavanje sa parametrom *k* kada je ono veće od 1 nije mnogo ispitivano. Ono daje lošije rezultate za slike čija forma mnogo odstupa od slika kojih nema u trening setu. Sa druge strane u mnogo više slučajeva za rezultat daje 100% tačnosti, ali događa se da daje 100% za više od jedne slike. Bitno je napomenuti da je trajanje obučavanja oko 30% duže. Generalno, optimizacija sistema sa takvom sigmoidnom funkcijom ima svoje prednosti i mane.



Slika 5. Izgled aktivacione funkcije pri različitom parametru k
a.) k = 1 b.) k = 3 c.) k = 0.333

Jedan od problema koji se javlja je to što se mreži moraju prosleđivati dobro izdvojene slike. Što podrazumeva fiksna rastojanja izmedju objekta koji je bitan na slici i ivica slike. Na slici 6 vidimo dve naizgled iste slike, ali mreža daje veoma različite rezulate zato što je slika 6*a* uvećana, odnosno razdaljina od lika do ivice slike a je manja nego u trening setu kod slika u trening setu, pa pri promeni veličine slike lik na jednoj slici izgleda veći. Slika 6*b* je primer bolje izdvojene slike pa je veličina glave približnija veličini iz trening seta.

Izlaz mreže za sliku *a* je

0.0000 0.2451 0.0000 0.0000 0.0006 0.3729 0.0018 (k = 1)
0.0000 0.1085 0.0000 0.0000 0.0000 0.2627 0.0000 (k = 3)

Izlaz mreže za sliku *b* je

0.0000 0.0006 0.0000 0.0000 0.0000 0.5160 0.9867 (k = 1)
0.0000 0.0000 0.0000 0.0000 0.0000 0.7650 1.0000 (k = 3)

Iz ovih rezulata vidi se da prvu sliku mreža ne prepoznaje ispravno i vidi se još jedan primer kako menjanje parametra *k* utiče na rezulate. Slika *b* je prepoznata bolje, mada je došlo do mešanja sa grupom 6, kao i u primeru frontalnog pogleda slike iz grupe 6 koja se mešala sa grupom 7. Vidi se i kako je promena parametra *k* povećala tačnost ali i grešku.

Slika 6. Problem prepoznavanja zbog promene plana

a.) b.)

Slike 6a i 6b su slikane na pozadini koja je drugačije osvetljena nego slike iz trening seta što takođe predstavalja problem pri prepoznavanju slike mada ne dolazi mnogo do izražaja jer se radi o crno-belim slikama. Ipak važno je naglasiti, da pozadina takođe utiče na tačnost rezultata mreže.

VI. Zaključak

Neuronske mreže nisu dobro istražena oblast. Zbog toga se u dostupnoj literaturi teško pronalaze konkretni podaci o osobinama mreže i zavisnostima izlaza od menjanja parametara. Tako da se svaki početak istraživanja u ovoj oblasti zasniva na parametrima koji se često određuju empirijski .

Jedna od mana ovog rada je to što radi sa crno belim slikama. Ne postoji problem da se ova mreža unapredi tako da radi sa slikama u boji. Najjednostavnije i najbrže rešenje je da se naprave 3 paralelne mreže. Svaka za jednu komponentu RedGreenBlue sistema. Tako da svaka mreža sa 33.33% učestvuje u krajnjem odgovoru.

Testiranja su bazirana na odzivu mreže za ulaze koji su neobični i u većoj meri različiti od slika iz trening seta. Nije određen kapacitet slika koje neuronska mreža

može da nauči, ukoliko bi došlo do zasićenja mreže. Što znači da za veliki broj različitih slika u trening setu mreža ne može da postigne odziv u granicama dozvoljene greške. Tada treba bazirati istraživanje na radu srednjeg sloja, njego proširivanje ili uvođenje više slojeva u skriveni sloj trebalo bi dovesti do povećanja kapaciteta mreže. Drugo rešenje tog problema bi bilo da se napravi više manjih mreža koje bi obrađivale određeni broj grupa slika iz trening seta. Pri treniranju ovakvih mreža veoma je bitno da se isti trening set koristi za svaku mrežu, ali da se izlazi ograniče tako da svaka mreža prepoznaje određeni broj različitih grupa, ili da se svaka grupa pojavljuje u bar 2 mreže, ali tako da ne postoje 2 identične mreže. Na ovaj način bi svaki konačan odziv mreže bio baziran na odzivu bar 2 manje mreže. Isti trening set bi se koristio za sve mreže kako bi se izoštrili izlazi mreže.

Bitno unapređenje za ovaj program bilo bi napisati pred program koji pri učitavanju slike vrši kvalitetno izdvajanje objekta koji želimo da mreža prepozna ili izvajanje samo karakterističnih delova svakog lica. Podizanje sistema na sledeći nivo značilo bi da se neuronska mreža poveže sa kamerom, bilo preko PC računara ili da se napravi nezavisan sistem sa nekim FPGA čipom, DSP ili CELL procesorom. Tada bi se sistem testirao u realnom vremenu.

Prepoznavanje lica je jedan od najkompilkovanijih primera prepoznavanja slika, tako da ako bi se program koristio za jednostavnije prepoznavanje, na primer saobraćajnih znakova, ili kvaliteta robe u nekom sistemu proizvodnje, mreža bi davala još sigurnije rezultate pošto se radi o mnogo jednostavnijim slikama.

VII. Literatura

[1]Jocković M., Ognjanović Z., Stankovski S. 1997. *Veštačka inteligencija*. Beograd: Krug
[2]Tadeusiewcz R, 1992 "Sieci neuronowe". Kraków
[3]ftp://ftp.sas.com/pub/neural/FAQ.html
[4]http://solair.eunet.yu/~ilicv/ANN.html

Positional Servo-System Design Using Digital PI Controller

I. Bosijoković and A. Marinković

Abstract - The paper presents DC motor position control using digital proportional-integral (PI) controller. In order to avoid generation of abrupt and high-magnitude control effort in case of step changes of reference signal, only the integral action of the controller is located in the direct path. Proportional action uses system output only. The controller design is carried out in discrete-time domain by employing pole placement technique. The developed positional control system has been experimentally tested on a laboratory prototype.

I. Introduction

Position control is an important issue in servo systems designs. DC motors are widely used in these systems because of their linear characteristics and good regulation properties in spite of many shortcomings. These are wear out of the moving mechanical part, presence of brushes and sparking, need for constant maintenance and high costs. Position control of DC servo motor is requested to provide fast and accurate response with respect to position reference. Digital control has the large potential in improving system performances. The advantages of this technology are: easy implementation of more complex control algorithms, flexibility and low costs.

There are a lot of papers in the literature dealing with DC motor control and positional servo systems design. Nowadays, most of servo systems are realized using digital microprocessors in controller implementation, for example [1]-[3]. Sliding mode control methodology is very often used in servo system synthesis, since it ensures great robustness to parameter perturbations and external disturbances [4], [5].

This paper considers positional servo-system design with a DC motor. The control system is implemented using digital controller. The conventional linear controller with proportional and integral (PI) action is adopted, because it is widely used in industry. The controller is designed in discrete-time domain and its parameters are tuned according to pole placement technique [6], [3]. Plant parameters are identified using angular velocity step response. The designed control system has been experimentally tested on a real servo-system [7].

I. Bosijoković and A. Marinković are students at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia.

E-mail: ikabossy@yahoo.com

II. Mathematical Model of DC Motor

DC motors are the most common actuators in positional servo systems due to their linear characteristic and excellent regulation properties. The well known mathematical model of a DC motor my be expressed as [7]

$$\begin{aligned} u_r(t) &= R_r i_r(t) + L_r \frac{di_r(t)}{dt} + c\frac{d\theta(t)}{dt} \\ c i_r(t) &= B\frac{d\theta(t)}{dt} + J\frac{d^2\theta(t)}{dt^2} + M_o(t) \end{aligned} \tag{1}$$

where θ is the angular position; u_r is the rotor (input) voltage; i_r is the rotor current; M_o is the load torque; R_r is the rotor resistance; L_r is the rotor inductance; B is the friction coefficient, J is the moment of inertia and c is the motor constant. State space model of DC motor may be derived by choosing angular position, velocity and rotor current as state variables. The control signal $u(t)$ generates the rotor voltage $u_r(t) = ku(t)$ by means of a power converter with the amplification factor k. The state space model is obtained as

$$\begin{aligned} \frac{d\theta(t)}{dt} &= \omega(t) \\ \frac{d\omega(t)}{dt} &= -\frac{B}{J}\omega(t) + \frac{c}{J}i_r(t) - \frac{1}{J}M_o(t) \\ \frac{di_r(t)}{dt} &= -\frac{c}{L_r}\omega(t) - \frac{R_r}{L_r}i_r(t) + \frac{k}{L_r}u(t) \end{aligned} \tag{2}$$

The time constant of electrical subsystem is usually neglected, because it is much smaller than the one of moving mechanical parts. Thus the model may be reduced to a second order system:

$$\begin{aligned} \frac{d\theta(t)}{dt} &= \omega(t) \\ \frac{d\omega(t)}{dt} &= -\frac{BR_r + c^2}{JR_r}\omega(t) + \frac{kc}{JR_r}u(t) - \frac{1}{J}M_o(t) \end{aligned} \tag{3}$$

According to (3), angular position in terms of control signal and load torque may be described in complex domain by

$$\Theta(s) = \frac{b}{s(s+a)}U(s) - \frac{1/J}{s(s+a)}M_o(s) \tag{4}$$

where $a = (BR_r + c^2)/JR_r$ and $b = kc/JR_r$. Therefore, the resulting transfer function in case of no load torque is

$$G(s) = \frac{\Theta(s)}{U(s)} = \frac{k_m}{s(1+sT_m)} \tag{5}$$

where is $k_m = b/a$ and $T_m = 1/a$.

Since intention is to design a digital controller it is necessary to derive the discrete-time representation of the transfer function (5). Using zero order hold

$$G_{ho}(s) = \frac{1-e^{-sT}}{s} \tag{6}$$

the discrete transfer function is calculated as

$$\begin{aligned} G(z) &= Z\,\mathrm{L}^{-1}\{G_{h0}(s)\cdot G(s)\} = \\ &= (1-z^{-1})\cdot Z\,\mathrm{L}^{-1}\left\{\frac{k_m}{s^2(1+sT_m)}\right\} \end{aligned} \tag{7}$$

The obtained transfer function may be written in the form

$$G(z) = \frac{b_0 z + b_1}{z^2 + a_1 z + a_2} \tag{8}$$

where coefficients b_0, b_1, a_1 and a_2 are given as

$$\begin{aligned} b_0 &= k_m T_m (e^{-T/T_m} - 1 + Te^{-T/T_m}/T_m) \\ b_1 &= k_m T_m (1 - e^{-T/T_m} - Te^{-T/T_m}/T_m) \\ a_1 &= -(1+e^{-T/T_m}) \\ a_2 &= e^{-T/T_m} \end{aligned} \tag{9}$$

III. Control System Design

The proposed digitally controlled DC motor positional servo system is shown in Fig. 1, where $k_T = kc/R_r = Jb$. In order to avoid generation of abrupt and high-magnitude control effort in case of step changes of reference signal, only the integral action of the controller is located in the direct path. P action uses system output only [6]. Closed loop transfer function of the inner control loop, which involves P action, is given with

$$G_1(z) = \frac{G(z)}{1+k_p G(z)} = \frac{b_0 z + b_1}{z^2 + (a_1 + b_0 k_p)z + a_2 + b_1 k_p} \tag{10}$$

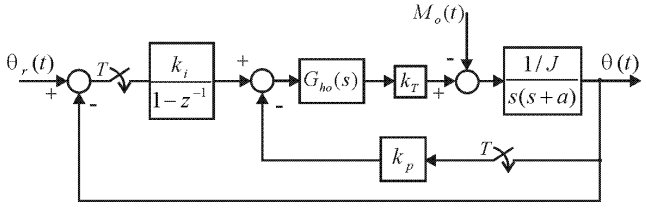


Fig. 1. Block scheme of digitally controlled positional system.

The open loop transfer function of the overall system, Fig. 1, is evaluated as

$$W_p(z) = \frac{k_i z}{z-1} G_1(z) \tag{11}$$

Characteristic equation of the proposed system is defined by

$$1 + W_P(z) = 0 \tag{12}$$

which yields by virtue of (10), (11) and (12)

$$\begin{aligned} &z^3 + (a_1 + b_0(k_p + k_i) - 1)z^2 + \\ &+ (-a_1 + a_2 - b_0 k_p + b_1(k_p + k_i))z - (a_2 + b_1 k_p) = 0 \end{aligned} \tag{13}$$

The control objective is to ensure desired closed loop dynamic behavior, which is determined by predefined values of relative dumping factor ζ and undumped natural frequency ω_n. This goal is provided using pole placement technique. Namely, the dominant closed loop complex conjugated poles z_1 and z_2 should be

$$z_{1,2} = e^{-s_{1,2}T}, \quad s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} \tag{14}$$

Since the system dynamics is described by the third order equation (13), the desired characteristic equation may be expressed then as

$$(z - z_0)(z - z_1)(z - z_2) = 0 \tag{15}$$

that becomes

$$\begin{aligned} &z^3 + (-(z_1 + z_2) - z_0)z^2 + \\ &+ (z_1 z_2 + z_0(z_1 + z_2))z - z_0 z_1 z_2 = 0 \end{aligned} \tag{16}$$

The third unknown real pole z_0 must not be dominant, i.e., it should be located closer to the origin in the z-plane than the other two.

By comparing the coefficients of equations (13) and (16), one may get a system of three equations with respect to three unknown parameters: k_p, k_i and z_0. Solution of the system is obtained as

$$\begin{aligned} k_p &= -\frac{z_1 z_2 (a_1(b_0+b_1) + b_1(-1+z_1) + (b_1 + b_0 z_1)z_2)}{(b_1 + b_0 z_1)(b_1 + b_0 z_2)} + \\ &+ \frac{a_2(b_1 + b_0(z_1 + z_2 - z_1 z_2))}{(b_1 + b_0 z_1)(b_1 + b_0 z_2)} \\ k_i &= \frac{(-1+z_1)(-1+z_2)(-a_2 b_0 + b_0 z_1 z_2 + b_1(a_1 + z_1 + z_2))}{(b_1 + b_0 z_1)(b_1 + b_0 z_2)} \\ z_0 &= \frac{a_2 b_0 (b_0 + b_1) - b_1(a_1(b_0+b_1) + b_1(-1+z_1)}{(b_1 + b_0 z_1)(b_1 + b_0 z_2)} + \\ &+ \frac{(b_1 + b_0 z_1)z_2)}{(b_1 + b_0 z_1)(b_1 + b_0 z_2)} \end{aligned} \tag{17}$$

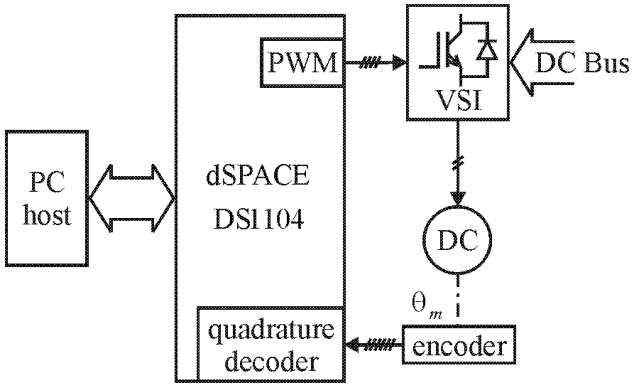


Fig. 2. Block diagram of the experimental platform.

IV. EXPERIMENTAL INVESTIGATION

Experimental investigation of the designed controller has been carried out on a real servo system, whose diagram is shown in Fig. 2. [7] The control subsystem of the positional servo-system is implemented by dSPACE DS1104 R&D controller board, installed on a host computer. This board is suitable for rapid control prototyping, i.e. for development of real-time control applications of multivariable systems. Board programing is carried out under user-friendly MATLAB enviroment.

Low power permanent magnet DC motor is used, with incremental optical encoder for position measurement mounted on its shaft. Motor parameters k_m and T_m has been experimentally identified using angular velocity step response. This response may be approximated by a first order transfer function $G(s)=\frac{\Omega(s)}{U(s)}=\frac{k_m}{1+sT_m}$. If a step signal $u(t)=U_0h(t)$ is applied at the system input it can be easily proven that velocity time response is

$$\omega(t)=k_mU_0(1-e^{-t/T_m}) \tag{18}$$

Motor gain k_m can be determined from the steady state, as t tends to infinity, as

$$k_m=\omega(\infty)/U_0 \tag{19}$$

Hence, in $t=T_m$ (18) may be written as

$$\omega(T_m)=\omega(\infty)(1-e^{-1})=0.63\omega(\infty) \tag{20}$$

which may serve for the evaluation of time constant T_m.

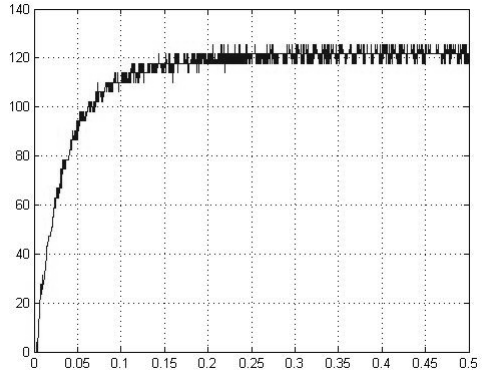


Fig. 3. Velocity step response.

Velocity step response is shown in Fig. 3, for $U_0=5$. Using (19) and (20), data analysis that $k_m=124/5=24.8$ and T_m=0.0379.

The desired closed loop dynamics is set by the following selection of damping factor and natural frequency, respectively: $\zeta=0.5$ and $\omega_n=14\text{rad/s}$. Hence, the desired closed loop dominant poles are found, according to (14),

$$z_{1,2}=0.9972\pm j0.0048 \tag{21}$$

Using (9), (21) and (17), the PI controller parameters as well as the third real pole are calculated as $k_p=0.562939$, $k_i=0.001475$ and $z_0=0.995088$.

Fig.4. depicts referent position (dashed line) and servo system output. The designed system is of type one, since it has an integrating property within controller, and should track ideally step position references. A certain error and deviation from the theoretically predicted response may be noticed in Fig. 4. The main causes are: dry friction, unmodeled dynamics and nonlinearities present in a real system. These effects can be visible only in experimental study.

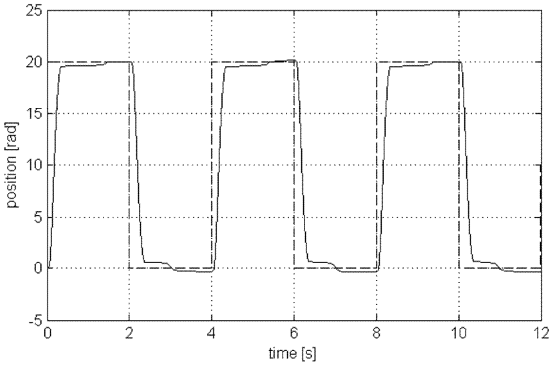


Fig. 4. Position response of the designed system.

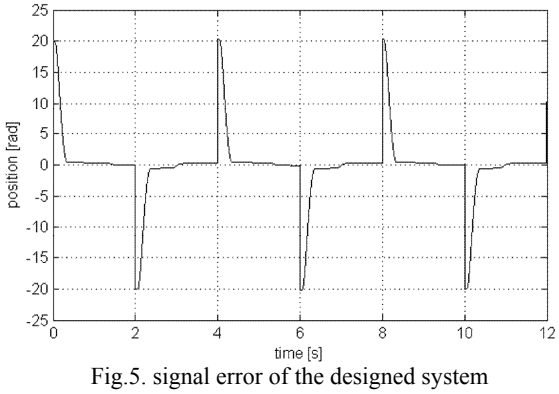


Fig.5. signal error of the designed system

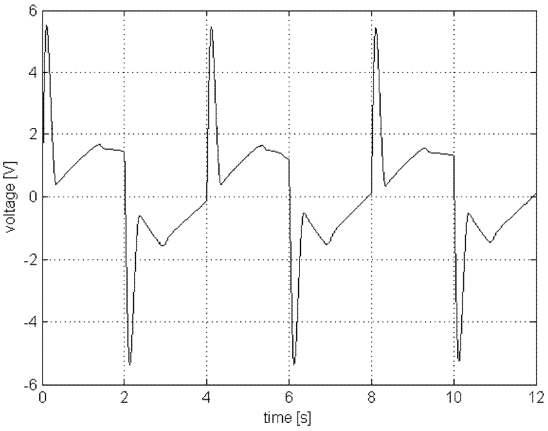


Fig.6. control signal of the designed system

Positioning error is presented in Fig. 5. Error magnitude rapidly decline close to zero value, but it never really sustain zero due to aforementioned nonidealities. Control signal is shown in Fig. 5.

V. CONCLUSIONS

The paper presents the design of positional servo system, by using digital PI controller with displaced P action. Displacing of P-action avoids abrupt high magnitude control signal in case of step input, whereas integral action eliminates the steady-state error. Controller parameters are tuned according to pole placement technique. According to the experimental results, a certain differences may be noticed between experiments and theoretically expected behavior, which are caused by an un-modeled dynamics, nonlinearities and especially by dry friction. The realized positional servo system run very well in all operating conditions, which verifies digital control system design.

REFERENCES

[1] Kiyoshi Onishi, Masato Nakao, Kouhei Ohnishi and Kunio Miyachi "Microprocessor-Controlled DC Motor for Load-Insesitive Position Servo System", IEEE transactions on industrial electronics. Vol. IE-34. 1, February 1987.

[2] Thierry Castagnet and Jean Nicolai "Digital Control for Brush DC Motor", IEEE transactions on industry applications, vol. 30, NO. 4, July/August 1994

[3] Milic R. Stojic "Design of the Micropprocesor-Based Digital System for DC Motor Speed Control", IEEE transactions on industrial electronics. Vol. IE-31. 3, August 1984.

[4] Ali J. Koshkuoei and Keith J. Burnham "Control of DC Motors Using Proportional Integral Sliding mode", Control Theory and Applications Centre, Coventry University, Coventry CV1 5FB, UK

[5] Fu-Juay Chang, Hsiang-Ju Liao and Shyang Chang "Position Control of DC Motors via Variable Structure Systems Control: A Chattering Allevation Approach", IEEE Transcations on industrial electronics, vol. 37, No. 6, December 1990

[6] Milica Naumovic "Control System Design", Faculty of electronic engineering and WUS Austria, Nis, 2005, (in Serbian)

[7] Boban Veselic, "Application of Digital Sliding Mode in Sinthesys of Robust for Coordinated Tracking of Complex Trajectories", Doctoral Dissertation, Faculty of Electronic Engineering, Nis, 2005 (in Serbian)

Semantic Web as an Infrastructural Support for Intelligent Search

M. Đelić, N. Jovanović

Abstract - In this paperwork some of the techniques of implementing semantic web will be presented and also explained through the examples as extracts of the code, as well as those that already exist on the Web and are fully functional. Further more, it will be discussed about the necessity of the semantic web today and its influence on what will Web look like and in which direction it will be developing tomorrow.

I. Introduction

The World Wide Web is the largest information resource humanity has ever produced. Despite the fact it is operating on computers, the most information is only understandable by humans and not by computers. While the syntax of the HTML documents can be easily understood by computers and displayed in a web browser, the thing that computers can not understand is the content or the semantics.

Although not completely realized, the Semantic web already has enough building blocks that are in place to enable taking advantage of several semantic web technologies to be implemented in a web page. Some of these are Resource Description Framework (RDF) and Web Ontology Language (OWL). The goal of the Semantic Web is to expose the vast information resource of the Web as data that computers can automatically interpret.

II. Elements of the Semantic Web

Semantic web can be considered as a stack of layers, as shown in Figure 1, where each layer is resting on and extending the functionality of the layers beneath it.

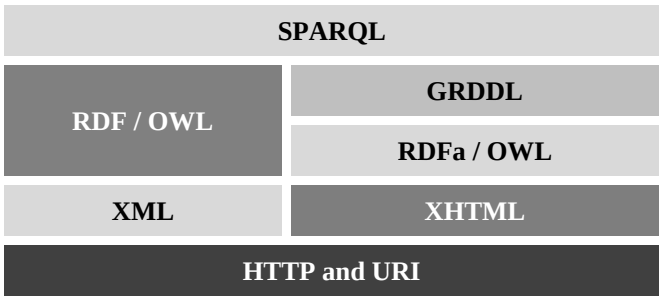


Fig. 1. Semantic Web technology stack

M. Đelić and N. Jovanović are students at the Faculty of Electronic Engineering, University of Niš, A. Medvedeva 14, Niš, Serbia,

E-mail: miloss.djelic@gmail.com, nikolabfs@yahoo.com

Here is important to understand that Semantic Web is an extension and enhancement of the existing Web rather then its replacement.

Although XML represents the basic element of constructing Semantic web it is far less ambiguous if the RDF is used instead. In general, RDF is triple with the following components: a subject, a predicate (or verb), and object. Each can be expressed as a resource on the Web, which is a URI. In the following Listings 1 and 2 this will be shown.

```
<author>
    <uri>page</uri>
    <name>Rob</name>
</author>

<person name="Rob">
    <work>page</work>
</person>

<document
href="http://www.example.org/test/page"
author="Rob" />
```

Listing 1. Relationship in XML

```
<page> <author> <Rob>.
```

Listing 2. Relationship in RDF

It would be very difficult to build software that can derive this relationship from all the possible ways to express it in XML. But an RDF expresses that relationship in only one way, so it becomes feasible to build generic parsers.

III. Metadata

In most cases data on the web is not adequatelly represented just by its name or address. The best example of this is searching for the images, which are often explored only by the surrounding content of the page or their title. This can lead to the confussing and unwanted results given from the search engines. Fot the purpose of improving this process, informaton about the element of the search should be better explained and with more details. Using “data about data”, how the metadata is often

explained, allows us to make content with descriptive terms, or tags, and simplifies organizing content for future navigation, filtering and search. There are several types of metadata which can be applied to some object on the web, and those are:

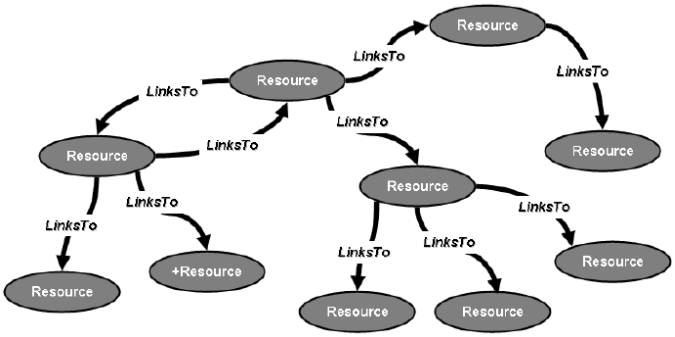
- Descriptive metadata describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
- Structural metadata indicates how compound objects are put together, for example, how pages are ordered to form chapters.
- Administrative metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

The administrative metadata is sometimes dispayed as a two types of metadata:

- Rights management metadata, which deals with intellectual property rights
- Preservation metadata contains information needed to archive and preserve a resource.

Existence of the strucutral metadata brings a lot of improvements in the structure of the web and makes it obvious that if using semantic web pages and other elements on the web will not be simply linked and connected but the relationship between them will be also presented and used as fundamental information for intelligent search.

Figure 2. and Figure 3. show the importance of the structural metadata. In the first case, we know nothing about the links and resourses, except the address of the page, but using metadata we have more information about the elements and their relations.



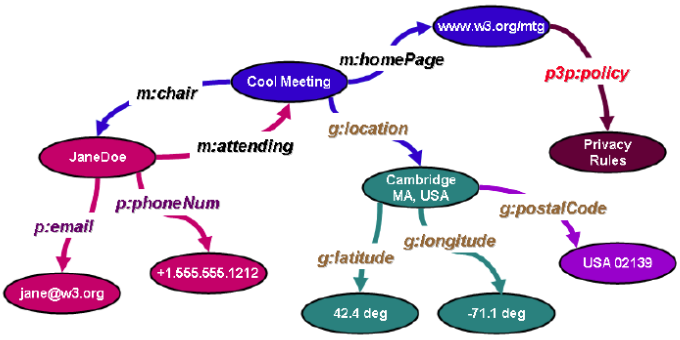


Fig. 2. Minimal machine-processable information

Fig. 3. More machine-processable information: data is connected by relationships

Metadata can be embedded in a digital object or it can be stored separately. Metadata is often embedded in HTML documents and in the headers of image files. Storing metadata with the object it describes ensures the metadata will not be lost, obviates problems of linking between data and metadata, and helps ensure that the metadata and object will be updated together. However, it is impossible to embed metadata in some types of objects (for example, artifacts). Also, storing metadata separately can simplify the management of the metadata itself and facilitate search and retrieval. Therefore, metadata is commonly stored in a database system and linked to the objects described.

IV. Metadata Properties

An important reason for creating descriptive metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration, provide digital identification, and support archiving and preservation.

A. Resource Discovery

Metadata serves the same functions in resource discovery as good cataloging does by: allowing resources to be found by relevant criteria; identifying resources; bringing similar resources together; distinguishing dissimilar resources; and giving location information.

B. Organizing Electronic Resources

As the number of web-based resources grows exponentially, aggregate sites or portals are increasingly useful in organizing links to resources based on audience or topic. Such lists can be built as static webpages, with the names and locations of the resources “hardcoded” in the HTML. However, it is more efficient and more common to build these pages dynamically from metadata stored in databases. Various software tools can be used to

automatically extract and reformat the information for Web applications.

C. Interoperability Describing

A resource with metadata allows it to be understood by both humans and machines in ways that promote interoperability. Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality. Using defined metadata schemes, shared transfer protocols, and crosswalks between schemes, resources across the network can be searched more seamlessly.

D. Digital Identification

Most metadata schemes include elements such as standard numbers to uniquely identify the work or object to which the metadata refers. The location of a digital object may also be given using a file name, URL (Uniform Resource Locator), or some more persistent identifier such as a PURL (Persistent URL) or DOI (Digital Object Identifier). Persistent identifiers are preferred because object locations often change, making the standard URL (and therefore the metadata record) invalid. In addition to the actual elements that point to the object, the metadata can be combined to act as a set of identifying data, differentiating one object from another for validation purposes.

E. Archiving and Preservation

Most current metadata efforts center around the discovery of recently created resources. However, there is a growing concern that digital resources will not survive in usable form into the future. Digital information is fragile; it can be corrupted or altered, intentionally or unintentionally. It may become unusable as storage media and hardware and software technologies change. Format migration and perhaps emulation of current hardware and software behavior in future hardware and software platforms are strategies for overcoming these challenges. Metadata is key to ensuring that resources will survive and continue to be accessible into the future. Archiving and preservation require special elements to track the lineage of a digital object (where it came from and how it has changed over time), to detail its physical characteristics, and to document its behavior in order to emulate it on future technologies.

V. COLLABORATIVE TAGGING

Although organizing electornic content this way is not new, a collaborative form of this process, also called "tagging", is gainig popularity on the web. Document repositories or digital libraries often allow documents to be organized by assigned keyword. Traditionally, this kind of catalogizing or indexing was performed by an athority or librarian, or derived from the mateiral provided by the authors of the document. In contrast, collaborative tagging allows anyone, espacially customers, to freely attach keywords or tags to content. Collaborative tagging is the most useful where there is nobody in the "librarian" role, or there is simply too much content for a single authorty to classify. Both of these to statements are true of the web, where collaborative tagging has grown popular.

VI. COLLABORATIVE TAGGING ON THE WEB

There is planty of sites benefit from using public way of tagging the information. Delicious allows for collaborative tagging and sharing the website bookmarks. Yahoo's MyWeb does this as well, and CiteULike and Connotea do the same to the accademic publications.

fig.4. Delicious bookmarking

VII. COUNCLUSION

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is based on the Resource Description Framework and therefore undestoodable not only by humans but also by machines enbling them to better use and process the given data. Since the infrastructure and technologies change rapidly this method of using, connecting and exploring data should ensure faster and more intelligent way of gathering information in the future.

VIII. REFERENCES

[1] Rowley, J. (1995), *Organizing Knowledge*, 2nd Edition Brookfield,VT: Gower.
[2] W3C Semantic Web Activity, http://www.w3.org/2001/sw/
[3] Scott A. Golder and Bernardo A. Huberman,*The Structure of Collaborative Tagging Systems*,http://arxiv.org/ftp/cs/papers /0508/0508082.pdf
[4]Wu, F. & Huberman, B. (2005). *Social Structure and Opinion Formation*. http://www.hpl.hp.com/research/idl/papers/opinions/ opinions.pdf

[5] *Understanding Metadata*
http://www.niso.org/publications/press/
UnderstandingMetadata.pdf
[6] CiteULike. http://www.citeulike.org/
[7] Connotea. http://www.connotea.org/
[8] Delicious ,http://www.delicious.com/
[9]Yahoo MyWeb 2.0. http://myweb2.search.yahoo.com/myweb

ArchiViewer: "Software for Viewing Three-dimensional Scenes"

A. Tošović

Abstract – ArchiViewer is software that enables viewing of three-dimensional scenes and objects that were designed with 3D Studio MAX application, from first person camera point of view. Actual need for this kind of software, technologies used for its realization, details about different functional blocks and future plans will be stated and explained in the following text.

I. INTRODUCTION

In last ten years, as technology kept improving itself and introducing new computer hardware with faster and more functional components, the great development steps was made in 3D (three-dimensional) design and visualization software. Civil design firms and other companies that sell design-based products use this software as main tool in their work. The main idea is to convert clients request into realistic picture and in that way convince them to buy project. In mind to do that, civil engineers invest money in programs that will make their ideological solutions more realistic. The most common way for representing their work is in form of image or animation. This is done by rendering of 3D scene and saving it as a file. This is the most common way for demonstrating civil projects, but it has one big disadvantage: disability of free moving trough scenes and buildings. It is possible to do it but the customer then need 3000$ expensive 3D Studio MAX or Maya and great knowledge to use these applications. The problem becomes more complicated when we assume that civil engineers never accept to give their editable .max or .obj files.

Main idea of ArchiViewer is to solve these problems and improve customer experience while viewing ideological solution. ArchiViewer represents simple 3D engine that puts user into first person camera and gives him controls to move around the scene and observe objects and buildings from positions he want. The application uses its own file format, named .arv, that can't be imported into 3D Studio MAX. As a result, engineer's data is secured from modification. Beside these functional solutions, ArchiViewer is free, portable, small and easy to use application. The user only needs to know how to open the file and how to use keyboard and mouse. The application is going to be downloadable via website.

A. Tošović is student at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia

E-mail:atosovic@gmail.com.

II. WORKING PRINCIPE

ArchiViewer is application that works as viewing tool in conjunction with the most popular 3D modeling software, like 3D Studio MAX, Maya, Blender, etc… We cannot speak about ArchiViewer as standalone application because it depends on models created in modeling program, in this case 3D Studio MAX. In this section will be explained relationship between these functionally related parts.

3D Studio is application developed by Autodesk company and ArchiViewer is creation of totally independent developer. In order to allow communication between them, appropriate plug-in for 3D Studio MAX and common file format have to exist. Plug-in is created with Max SDK and acts as a one-way door, which lets data to go from 3D Studio to ArchiViewer. The plug-in task is to collect all information of interest from the scene and save it in a specially designed file with ARV extension. When file is created, connection between two applications can be established by opening that file in ArchiViewer. For now, communication from ArchiViewer to 3D Studio is disabled.

We can represent functionality of these applications with two nodes where one node is 3D modeling application (3D Studio MAX for now) and another is first-person viewing application (ArchiViewer). These nodes are connected with ARV file, which is created by modeling software and can be viewed with ArchiViewer. This is shown in figure 1.

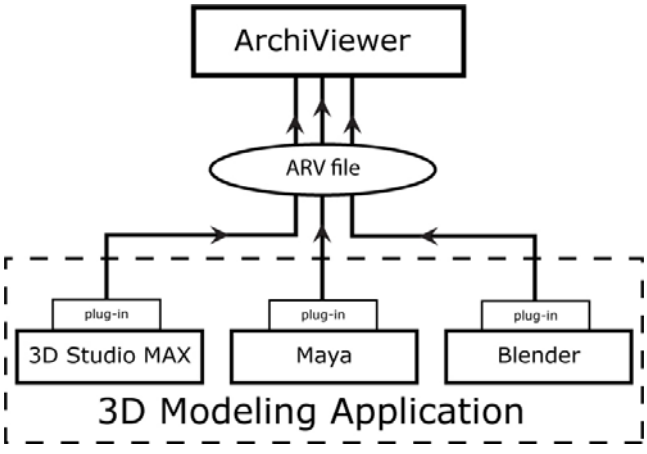


Fig. 1. Graphically represented connection between ArchiViewer and 3D modeling application.

If we want to view models made by other 3D editing tools, like Blender or Maya, development of plug-in for each application is needed. This is because every application developer provide user with its own API or SDK.

III. ARCHIVIEWER

Main task of ArchiViewer is to simplify and enrich user experience while viewing 3D scenes. This is done with implementation of simple engine that enables free walk trough 3D space with mouse and keyboard. This application is simple to use with very few options and commands, but good enough to afford everything that user needs.

Application is developed under Windows XP SP2 operating system. As programming environment, Visual Studio 2005 and C++ programming language with MFC framework are used. Therefore, this version of ArchiViewer works only on Windows machines. However, because user interface is coded separately from core classes it is not difficult task to implement ArchiViewer for other operating systems.

MFC is the C++ class library Microsoft provides to place an object-oriented wrapper around the Windows API [1]. Some of classes, that MFC provide, encapsulate Win32 application programming interface (API). Other classes encapsulate application concepts such as documents, views, and the application itself. The cornerstone of MFC's application framework is the document/view architecture, which defines a program structure that relies on document objects to hold an application's data and on view objects to render views of that data [2]. It is possible to develop two kinds of application using Document/View Architecture: SDI (Single- Document Interface) and MDI (Multi-Document Interface). SDI applications allow only one open document frame window at a time and MDI allow multiple documents. Inside Document/View structure and CView class methods of ArchiViewer, OpenGL calls are made. OpenGL (Open Graphics Library) is a standard specification defining a cross-language cross-platform API for writing applications that produce 2D and 3D computer graphics.

In ArchiViewer MFC is only used for managing windows, resources and program structure. Core classes and functions for managing 3D scenes and working with ARV files are developed separately. Therefore, this viewing tool can be ported to other operating systems, only by reprogramming UI (user interface). One more reason for choosing MFC is fast program execution, because it represents just a wrapper around Win32 API. For ArchiViewer purposes, SDI is used because it demands less memory. Let assume that application uses MDI and user have opened two scenes. For each scene then, application needs to allocate memory for geometry and texture data. As memory has a big influence on frame rate (number of images per second), it is better to restrict opening multiple documents.

OpenGL calls are made inside instance of CView class that is in charge of redrawing window content. Code inside methods of CView object reads geometry data from structures, which are contained inside CDocument, and draws appropriate polygons and textures. CDocument class is also part of Document/View architecture and its duty is to hold application data. In figure 2 we can see architecture of ArchiViewer application.

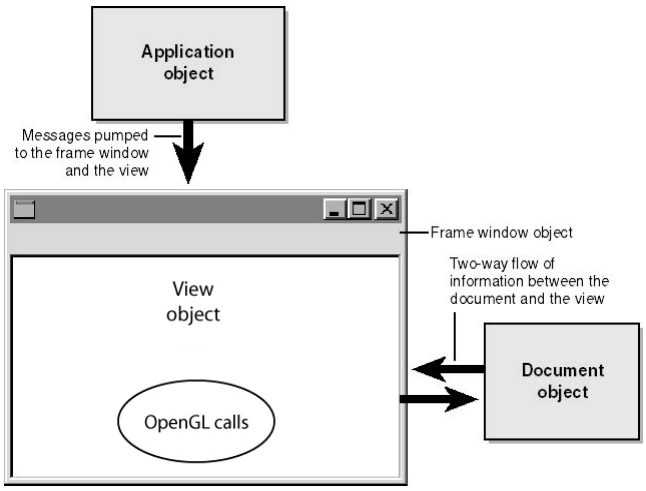


Fig. 2. Architecture of ArchiViewer application.

On top of this complex layer, simple and functional user interface exists. It consists of simple window with menu bar, and keyboard and mouse controls. Menu bar provides user with options to open 3D scene with ARV extension, clear the screen, show information about the scene, turn on and off textures, set anti-aliasing options, switch between smooth and faced normals and show help. Figure 3 shows user interface of ArchiViewer.

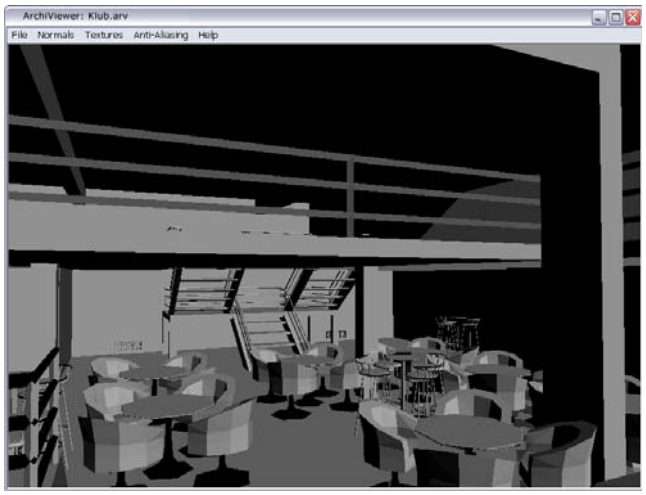


Fig. 3. User interface of ArchiViewer application.

Textures are by default loaded into the scene. However, if the scene contains large number of polygons they can be turned off to improve frame rate while walking trough the scene.

Anti-aliasing is the technique of minimizing the distortion artifacts known as aliasing when representing a high-resolution signal at a lower resolution [3]. In 3D computer graphics, high-resolution signal represents geometry data and lower resolution represent monitor screen. Task of computer program is to represent very precise geometry data on discrete two-dimensional array of pixels. ArchiViewer uses anti-aliasing technique to blur distortions that appears on screen. User can set different levels of anti-aliasing, but he need to be careful because higher level of anti-aliasing demands more resources. Figure 4 shows 3D object in ArchiViewer with anti-aliasing options turned on.

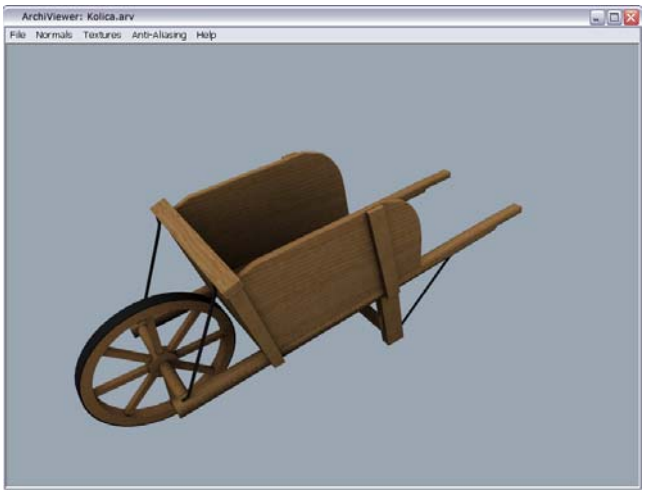


Fig. 4. 3D model with anti-aliasing options turned on.

Surface normals can be manipulated too. There are two options for setting normals: faced and smooth. If faced option is selected, than every vertex normal that belongs to one polygon is perpendicular to its surface. If smooth option is selected, than information about smoothing groups that is contained in ARV file is used. This option provides better results.

ArchiViewer include user reference that can be accessed via Help menu. User reference contains documentation about functionality of every option in ArchiViewer and instructions for viewing the scene.

When ARV file is opened and scene loaded, user can walk trough the scene. Arrows on the keyboard are used for moving and mouse for rotating the camera. Table I shows controls for viewing the scene.

TABLE I
CAMERA CONTROLS

Move left	Left arrow
Move right	Right arrow
Move forward	Up arrow
Move backward	Down arrow
Rotate left	Mouse left
Rotate right	Mouse right
Rotate up	Mouse up
Rotate down	Mouse down

IV. PLUG-IN FOR 3D STUDIO MAX

Because new file format was introduced, it was necessary to develop small program that enables modeling application to export 3D scene into ArchiViewer. It is actually MacroScript plug-in written in MAXscript language. It is visible in 3D Studio MAX toolbar as a button and it starts to execute the script on user left click. Figure 5 shows print screen of 3D Studio MAX toolbar.

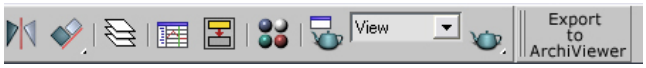


Fig. 5. 3D Studio MAX toolbar with button that executes plug-in

Plug-in is designed to be exporter that converts some of 3D Studio MAX objects into ArchiViewer readable file. As every 3D exporters task, it collects geometry data, vertex coordinates and faces, smoothing group information and normal coordinates. What make this plug-in different and unique is that it helps MAX designer to bake textures assigned to objects and apply UnwrapUVW modifier. In fact, it does not help; it automates this complicated and time-consuming task. Therefore, plug-in also exports image files and texture coordinates.

It is important to mention that before automated rendering to texture and applying UnwrapUVW modifier, plug-in creates snapshot of 3D model in memory. Considering that, plug-in doesn't modify scene models, which are edited by designer. Script frees allocated memory after execution.

In the following subsections MAXScript, rendering to texture and UnwrapUVW modifier are introduced.

A. MAXScript

MAXScript is the built-in scripting language for 3DS Max and related products published by Autodesk [4]. Therefore, MAXScript can't be used outside 3DS Max environment.

It provides users with ability to script almost every aspect of the program's use (geometry, lighting, rendering, animation, texturing …). MAXScript provides command-line listener that can be used for controlling the program interactively. In addition, it gives an opportunity to package scripts within utility panel rollouts or modeless dialogs. Plug-in for ArchiViewer purposes is written as MacroScript and installed as toolbar button. This represents one more ability of MAXScript.

With MAXScript user can access and manipulate objects on the scene with collection of classes and functions that are provided by Autodesk. Programmer or designer can work with geometric objects, access its sub-object elements (like vertices, faces, edges …), control animation flow, change light settings and rendering parameters. As a result of such great capabilities of MAXScript, great portion of 3DS Max was built. Because MAXScript is such a powerful tool, it is used as solution for developing plug-in that will work as exporter to ArchiViewer.

B. Render to texture

Rendering to texture, or "texture baking," allows designer to create texture maps based on an object's appearance in the rendered scene [5]. The textures are then "baked" into the object: that is, they become part of the object via mapping, and can be used to display the textured object rapidly on Direct3D and OpenGL devices such as graphics display cards or game engines.

The task of MacroScript plug-in (exporter) is to bypass manual settings of texture baking and automate the process. If automation is omitted the user needs to set rendering options to each object according to its size and "bake" textures in separate files. In 3DS Max this is done via Render to texture dialog, that is activated via "Rendering" menu and "Render to texture" submenu.

Beside automation, plug-in solves problem regarding to the size of texture map that needs to be baked. Size of baked texture depends on geometric object's size. If small dimensions are used for maps that are going to be applied on large objects in the scene, then quality of applied textures will be low. If large maps are used for small objects then ARV file and allocated memory will be unnecessarily bigger. In that manner, plug-in chooses appropriate texture size according to the size of scene object.

When "baking" process is finished, texture is saved as bitmap and new texture coordinates are calculated. Texture coordinates need to be saved in order to be properly assigned during UnwrapUVW phase.

C. UnwrapUVW modifier

The Unwrap UVW modifier is used to assign planar maps to sub-object selections and to edit the UVW coordinates of those selections [6].

In order to use baked textures on appropriate object, UV texture coordinates need to be memorized and then used with Unwrap UVW modifier. This modifier maps geometric object with texture saved (baked) as bitmap according to coordinates saved during rendering to texture phase.

This modifier is often used for preparing 3D models in video games. It provides game programmer with all data necessary to easily draw textured 3D models in game scene. The only request that needs to be fulfilled is common file format for 3DS Max designer and OpenGL (or DirectX) programmer. This is done by programming or scripting 3DS Max plug-in that will work as exporter.

V. ARV FILE FORMAT

As mentioned before connection between applications is established with ARV file format. This file keeps information about geometric objects, textures and camera. Face data and coordinates of vertices, normals and textures are implied by geometric objects. ArchiViewer uses values of these data to draw models in 3D space. Textures are images that are assigned to appropriate geometry by using suitable texture coordinates. Data about camera tells ArchiViewer eye position, target spot coordinates and field of view. This is shown in figure 6.

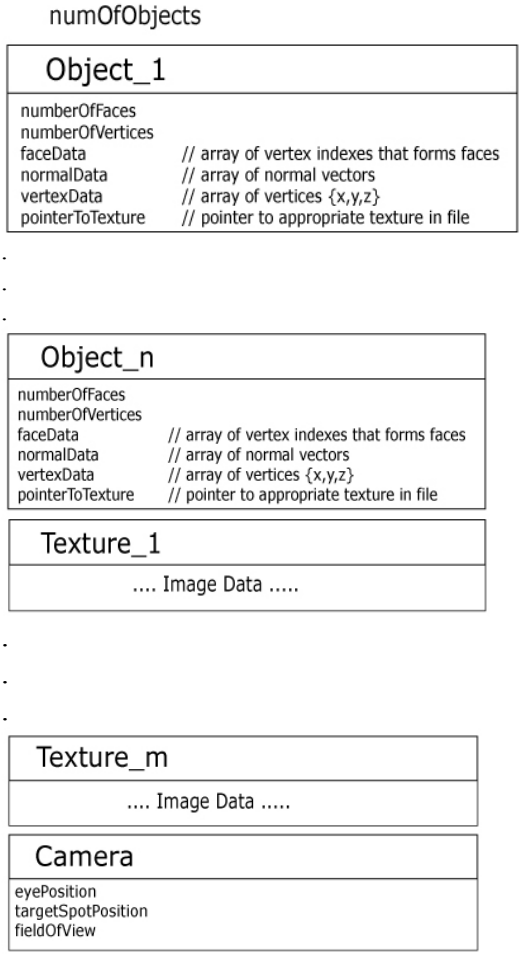


Fig. 6. Structure of ARV file format

ARV file format is built to eliminate unnecessary and recourse-consuming data included in other file formats. Also, it is optimized for reading in ArchiViewer to satisfy walk trough engine demands and civil engineer needs.

VI. FUTURE PLANS

Considering future plans, ArchiViewer tends to be multi-platform application. Thus, the users will not be limited by operating system they use.

Plans are also to develop plug-ins for other modeling applications such as Maya, Blender etc…

Maybe it is a good idea to enable users of ArchiViewer to add comments on some places in 3D scene so civil engineer or designer would now exactly on which details in scene to make modifications or pay attention. This automatically means that ArchiViewer must implement function for saving files and consider their password protection.

If the way of viewing scene with mouse and keyboard find acceptance in users, then next step is to include collision detection. Therefore, camera will not be able to pass trough walls and other objects and this will contribute to richer user experience.

One more feature for future versions of ArchiViewer could be option for manipulating light in 3D scene.

VII. CONCLUSION

ArchiViewer with plug-ins represent free and simple gateway for first person viewing of 3D scenes made in most popular and the highest rated applications for 3D modeling. It tends to represent standard for viewing 3D content and find users in students, civil engineers, designers and ordinary PC users.

REFERENCES

[1] J. Prosise, *Programming Windows with MFC 2nd edition*, Washington: Microsoft Press, 1999.

[2] Microsoft, *Document/view architecture*, MSDN Library [online], Available: http://msdn.microsoft.com/en-us/library/4x1xy43a(VS.80).aspx

[3] Wikipedia, the free encyclopedia, *Anti-aliasing*, [online], Available: http://en.wikipedia.org/wiki/Anti-aliasing

[4] Autodesk, *MAXScript overview*, MAXScript reference 9.0 [compiled HTML help file]

[5] Autodesk, *Render to texture*, 3D Studio MAX User Reference [compiled HTML help file]

[6] Autodesk, *Unwrap uvw modifier*, 3D Studio MAX User Reference [compiled HTML help file]

Senzor Sile S-tipa – Karakteristike i Praktična Primena

Ivan Paunović

Sadržaj – *Kako su senzori postali svakodnevna pojava i imaju najrazličitije primene, javila se potreba za upoznavanjem njihovog rada i inplementacije u sisteme. Ovaj rad opisuje senzor za detektovanje sile i prikazuje faze kroz koje se prolazi od samog odabira senzora, preko detektovanja signala, do njegove konačne praktične realizacije u sistem.*

I UVOD

Senzor sile (load cell) je uređaj koji služi za pretvaranje neke sile (opterećenja) u električni signal, tj. pretvaranje neelektrične veličine u lako merljivu električnu veličinu. Ovo pretvaranje (konverzija) je indirektno i odigrava se u dve faze. U mehaničkom sklopu, sila se detektuje preko deformacije materijala na detektoru deformacija, najčešće promenljivom otporniku. Promenljivi otpornik konvertuje deformaciju u električni signal. Senzor sile se obično sastoji od četiri otpornika (od kojih je jedan promenljivi) postavljena u konfiguraciju Vinstonovog Mosta (Wheatstone Bridge). Postoje, naravno i senzori za merenje sile sa dva ili više promenljiva otpornika, ali se u industrijske primene najčešće koriste gore pomenuti. Izlazni signal sa senzora je najčešće u opsegu od nekoliko milivolti i kao takav zahteva pojačanje instrumentacionim pojačavačem pre nego što se počne koristiti. Pri upotrebi, signal koji se dobije sa senzora (pretvarača), se koristi u jednačinama za izračunavanje sile kojom je opterećen senzor.

I ako su senzori sile sa promenljivim otpornikom uobičajeni, postoje i drugi tipovi senzora sile. U industrijskoj primeni, hidraulični (hidrostatički) senzori sile su takodje veoma zastupljeni, a koriste se da eleminišu neki probleme koji se javljaju kod senzora sile sa promenljivom otpornošću. Na primer, hidraulični senzori su imuni na kratkotrajne promene napona, pa mogu biti mnogo efikasniji za upotrebu u spoljašnjim uslovima. U ostale vrste senzora spadaju piezo-električni senzor sile (koristan za dinamičko merenje sile), kao i senzor sile sa vibrirajućom žicom koji je koristan za merenje u geomehaničkim aplikacijama sa malim vrednostima otklona.

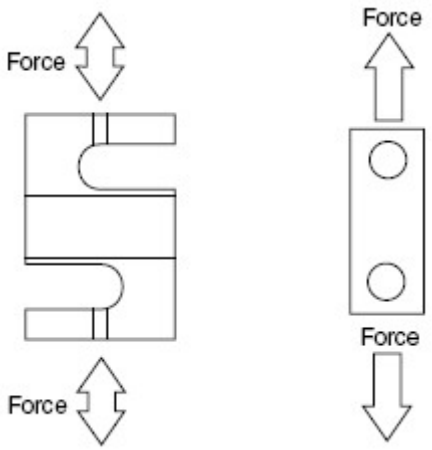
Svaki senzor sile je uređaj koji alarmira kada dođe do promene opterećenja u sistemu, tako da se može koristiti u najrazličitijim sistemima u kojima se može iskoristiti takva njegova namena.

U narednim poglavljima biće opisane opšte karakteristike senzora sile kroz primer senzora sile S-tipa. Biće opisan postupak odabira odgovarajućeg senzora, način praktične realizacije, kao i primer industrijskog sistema u kome je senzor inplementiran.

I. Paunović is student at the Faculty of Electronic Engineering, University of Niš, A. Medvedeva 14, Niš, Serbia,
E-mail: miloss.djelic@gmail.com, nikolabfs@yahoo.com

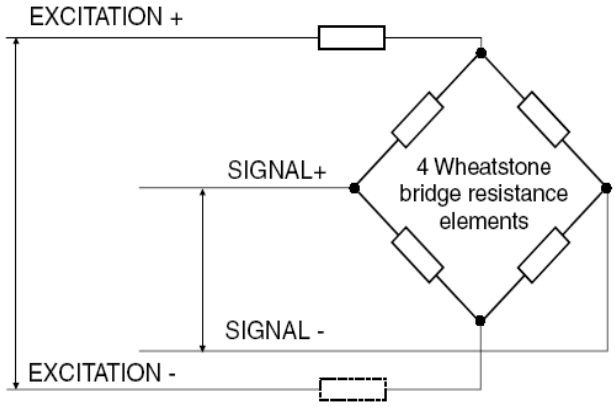
II OPŠTE KARAKTERISTIKE SENZORA SILE

Da bi se senzor sile inplementirao u neki sistem potrebno je poznavati njegove karakteristike i osobine. U tabeli I se nalaze svi podaci potrebni za rad i razumevanje senzora. Senzor, čiji su parametri dati, je senzor S-tipa. Kao što se vidi sa slike 1. ovaj tip senzora se postavlja u pravcu prostiranja sile, tako da bi, prilikom delovanja, došlo do deformacije materijala, tj. do njegovog istezanja, i na taj način do detektovanja sile u vidu izlaznog napona. Ovakvi senzori mogu da funkcionišu kako pri istezanju tako i pri sabijanju samo što u tom slučaju dolazi do promene polariteta izlaznog napona.



Sl.1. Senzor sile S-tipa

Senzor sile S-tipa, kao i većina senzora sile, detektuje opterećenje uz pomoć deformacije materijala, odnosno istezanja. Promenljivi otpornik u ovom senzoru je postavljen u konfiguraciju Vinstonovog mosta, slika 2. Napajanje (excitation) koje se dovodi mostu, predstavlja napajanje senzora, a signali + i – koje dobijamo na izlazu Vinstonovog mosta, su signali koje dobijamo na izlazu senzora. Prilikom istezanja materijala dolazi do promene otpornosti elementa, što dovodi do narušavanja ravnoteže u mostu, a sa tim i do promene signala na izlazu. Pogodnim iskorišćenjem tog signala stiče se uvid u silu kojom je senzor opterećen.



Sl.2. Vinstonov most

Prilikom izbora bilo kog senzora sile, najbitnije je da senzor može da podnese opterećenja za koja je namenjen. Najčešće se bira takav senzor kod koga je maksimalno dozvoljeno opterećenje, bar za 30% veće od maksimalne sile koja se očekuje da će opteretiti senzor. Proizvođači uglavnom predviđaju ovakvu vrstu rezerve, pa je, recimo, za ovaj senzor granica 20% ili ekstremno čak 50% iznad dozvoljnog opterećenja.

Pri projektovanju sistema za merenje sile, mora se imati u vidu činjenica da svaki senzor sile na svom izlazu daje promene napona reda veličine nekoliko milivolta. Zbog toga je neophodno projektovati instrumentacioni pojačavač kako bi se ovaj izlazni signal pojačao i postao upotrebljiv za korišćenje u realnim sistemima. Opseg signala koji dobijamo na izlazu senzora, pre nego što prođe kroz instrumentacioni pojačavač, je veličina koja zavisi od dva parametra. Parametar *nominalni izlaz* (rated output) se izražava u mV/V i označava koliko će se puta, po svakom dovedenom voltu za *napajanje senzora***,** povećati izlazni signal. Na primer, ako je napajanje senzora 10V, izlazni signal će biti 20mV pri maksimalnom opterećenju (10V*2.0027mV/V≈20mV).

TABELA I

O*snovni parametri senzora sile*

Rated output	mV/V	2.0027
Precision	mV/V	0.03
Non-linearity	%F.S	0.028
Histeresis	%F.S	0.028
Repeatability	%F.S	0.029
Creep	%F.S/30MIN	0.03
Temp. effect on zero	%F.S/10°C	0.029
Temp. effect on span	%F.S/10°C	0.020
Zero balance	%F.S	±1
Input impedance	Ω	350±3
Output impedance	Ω	350±3
Insulation resistance	MΩ	≥5000(50VDC)
Recommennded excitation voltage	VDC	9~12
Allowed excitation voltage	VDC	5~18
Compensated temp. range	°C	-10°C~+40°C
Operating temp. range	°C	-20°C~+55°C
Safe overload	%F.S	120
Ultimate overload	%F.S	150
Connection	Excitation Signal	Red: + Black: - Green: + White: -

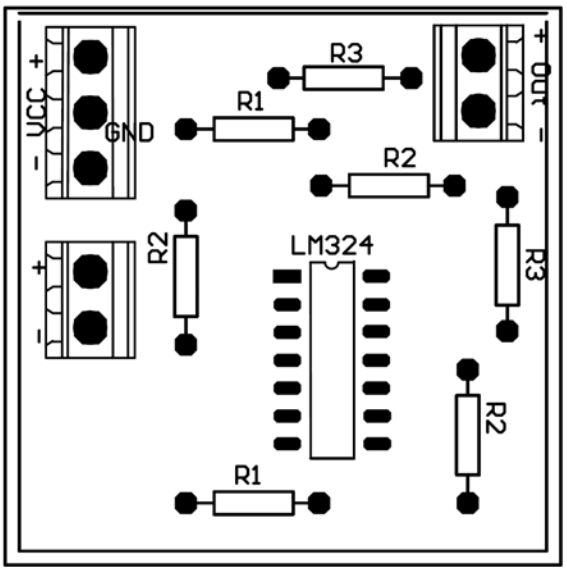
III INSTRUMENTACIONI POJAČAVAČ

Kao što je ranije pomenuto, na izlazu svakog senzora javlja se izlazni napon reda milivolta. Da bi se taj signal mogao upotrebiti potrebno ga je pojačati instrumentacionim pojačavačem na vrednost napona koja je pogodna za praktičnu primenu. Instrumentacioni pojačavači postoje kao gotove komponente ali se u ovom slučaju odlučilo za projektovanje jednog takvog, kako bi se mogao izvršiti niz eksperimenata, za različite vrednosti otpornika, u svrhe pogodne inplementacije u različite industrijske sisteme, kao i za obuku studenata i inženjera početnika u ovom polju.

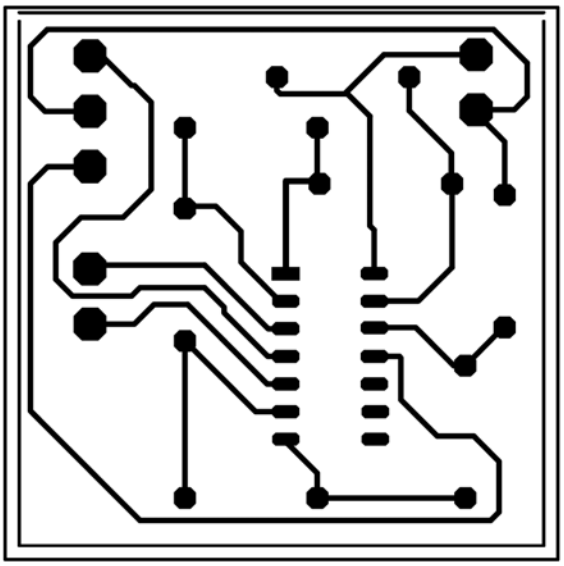
Instrumentacioni pojačavač predstavlja vrstu diferencijalnog pojačavača koja služi za pojačavanje malih nivoa ulaznih signala. Realizovan je sa tri operaciona pojačavača i nekoliko otpornika, od kojih je jedan promenljivi, kao što je to prikazano na slici 5.

Pojačanje instrumentacionog pojačavača se računa prema formuli:

$$\frac{V_{\text{out}}}{V_2 - V_1} = \left(1 + \frac{2R_1}{R_{\text{gain}}}\right)\frac{R_3}{R_2}$$

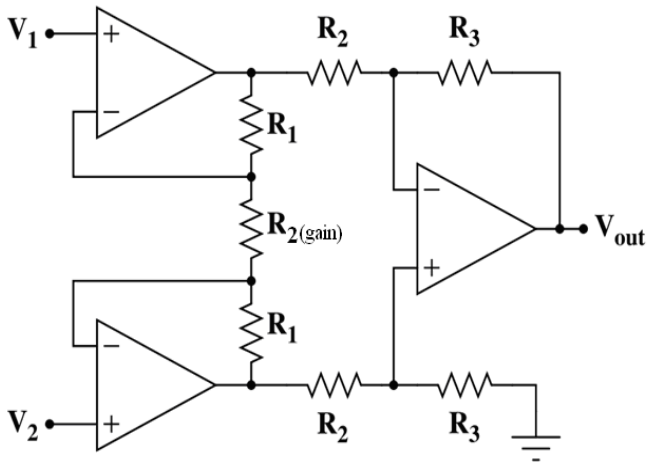


Sl.3. Layout – top



Sl.4. Layout – bottom

V_{out} je napon koji se dobija na izlazu instrumentacionog pojačavača, tj. pojačani napon, a naponi V_1 i V_2 su naponi koji se dobijaju sa senzora i njihova razlika se pojačava. Pri projektovanju ovog instrumentacionog pojačavača promenljivi otpornik R_{gain} je zamenjen otpornikom poznate otpornosti R_2, da bi se olakšalo izračunavanje pojačanja u svim slučajevima kada dolazi do promena otpornosti pri različitim merenjima.



Sl.5. Električna šema instrumentacionog pojačavača

Pošto se javila potreba za dobijanjem izlaznog napona u opsegu od -10V do +10V, uzeto je da vrednosti otpornika budu sledeće: R_1=10kΩ, R_2=1kΩ i R_3=22kΩ. Za operacione pojačavače korišćeno je integralno kolo LM324 koje u sebi sadrži četiri operaciona pojačavača, a za ovu primenu su iskorišćena tri. Ovi operacioni pojačavači moraju da se napajaju sa +12V i -12V.

Pored ovih komponenti korišćene su redne kleme za priključke napona napajanja, kao i ulaznih i izlaznog napona. U sledećoj tabeli II prikazane su komponente koje su korišćene pri izradi instrumentacionog pojačavača.

TABELA II

Spisak komponenti

LM324	1 kom	
R1	2 kom.	10kΩ
R2	3 kom	1kΩ
R3	2 kom.	22kΩ
Redna klema	6 kom.	

Pored potrebnih izračunavanja i određivanja komponenti, odrađena je simulacija u programskom paketu Orcad Family Release 9.2 (Schematics) sa korišćenjem odabranih komponenti. Pošto su rezltati simulatora pokazali da je odziv pojačavača u traženom opsegu, pristupilo se projektovanju layout-a uz pomoć Protel 99 SE programa. Layout-i su prikazani na slikama 3 i 4.

Napajanje pločice VCC ima vrednosti ±12V. Na ulaze ‘’+’’ i ‘’-‘’ se dovode signali sa senzora koji se pojačavaju. Sa izlaza ‘’±Out’’ se očitava vrednost pojačanog napona.

U sledećem poglavlju će biti opisan jedan industrijski sistem u okviru koga je inplementiran senzor sile.

IV PRIMER INDUSTRIJSKOG SISTEMA SA UGRAĐENIM SENZOROM SILE

Kao primer industrijskog sistema razmotrićemo sistem za podizanje i spuštanje ruke robota po vertikalnoj osi. Naime, oko metalnog stuba je postavljen nosač prstenastog oblika, a na taj nosač se vezuje ruka robota za najrazličitije primene. U samom metalnom stubu se nalazi motor koji uz pomoć sajle podiže i spušta nosač ruke. Na sajlu se kači senzor sile sa ciljem da se detektuje eventualno pucanje sajle ili udaranje ruke robota prilikom vertikalnog kretanja.

Ovim sistemom se upravlja uz pomoć PLC-a (programmable logic controller), preko kontrolnog panela sa ekranom na kome se ispisuje trenutni status sistema. PLC daje motoru komandu za kretanje u željenom pravcu, bilo na gore ili na dole, na unapred programiran način. Pored toga, programom smeštenim u memoriju PLC-a, može se zadati i brzina kretanja motora, a samim tim i nosača ruke.

Senzor sile, zakačen na sajlu kojom motor pokreće nosač ruke, je takođe povezan sa PLC-om. Pošto se mehanički postavlja na mesto na kome je predviđeno da stoji, mora se izvršiti povezivanje. Potrebno je dovesti napajanje, koje, u konkretnom slučaju, iznosi +12V (iz tabele I. se vidi da se napajanje dovodi izmedju crvene i crne žice). Zatim se vrši povezivanje sa instrumentacionim pojačavačom, tako što se na ulaz pojačavača dovede pozitivan i negativan signal sa senzora (tabela I. zelena i bela žica). Izlaz pojačavača se povezuje sa PLC-om.

Signal koji stiže u PLC je napon koji se menja, u zavisnosti od opterećenosti senzora, u opsegu od nekoliko volti. Da bi senzor uspešno radio, potrebno je izvršiti kalibraciju. Za različite poznate težine očitava se vrednost napona koji stiže u PLC. Tim vrednostima napona se po analogiji dodeljuju vrednosti težine koje su poznate da senzor u tom trenutku nosi, pa se pogodnim izračunavanjem tako izvršava kalibracija, odnosno prilagodi se PLC da se, na osnovu vrednosti napona koje se očitavaju, prikazuje težina. Ta težina se prikazuje na ekranu kontrolnog panela.

Pošto se izvrši kalibracija senzora, potrebno je podesiti program PLC-a da na osnovu očitane težine obavlja određenu funkciju. U ovom slučaju, senzor bi trebalo da očitava težinu robotske ruke. PLC treba da zaustavi svako kretanje ruke ako, prilikom kretanja, dođe do neke promene težine. Ako, recimo, dođe do naglog smanjenja težine, najverovatnije je došlo do pucanja sajle i u tom slučaju treba automatski aktivirati sve sisteme za kočenje. Prilikom kretanja može doći do udara ruke u neku nepokretnu prepreku, što se detektuje takođe smanjenjem težine kada ruka ide na dole ili povećanjem ako je kretanje bilo na gore.

Na ovaj način je senzor sile iskorišćen u svrhe zaštite uređaja od eventualnog kvara. Takođe se u svakom trenutku može očitavati težina koju motor podiže ili spušta što u nekim primenama može biti od velikog značaja. Na osnovu ovog može se steći sasvim lep uvid u način primene i inplementacije senzora sile u neki od sistema.

V ZAKLJUČAK

Senzori su danas sveprisutni u svakodnevnom životu. Postoje najrazličitije vrste senzora za najrazličitije namene, tako da je njihovo poznavanje kao i način korišćenja za bilo koga ko se bavi elektronikom od izuzetnog značaja. Takav je slučaj i sa senzorom sile.

Prilikom korišćenja senzora sile, neophodno je poznavati parametre koji ga karakterišu, princip rada i način primene, kako bi se mogle iskoristiti sve prednosti koje poseduje, prilikom inplementacije u neki sistem. Tako se, poznavajući parametre, može izvršiti adekvatan odabir samog senzora za sistem u koji želimo da ga postavimo. Ukoliko se zna način primene može se projektovati sistem koji bi maksimalno koristio signale sa senzora za traženu namenu.

Senzor sile koji je bio predmet ovog rada, je korišćen u svrhe industrijskog sistema, čiji je deo opisan u prethodnom poglavlju. Prilikom projektovanja tog sistema (u pitanju je automatizovani rentgen aparat), vodilo se računa o bezbednosti ljudi koji bi trebali da ga koriste, kao i o bezbednosti uređaja. Iz navedenih razloga se javila potreba za primenom ovog senzora. Pošto je prototip ovog sistema zahtevao niz eksperimenata sa različitim vrstama komponenti, a i samih uređaja u sistemu, prilikom primene senzora projektovan je instrumentacioni pojačavač kako bi se olakšala upotreba i povećao opseg testiranja.

Zbog svega toga, ovakva primena senzora je korisna, kako za upotrebu u industriji, tako i za obuku početnika u ovoj oblasti jer obuhvata korišćenje različitih alata za projektovanje i praktičnu realizaciju. Pored ovoga, korisno je poznavati način inplementacije senzora u neki sistem.

LITERATURA

[1] Read Research Group, " Industrial Scales & Load Cells", October 2007.

[2] '' http://en.wikipedia.org/wiki/Load_cell_'', Avgust 2008

[3] Jon S. Wilson, "Sensor Technology Handbook", Decembar 2004.

[4] David M. Scott, " Industrial Process Sensors", Novembar 2007.

Abstract – As sensor became regular occurence and have the most various function, there was a need to introduce their work and inplementation into systems. This work describes force detecting sensor (load cell) and presents which phases are occuring from simple sensor selection, through detecting signal, till his final practical realisation into system.

Load cell S-type – Characteristic and Practical Realisation

Ivan Paunović

Sampling over lattices and sampling efficiency of a different sampling lattices

Ivan Ristić

Abstract –This paper describes how lattices can be used to sample a video signal and compares sampling efficiency when three different lattices are used to sample signal with a sphere spectrum support region.

I. BASICS OF A LATTICE THEORY

When video signal is sampled, samples are normally taken on a rectangular grid. Non-rectangular grid can also be used for signal sampling as long as the grid has a structure that allows specification of the grid points using integer vectors. This type of grid is also known as a lattice. The theory of sampling multi-dimensional signal on a lattice was first presented by Petersen and Middleton [1]. For a comprehensive treatment of the lattice theory, the readers are referred to [2].

A. Mathematical Apparatus

A lattice Λ, in the real K D space, $\mathcal{R}^K$ is the set of all possible vectors that can be represented as integer-weighted combination of a set of K linearly independent basis vectors,

$$v_k \in \mathcal{R}^K, k \in \mathcal{K} = \{1, 2, \dots, K\} \quad (1)$$

That is,

$$\Lambda = \left\{ x \in \mathcal{R}^K | x = \sum_{k=1}^{K} n_k v_k, \forall n_k \in \mathcal{Z} \right\} \quad (2)$$

The matrix $V = [v_1, v_2, \dots, v_K]$ is called the generating matrix. Generating matrix is also called as the sampling matrix.

B. Example

Consider two lattices in $\mathcal{R}^2$ with the following generating matrices:

N. Dojčinović is student at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia
E-mail: ivanristich@hotmail.com

$$V_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \; V_2 = \begin{bmatrix} \sqrt{3}/2 & 0 \\ 1/2 & 1 \end{bmatrix} \quad (3)$$

We first draw the two points corresponding to the basis vectors, and then determine points corresponding to typical integer combinations of the basis vectors. Based on these points, we can extrapolate all the other possible points usually by visual inspection.

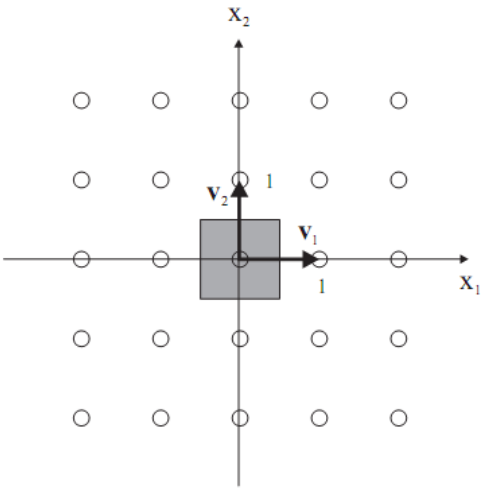


Fig. 1: Rectangular lattice generated using V_1 generating matrix

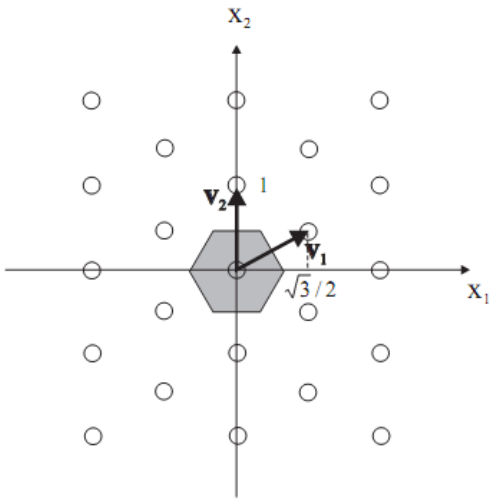


Fig. 2: Hexagonal lattice generated using V_2 generating matrix

The fact that any point in a lattice can be indexed by an integer vector makes the lattice an elegant tool for sampling a continuous signal with a regular geometry.

II. QUANTIZATION OF A SPACE USING A LATTICE

Given lattice Λ, unit cell $\mathcal{U}(\Lambda)$ can be found so that its translations to all lattice points form a non-overlapping covering of entire space $\mathcal{R}^K$

$$\bigcup_{x \in \Lambda}(\mathcal{U} + x) = \mathcal{R}^N \wedge (\mathcal{U} + x) \cap (\mathcal{U} + y) = \emptyset, x \neq y \quad (4)$$

Where $\mathcal{U} + x = \{p + x | p \in \mathcal{U}\}$ denotes the translation of $\mathcal{U}$ by x.

The above theorem tells us that the space $\mathcal{R}^K$ can be represented as a tiling by a chosen unit cell and its translations. This representation is useful when we consider the quantization of the space $\mathcal{R}^K$. Among several interesting unit cells, fundamental parallelepiped and the Voronoi cell are most useful.

A. *Fundamental parallelepiped*

The fundamental parallelepiped of a lattice with basis vectors $v_k, k \in \mathcal{K}$, is set defined by

$$\mathcal{P}(\Lambda) = \{x \in \mathcal{R}^K | x = \sum_k \alpha_k v_k, \forall\, 0 \leq \alpha_k \leq 1\} \quad (5)$$

This is the polygon enclosed by the vectors corresponding to the basis vectors.

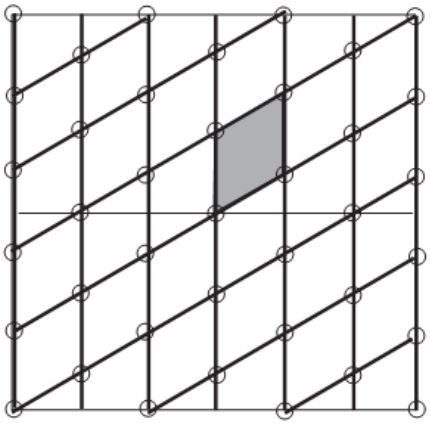


Fig. 3: Fundamental parallelepiped

Fundamental parallelepiped and its translations to all the lattice points form a partition of the space $\mathcal{R}^K$, and therefore the fundamental parallelepiped is a unit cell. Obviously, there are many fundamental parallelepipeds associated with a lattice, because of non-uniqueness of the generating basis

B. *Voronoi Cell*

The Voronoi cell of a lattice is the set of points which are closer to the origin than any other points in the lattice. That is:

$$\mathcal{V}(\Lambda) = \{x \in \mathcal{R}^K | d(x, 0) \leq (x, p), \forall p \in \Lambda\} \quad (6)$$

As the fundamental parallelepiped, the Voronoi cell and its translations to all the lattice points also form a partition of the space $\mathcal{R}^K$. Therefore, the Voronoi cell is also a unit cell. Voronoi cell is very useful for analyzing the sampling process.

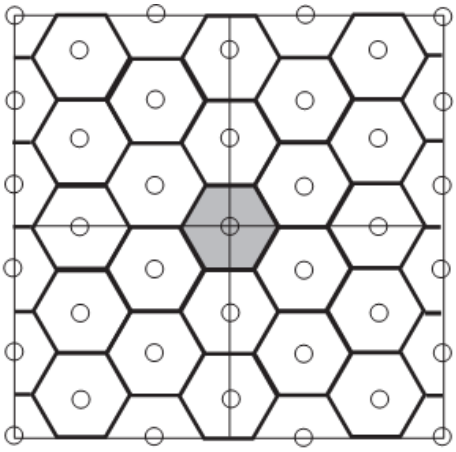


Fig. 4: Voronoi cell

III. VOLUME OF A UNIT CELL AND SAMPLING DENSITY

Although the unit cell associated with a lattice is not unique, the volume of the unit cell is unique. This is because the same number of unit cells is required to cover a finite subspace of $\mathcal{R}^K$ regardless the shape of the unit cell. The volume of the fundamental parallelepiped unit cell is $|detV|$. Obviously, the smaller is the unit cell, the more lattice points exist in a given volume. Therefore, the inverse of the volume of the unit cell measures the sampling density, which will be denoted by:

$$d(\Lambda) = \frac{1}{|detV|} \quad (7)$$

This variable describes how many lattice points exist in a unit volume in R^K.

IV. SAMPLING EFFICIENCY

To avoid aliasing, the sampling lattice must be designed so that the Voronoi cell of its reciprocal lattice completely covers the signal spectrum. We can always design a very dense lattice to satisfy this requirement. This however will require a very high sampling density. To minimize the sampling density, we should design the lattice so that its reciprocal Voronoi cell covers the signal spectrum as tightly as possible. The solution will depend on the signal spectrum, which in general could have an arbitrarily shaped support region. Fortunately, most real-world signals are symmetric in frequency contents in different directions (after properly scaling the frequency axes) so that their spectrum support regions can be approximated well by a sphere.

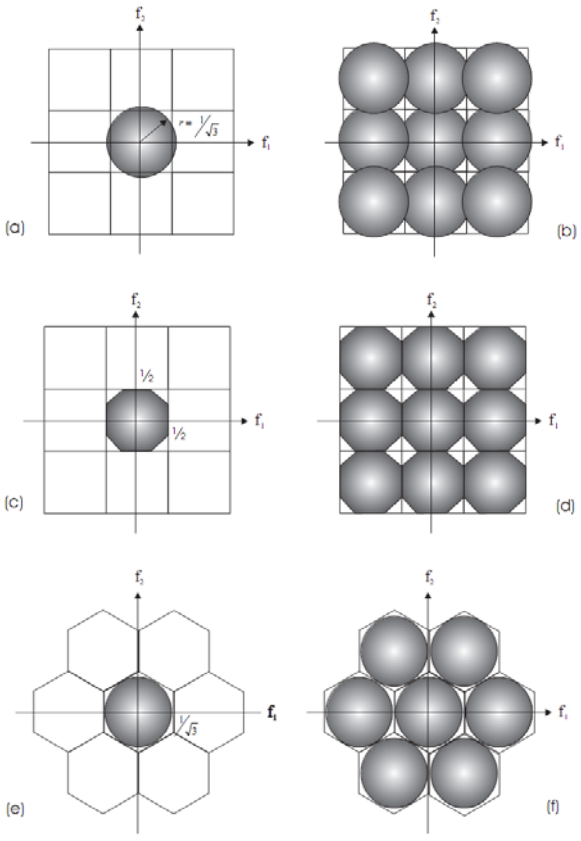


Fig. 5: Illustration of the sampling process in the spectral domain (a) the original signal; (b) the sampled signal using the rectangular lattice; (c) the prefiltered signal; (d)the sampled signal after pre filtering; (e) original signal; (f) the sampled signal using the hexagonal lattice.

To compare the sampling efficiency of different lattices, we can evaluate the sampling densities required for a signal that has a spherical support with radius one. To avoid aliasing, the Voronoi cell of the reciprocal of the sampling lattice, $\mathcal{V}(\Lambda^*)$, must enclose the spehre. The tighter fit is $\mathcal{V}(\Lambda^*)$ to the sphere, the lower is the required sampling density. Tis leads to the definition of the sampling effeiciency by:

$$\rho(\Lambda) = \frac{\text{volume(Unit Sphere)}}{\text{volume}\ (\mathcal{V}(\Lambda^*))} = \frac{\text{volume(Unit Sphere)}}{d(\Lambda)} \qquad (8)$$

The closer is the above value to 1, the more efficient is the lattice. Figure 5 shows the efficiency in the sphere covering of the several 2D lattices. It can be seen that a hexagonal lattice can cover the spheremost tightly. Because of this, the sampling efficiency can also be defined relative to density of the hexagonal lattice. Then the efficiency of the hexagonal lattice becomes one.

V. SUMMARY

A sampling lattice is more efficient if the Voronoi cell of its reciprocal lattice covers the support region of the signal more tightlz. For example, for a signal with cicular spectrum support, the rectangular sampling lattice is less efficient than a hexagonal or a diamond lattice.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge my indebtedness to prof. Dr Saša Nikolić of Faculty of Electronic Engineering of Niš, for a most valuable discussion on the subject of this note.

REFERENCES

[1] D. P. Petersen and D. Middleton. - Sampling and reconstruction of wave-number limited functions in N-dimensional Euclidean spaces. Information Control, 5:279- 323, 1962.

[2] J. W. S. Cassels. An introduction to the Geometry of Numbers. Springer-Verlag, Berlin, Germany, 1959.

[3] N. A. Diakides. Phosphor screens. In D. G. Fink, editor, Electronics Engineers' handbook, pages 11-33-11-39. McGraw Hill, New York, 1975.

[4] E. Dubois. The sampling and reconstruction of time-varying imagery with application in video systems. Proceedings of IEEE, 73:502-522, 1985. Section 3.7. Bibliography

[5] D. E. Dudgeon and R. M. Mersereau. Multidimensional Digital Signal Processing. Prentice Hall, Englewood Cliffs, NJ, 1984.

[6] L. D. Miller. A new method of specifying the resolving power of television camera tubes using the RCA P-300 test chart. Journal of Society of Motion Picture Television Engineering, 89:249-256, Apr. 1980.

[7] R. A. Beuker and I. A. Shah. Analysis of interlaced video signals and its applications. IEEE Trans. Image Process. 3(5):501-512, Sept. 1994.
[8] G. J. Tonge. - The television scanning process. Journal of Society of Motion Picture Television Engineering, 93:657-666, July 1984.

The Schumann resonance and brainwaves

N. Jovanović, M. Đelić

Abstract - In 1952 Winfried Otto Schumann, published his first paper about electromagnetic waves in the waveguide which is formed by the earth's surface and the ionosphere. Since this time the study of these, later referred to as Schumann resonance waves, has become an interesting subject for research. In the last years, one of the most interesting topics, regarding the Schuman resonance, is the one about its possible correlations with the human psychobiology. Science is yet to determine whether this correlation is just a coincidence or it may be that these waves have a far more significant effect than it was primarily considered.

I SCHUMANN RESONANCE

The ionosphere is a layer in the earth's upper atmosphere where a large portion of the atoms and molecules have been ionized by exposure to the ultraviolet radiation of the sun. With so many charged particles free to roam around, the ionosphere is a reasonably good conductor of electricity. The surface of the earth is also a reasonably good conductor. This should be somewhat obvious since 70% of the earth's surface is covered in saltwater, which will short out electrical equipment as everyone knows, and the remaining 30% is exposed rock or soil, the stuff that electrical circuits are grounded to. The layer of atmosphere in between these two conductors is ordinary, non ionized air, which is transparent to radio waves. For extremely low frequency (ELF) radiation, the gap between the earth and its ionosphere acts as a spherical wave guide — a kind of racetrack for radio waves. Lightning and other natural phenomena generate ELF waves at all sorts of different frequencies. Those frequencies that are just right will travel around the earth, meet themselves in phase, and form standing waves. The set of frequencies that will do this are known as the Schumann resonances in honor of Winfried Otto Schumann (1888-1974, Germany), the scientist who predicted their existence in 1952. [1]

Each finite waveguide has its characteristic natural frequency (resonant frequency). Schumann recognized for the first time, that the space bounded by the highly conducting earth and the likewise highly conducting ionosphere represents such a waveguide (Fig. 1).

N. Jovanović and M. Đelić are students at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia,

E-mail: nikolabfs@yahoo.com, miloss.djelic@gmail.com

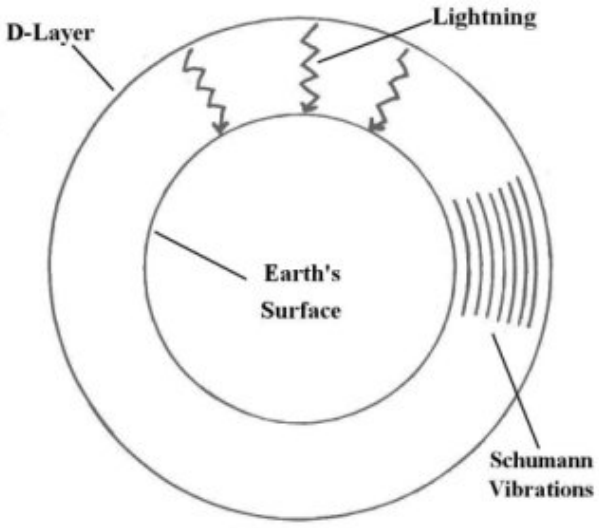


Fig. 1. Earth's cavity formed within earth surface and ionosphere

The fundamental frequency ought to be roughly the time it takes electromagnetic radiation to go all the way around the spherical shell. Since the speed of light is about 300,000 km/sec and one cycle is the circumference of the Earth, which is about 40,000 km/cycle, the fundamental frequency should be on the order of

$$f_n = \frac{v}{\lambda_n} = \frac{v}{2\pi r_E} n \approx 7.5n\ [Hz] \qquad (1)$$

where n=1 represents thereby the fundamental frequency, all larger n the higher harmonic waves.

The Schumann Resonances are actually observed by experiment to occur at several frequencies between 6 and 50 cycles per second; specifically 7.8, 14, 20, 26, 33, 39 and 45 Hertz, with a daily variation of about +/- 0.5 Hertz. The 7.8 Hz observed fundamental resonance is close to the rough theoretical estimate of 7.5 Hz.

Schumann's contribution consisted of the fact that he derived the eight frequencies of this waveguide mathematically in a very general form. Taking spherical geometry into account one obtains the resonant frequencies

$$f_n = \frac{V(\sigma)}{2\pi r_E}\sqrt{n(n+1)} \approx 6.0\sqrt{n(n+1)}\ [Hz] \qquad (2)$$

which also includes the damping of the waves because of the finite conductivity of the "upper" boundary of the waveguide, namely the ionospheric D-layer. Its

propagation speed V therefore amounts to - depending on the conductivity σ - about 80% of the speed of light. While the conductivity of the ionospheric lower boundary is very variable at a height of 70-90 km with values between 10-5 to 10-3 S/m, the average conductivity of the ground and the sea at approximately 10-3 S/m is practically constant and usually greater than the latter. It therefore does not contribute substantially to the damping of the waves. [3]

The first measurements were published by Schumann and König in 1954. Between 1960 and 1970 further measurement results were published as well as extended theoretical work on the topic. Schumann resonance experienced a renaissance at the beginning of the 1990's, after measurement methods had improved and new applications appeared. Some of the fields of its application are global lightning triangulation, global D-layer monitoring and space weather analysis, climate change research, biological effects etc.

II BRAINWAVES

Human brain is made up of billions of brain cells called neurons, which use electricity to communicate with each other. The combination of millions of neurons sending signals at once produces an enormous amount of electrical activity in the brain, can be detected using sensitive medical equipment (such as an EEG), measuring electricity levels over areas of the scalp. The culmination of electrical activity of the brain is commonly called a Brainwave pattern, because of its cyclic, "wave-like" nature.

With the discovery of brainwaves came the discovery that electrical activity in the brain will change depending on what the person is doing. For instance, the brainwaves of a sleeping person are vastly different than the brainwaves of someone wide awake. Over the years, more sensitive equipment has brought us closer to figuring out exactly what brainwaves represent and with what they mean about a person's health and state of mind. [4]

Researchers in the 1930's and 40's identified several different types of brain waves, ranging from the most activity to the least activity. Traditionally, they are categorized into four types (Fig. 2).

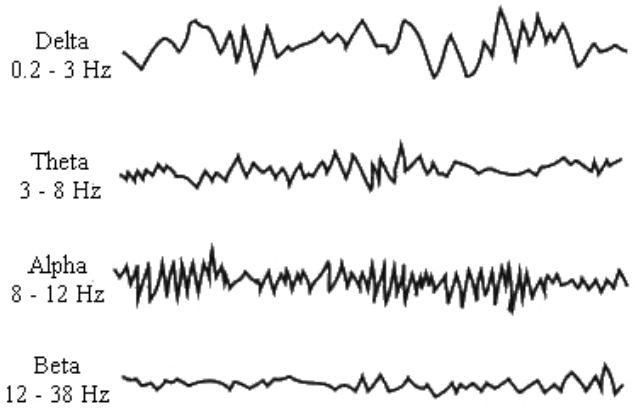


Fig. 2. Recordings of the different types of brainwaves

When the brain is aroused and actively engaged in mental activities, it generates *beta* waves. These beta waves are of relatively low amplitude, and are the fastest of the four different brainwaves. The frequency of beta waves ranges from 12 to 40 cycles a second. Beta waves are present when we are alert or even anxious, or when engaged in problem solving, judgment, decision making, information processing, mental activity and focus.

The next brainwave category in order of frequency is *alpha*. Where beta represented arousal, alpha represents non-arousal. Alpha brainwaves are slower, and higher in amplitude. Their frequency ranges from 8 to 12 cycles per second. Alpha waves aid overall mental coordination, calmness, alertness and learning.

The next state, *theta* brainwaves, are typically of even greater amplitude and slower frequency. This frequency range is normally between 3 and 8 cycles a second. Theta waves occur most often in sleep but are also dominant in the deepest states of meditation and thought.

The final brainwave state is *delta*. Here the brainwaves are of the greatest amplitude and slowest frequency. They typically center around a range of 0.5 to 3 cycles per second. These waves are generated in deepest meditation and dreamless sleep. Certain frequencies within the delta range trigger the release of a growth hormone which is beneficial for healing and regeneration.

Researchers have found that not only are brainwaves representative of mental state, but they can be stimulated to change a person's mental state, and even help treat a variety of mental disorders. Certain Brainwave patterns can be even be used to access many extraordinary experiences.

TABLE I
BRAINWAVE TYPES AND THEIR ASSOSIATED MENTAL STATES

Wave	Frequency	Mental State
Beta	12Hz-38Hz	Wide awake, alert mental state.
Alpha	8Hz-12Hz	Awake but deeply relaxed.
Theta	3Hz-8Hz	Light sleep or extremely relaxed state.
Delta	0.2Hz-3Hz	Deep, dreamless sleep.

It is well known that sound and light directly affect the brain through a complex neural process called Brainwave Entrainment. Brainwave Entrainment refers to the brain's electrical response to rhythmic sensory stimulation, such as pulses of sound or light. When the brain is given a stimulus, it emits an electrical charge in response. Entrainment is a principle of physics. It is defined as the synchronization of two or more rhythmic cycles. The human brain is the example of complex system to which rules of entrainment also apply. [4]

III CORRELATIONS

Lewis B. Hainsworth, Electrical engineer of Western Australia seems to be the first researcher to recognize the

relationship of brainwave frequencies to the naturally circulating rhythmic signals, known as Schumann's resonances. It was in back 1975. that Hainsworth was among the first to suggest that human health is linked with geophysical parameters by way of the naturally occurring Schumann's ELF. His hypothesis identified naturally occurring features which determine the frequency spectrum of human brain-wave rhythms: The frequencies of these electromagnetic signals, circulating in the electrically resonant cavity bounded by the Earth and the ionosphere, have governed or determined the 'evolution' or development of the frequencies of operation of the principal human brainwave signals. In particular, the alpha rhythm is so placed that it can in no circumstances suffer an extensive interference from naturally occurring signals. Hainsworth concluded that the frequencies of human brain waves evolved in response to these signals. If his hypothesis is correct, conditions for evolutionary changes in human brain-wave patterns have now been established. Furthermore, variations in these patterns can produce mild to disastrous health and behavioral changes. [5]

When a person is deeply relaxed, slow rhythmic sine-wave patterns can be detected in both the EEG and the heart/aorta resonating oscillator in the 7-8 Hz range. Resonance occurs when the natural vibration frequency of a body is greatly amplified by vibrations at the same frequency from another body. Oscillators alter the environment in a periodic manner. Thus, standing waves in the body, whether during meditation/relaxation or not, can be driven by a larger signal. Progressively amplified wave-forms, created by resonance, result in large oscillations entraining other circuits in the body tuned to those frequencies. A hierarchy of frequencies thus couples our psychophysical selves to the harmonic frequency of the electrical charge of the Earth, which naturally pulses at the same frequencies. This is hardly a coincidence, as we are adaptive products of our environment.

One of the foremost researchers in this field is Dr Wolfgang Ludwig, who has been investigating Schumann Resonance and its place in nature for many years, writes in his book *"Informative Medizin"* about the research carried out by E. Jacobi at the University of Düsseldorf. In this research, it is showed that the one sided use of Schumann wave simulation without the geomagnetic signal caused serious health problems. On the other hand, the absence of Schumann waves creates a similar situation. Professor R.Wever from the Max Planck Institute for Behavioral Physiology in Erling-Andechs, built an underground bunker which completely screened out magnetic fields. Student volunteers lived there for four weeks in this hermetically sealed environment. Professor Wever noted that the student's circadian rhythms diverged and that they suffered emotional distress and migraine headaches. After only a brief exposure to 7.8 Hz (the very frequency which had been screened out), the volunteers health stabilized again. The same complaints were reported by the first astronauts and cosmonauts, who, out in space, also were no longer exposed to the Schumann waves. Now modern spacecrafts are said to contain a device which simulates the Schumann waves. [6]

Some research has suggested that the frequency of the basic Schumann's resonance has recently been rising in value, possibly threatening the whole biosphere, human welfare and our evolutionary future. Their claims are based on the principle that all biological processes are a function of electromagnetic field interactions. There are number of scientist who firmly believe that the frequencies used in mobile telephony are harmful for humans because they interfere with the natural frequencies of earth magnetic field and thus changing it. If organisms do in fact respond to, and perhaps depend on, electromagnetic fields as weak as that produced by Schumann resonance at 0.22-1.12 mV/m, this is of major significance for the development of present and future wireless technologies.

IV Conclusion

Although it has been more than a half of century since they've been discovered, the knowledge on Schuman waves still haven't been fully exploited. The reason for that is mainly because of the complicated measuring methods. On the other side, the theories that link them with certain impact on living organisms are very controversial and reasonably doubtful in the first place. Nevertheless, if there is even a slightest chance that these theories might be true, it could be a breakthrough in science. For that reason, it is necessary to conduct more researches on this topic by using experiences both from the field of engineering and medicine. Only in this way we can discover whether the magnetic field of Earth and the living beings on it are synchronized in some way. If that could be the case, it is necessary to know whether we could harm that harmony and thus endanger our biological existence.

References

[1] Glenn Elert, "Standing Waves", The Physics Hypertextbook, 1998-2008, *http://hypertextbook.com/physics/waves/*

[2] Kristian Schlegel and Martin Füllekrug, "50 Years of Schumann Resonance", *Physik in unserer Zeit*, 256-26, 2002

[4] Transparent Corp., "Neuro Programmer 2™ Documentation", The Transparent Corporation™, 2004-2007

[3] Tony Smith, "Schumann Resonances, Geomagnetic Reversals, and Human Brain States", *http://www.valdostamuseum.org/*

[5] Richard Alan Miller and Iona Miller, "The Schumann Resonances and Human Psychobiology", *Nexus Magazine*, Volume 10, Number 3 (April-May 2003) O Box 30, Mapleton Qld 4560 Australia

[6] Ingrid Pastl-Dickenson, "Earth breathing", South West London, 2003, *http://www.earthbreathing.co.uk*

Mechanically scanned LED display

D. Vučković, N. Jordanov

Abstract – Description of a system that uses high-speed movement of one vertical array of LEDs to create an illusion of cylindrical display, which is furthermore used for displaying text.

I. INTRODUCTION

It is very well known that high-speed rotation of a single dot can create a circle. Putting several dots in horizontal plane and rotating them will make an illusion of a cylinder. By turning those dots "on" and "off" one can make different patterns on the cylinder. The following description gives an example of a device that uses this technique in order to write letters on the cylinder like canvas [1].

Description of the device will be divided into three parts:

- Mechanical construction
- Electronics and power transfer
- Software

II. MECHANICAL CONSTRUCTION

The system can be divided into two segments: base (stationary part) and propeller.

The base of the system holds the entire structure and houses the motor, primary coil, square wave generator and power supply.

In order to create an illusion of cylindrical display a motor with high rpm is needed. On the other hand, the motor must be silent. In order to fulfill both of these demands a CPU cooler fan is selected to be the working horse of the entire system. The only drawback of this type of motor is that it can't be loaded a lot, so the propeller must be very light.

The propeller is about two times longer than the diameter of the CPU fan. It is made of copper clad laminate so it is strong enough and it is also used for making the PCB for electronics placed on the propeller. On the end of propeller there is a vertical LED array.

D. Vučković and N. Jordanov are with Department of Microelectronics, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia,

E-mail: d.vuckovic85@gmail.com, nenadms5@gmail.com

In Fig.1 there is a picture of the device while operating. The letters seam to float because the propeller is not visible due to high rotation speed.

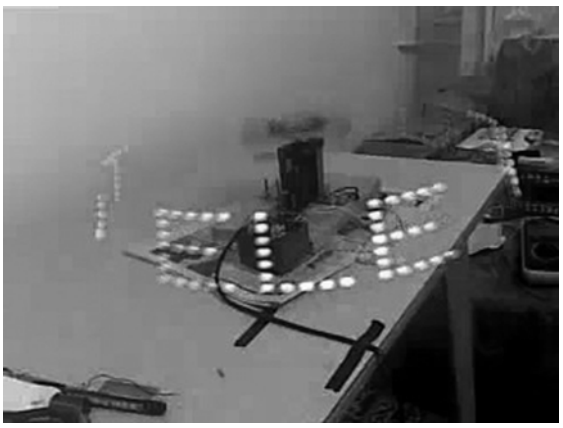


Fig 1. Picture of the device while operating

III. ELECTRONICS AND POWER TRANSFER

Well-designed electronics are necessary in order to achieve good performance of this system. Block diagram of the entire system is shown in Figure 2. All electric parts can be arranged into two main blocks: stationary electronics (base) and propeller. Bases primary objective is to transfer power from power source to the rotating propeller with minimal power loss. Secondary objective of the base is to provide stabile power supply for the motor. This is very important because if the power supply is unstable the motor doesn't provide a constant number of revolutions per minute (rpm) which is an imperative for getting a stable display. The base consists of transformer block, AC/DC converter block, two voltage regulators, square wave generator and a cylindrical primary coil that is used for transferring power to the propeller. The transformer lowers 220V AC voltage to 15V. Then that voltage is fed to two voltage regulators: one for the motor and one for the square wave signal generator. Motor is a standard CPU fan so it requires 12V. On the other hand square wave generator is working on 3V. All electronics that are situated on the propeller can be divided into five blocks: secondary coil, AC/DC converter, Hall sensor, Microcontroller block (MCU) and LED array. Secondary coil is used to collect energy from the primary coil, which is stationary, and then induced current is fed to AC/DC converter.

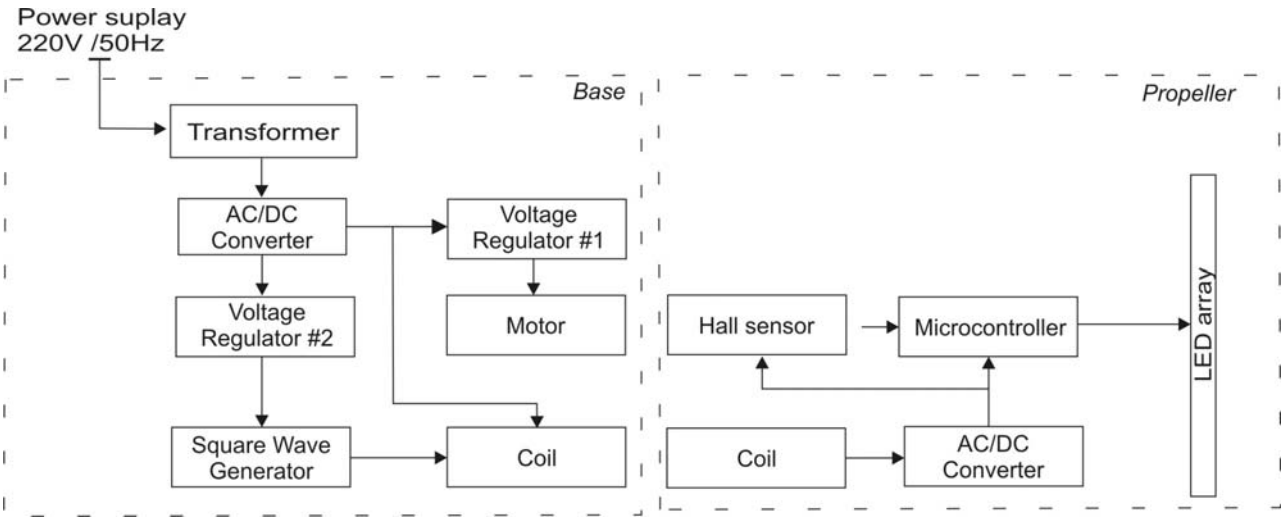


Fig 2.Block diagram of the entire device

After conversion to DC the current is going through a voltage stabilizer which provides stable voltage for MCU. Microcontroller is used to turn the LED array on and off at specific times. Hall sensor is used to trigger the MCU on every beginning of circle. A stationary magnet is located beneath the propeller so the sensor triggers the MCU always when the propeller above the magnet. Operation of the MCU will be discussed in more detail in the following sections.

A. Power transfer

Power transfer is done using two concentric cylindrical coils that have an air gap between them. Windings ratio between the primary and secondary coil is 1:1. The principle behind power transfer is Inductive coupling of two coils. Primary coil is driven by square signal generator in order to induce electric current in the secondary coil. In order to maximize the power transfer frequency of the square wave is 1 kHz. This particular system is capable of producing 13V AC on the output of the secondary coil when 15V signal is applied to the primary coil. All electronics on the propeller are using 5V and consuming 100mA when all LEDs are active.

IV. SOFTWARE

The heart of the system is the software that is driving the LED array. The task of software is rather complex. The MCU must keep track of the current location of the propeller, calculate the width of the "on" stage of the LEDs and shift the text in a way so it seems the displayed text is moving.

The crucial part is knowing the precise location of the propeller. Otherwise, stable display can't be accomplished. Even the minimal deviation between the point where the LED array should be turned on in two consecutive revolutions, will result in flickering.

A. Program explanation

Although PICs don't support multithreading this program can be analyzed as some sort of multithreaded application. There are two main threads:

- Measuring the duration of revolution
- Tracking propeller location and setting LED output

On the beginning of each revolution a timer is being started so it measures time before the beginning of the next revolution. When the sensor crosses the magnet, it sets its output to low signaling the MCU the start of the revolution. After the next revolution starts, the timer is stopped and the measured time is divided into a number of even time segments. These segments define how long will LED array be held active. Impression of wider LED lines is done by keeping LEDs longer active. When the calculation is complete the timer is being reset and started again. Duration of the revolution must always be monitored because it can vary.

Tracking propeller location is done using time segments previously described. Each segment represent one vertical "line" on the cylindrical display. By counting the number of elapsed time segments one can always know the location of the propeller. When the precise location is known specific pattern is sent to LED line. As propeller rotates it moves form one line to another, thus building up the display canvas. By controlling the patterns that are fed to LED array, one can alter what is displayed.

Rotating contents of the display is done simply by shifting what is displayed for one line on every beginning of the revolution.

Changing the contents of display is done rather simple by connecting the MCU to the computer using a special board that uses RS232 protocol to communicate with MCU. Next version of this device, which is currently under development, will have infrared communication between computer and MCU. Infrared communication will simulate RS232 protocol making it rather easy to use. This will allow the system to change what is displayed while the system is running allowing the display to become more versatile.

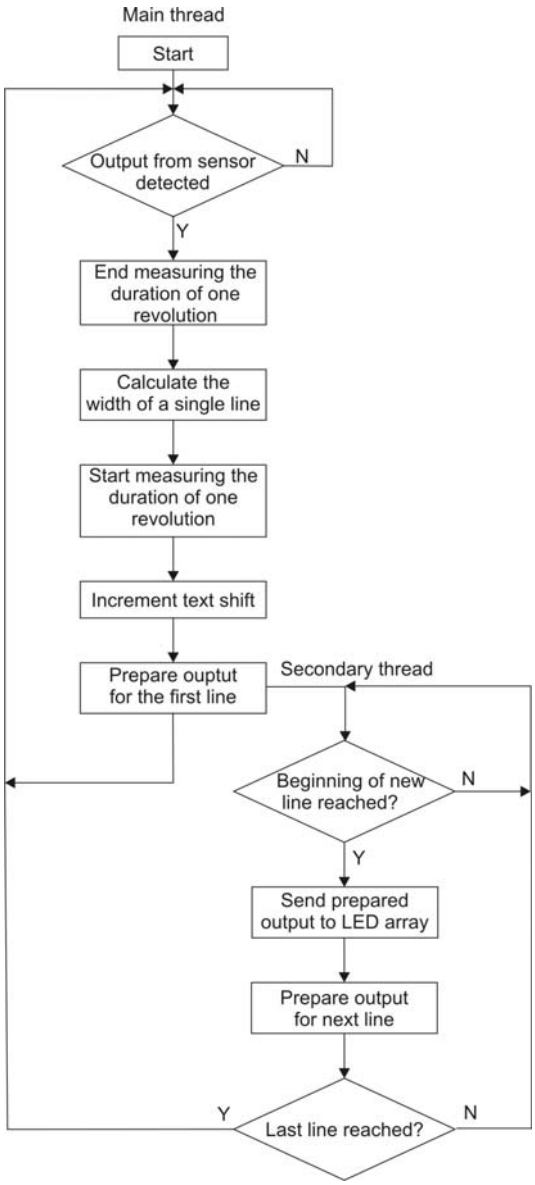


Fig 3. Pseudo-algorithm of program flow

B. Algorithm

The algorithm is as showed in Figure 3. This is some sort of pseudo algorithm because the first thread can interrupt execution of the secondary thread at any given moment.

In a system that is working with a motor that has stable rpm there will be no gap between the last line and the beginning of the next revolution allowing full utilization of display. In that case, the secondary thread will stop execution and there will be no interruption from the first thread.

IV. CONCLUSION

High-speed moving LEDs in conjunction with MCU and careful time management can make a simple but efficient display, which has a viewing angle of 360°. This kind of display is highly portable and has low power consumption and it can have a variety of applications in the first place for displaying advertisements and information.

REFERENCES

[1] Web site that contains another example of mechanically scanned LED display; www.bobblick.com/techref/projects/propclock/propclock.hhht

Mehatronički 3D skener
Principi i upravljanje

M. Božić

Abstrakt - 3D skener je mehatronički sklop za sakupljanje podataka o prostoriji u kojoj je postavljen. Obradom sakupljenih podataka računarom, stvara se geometriska slika prostorije u tri dimenzije. Tema rada je upoznavanje sa principima i upravljanjem mehatroničkim sklopom 3D skenera.

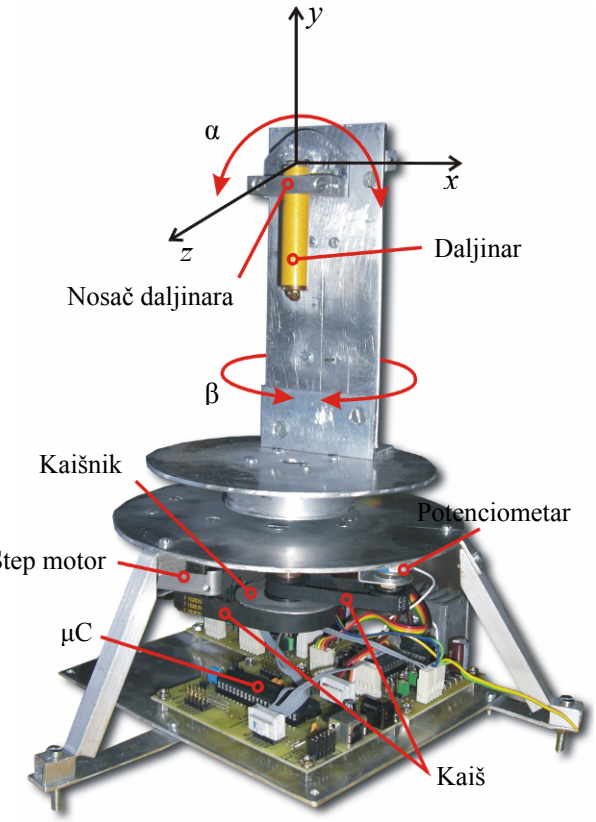
I. Uvod

Vrlo često je potrebno imati tačne podatke o geometriji neke prostorije. Prikupljanje podataka, odnosno niza koordinata tačaka u prostoru, može obaviti čovek merenjem. Preciznost dobijene slike zavisi od broja merenja, koje treba vršiti u raznim pravcima i pod različitim uglovima, što u nekim slučajevima može biti problem. Da bi vreme potrebno za skeniranje bilo što kraće, osmislili smo odgovarajući mehatronički sklop (slika 1), koji će sakupiti podatke o geometriji prostorije. U narednim poglavaljima će biti opisani svi elementi mehatroničkog sklopa ponaosob. Za sada je dovoljno reći da se sastoji od dva step motora, optičkog daljinara i mikrokontrolera koji upravlja radom uredjaja.

Najčešći problem koji se javlja pri merenju je nesavršenost izrade mehaničkog sklopa, čije se rešenje nalazi u većem broju merenja za svaku tačku i uzimanja srednje vrednosti merenja kao i u kalibraciji osa rotacije, o čemu će biti reči u nastavku teksta. Takodje se pri skeniranju nepravilnih površina ili prostorija u kojima se nalazi više objekata javlja problem medjusobnog preklapanja površina, koji se može otkloniti skeniranjem sa različitih pozicija. Ovaj problem predstavlja intersantnu temu za razmatranje i biće obradjen u nekom od narednih radova.

II. Princip rada

Rad skenera zasnovan je na merenju i prikupljanju rastojanja karakterističnih tačaka od početka koordinatnog sistema. Za dobijanje trodimenzionalne geometriske slike prostorije potrebno je da se skeniranje tačaka obavlja u više pravaca pod različitim uglovima, što zahteva rotiranje uredjaja za merenje. Rotacija se obavlja pomoću dva step motora, a svakoj snimljenoj tački se pridružuju odgovarajući uglovi α i β, (ose rotacije i uglovi su prikazani na slici 1). Upravljanje motorima se vrši pomoću odgovarajućih drajvera i mikrokontrolera koji spregnut sa računarom, a na osnovu dobijenih rastojanja izračunavaju se koordinate svake tačke, o čemu će biti reči u narednom odeljku.



Sl. 1. Mehatronika sistema i ose rotacije

A. Odredjivanje kordinata skenirane tačke

Podaci potrebni za iscrtavanje prostorije su koordinate skeniranih tačaka, koje se računaju na osnovu dobijenih rezultata merenja, trigonometriskim jednačinama. Princip računanja koordinata i uglovi pod kojima se meri rastojanje tačke su prikazani na slici 2.

Predpostavimo da se daljinar nalazi u centru koordinatnog početka, slika 1. Rastojanje tačke koje se

M. Božić is student at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia
E-mail: mbozic84@gmail.com

meri od početka trodimenzionalnog koordinatnog sistema označeno je sa c, dok je ugao pod kojim se meri rastojanje u odnosu na X - Z ravan označen sa α, slika 2. Y koordinata tačke se izračunava na osnovu jednačina:

$$a = c\sin\alpha$$

$$y = a$$

X i Z koordinata tačke računaju se na osnovu ugla β, koji predstavlja ugao pod kojim daljinar seče X - Z ravan i rastojanja b, koje se računa kao:

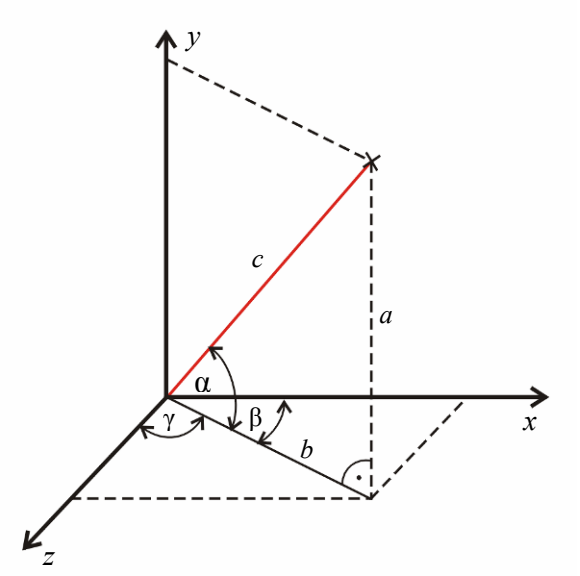
$$b = c\cos\alpha$$

pa je:

$$x = b\cos\beta$$

$$z = b\sin\beta$$

što je prikazano na slici 2.



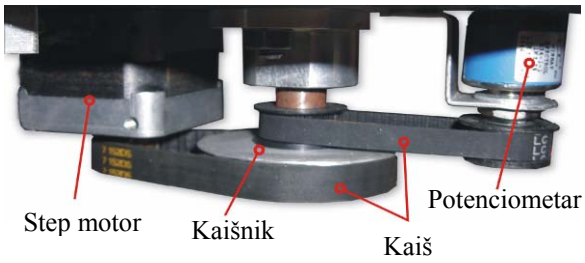
Sl. 2. Odredjivanje kordinata skenirane tačke

III. Mehatronika sistema

Osnovni elementi potrebni za funkcionisanje skenera su dva step motora, daljinar i upravljačka elektronika. Svaki motor je kaišem i kaišnikom spregnut sa osovinom višeobrtnog potenciometra, pa je napon na izlazu potenciometra proporcionalan uglu zaokretanja osovine motora sa kojim je spregnut. U narednim poglavljima će biti opisani svi elementi sklopa i objašnjen princip pozicioniranja pomoću potenciometra.

A. Pogon prve ose

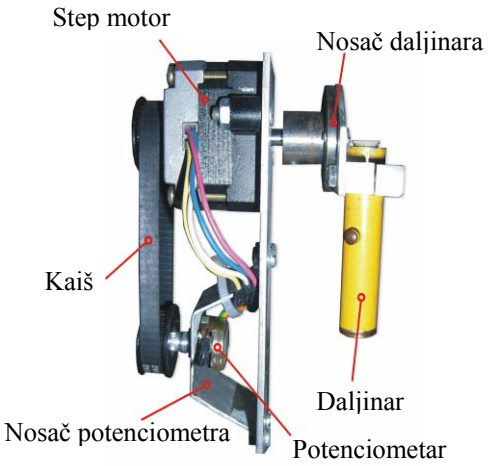
Mehanika za rotiranje sistema oko Y ose je prikazana na slici 3. Iskorišćeni motor ima 200 koraka za 360° rotacije, tako da je minimalni ugao za koji se sistem može zaokrenuti pri jednom koraku 1.8° u full step modu, odnosno 0.9° u half step modu. Pomoću kaiša i kaišnika motor je spregnut sa potenciometrom i osovinom koja vrši zaokretanje sistema po Y osi, za ugao β, slika 3.



Sl. 3. Pogon prve ose

B. Pogon druge ose

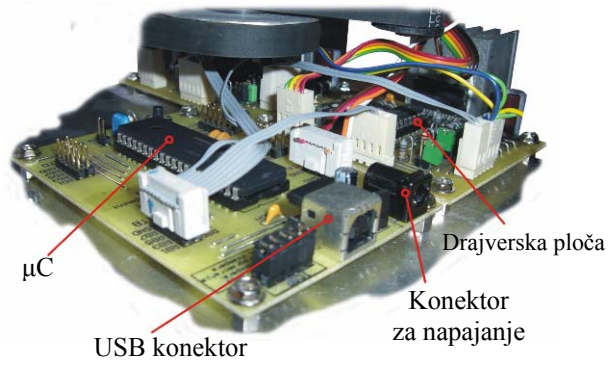
Rotiranje daljinara za ugao α, realizovano je takodje step motor. Iskorišćeni motor ima 400 koraka za 360° rotacije tako da je minimalni ugao za koji se sistem može zaokrenuti pri jednom koraku 0.9° u full step modu, odnosno 0.45° u half step modu. Informaciju o uglu zaokretanja osovine motora α daje potenciometar, a sprega sa njim je ostvarena pomoću kaiša i kaišnika, slika 4. Daljinar je montiran direktno na osovinu motora *nosačem daljinara*, slika 4, pa je ugao za koji se zaokrene osovina motora takodje i ugao pod kojim se meri rastojanje od odredjene tačke.



Sl. 4. Pogon druge ose

C. Upravljačka elektronika

Upravljačka elektronika se sastoji iz mikrokontrolerske ploče i drajverskih ploča. Drajverske ploče upravljaju radom step motora, a brzinom i smerom okretanja upravlja mikrokontrolerska ploča. Pozicija u kojoj se sistem nalazi odredjuje se na osnovu očitane vrednosti napona sa potenciometara. Za upravljanje drajverima iskorišćen je *Microchip* - ov PIC18F4550 mikrokontroler, koji je sa računarom spregunt putem USB porta. Sa računara se zadaje pozicija, odnosno uglovi pod kojima treba da se izvrši merenje rastojanja, a µC odgovara vraćanjem izmerene vrednosti. Drajverske ploče su realizovane parom integrisanih kola L297(translator), i L298(izlazni stepen). Upravljačka elektronika je prikazana na slici 5.



Sl. 5. Upravljačka elektronika

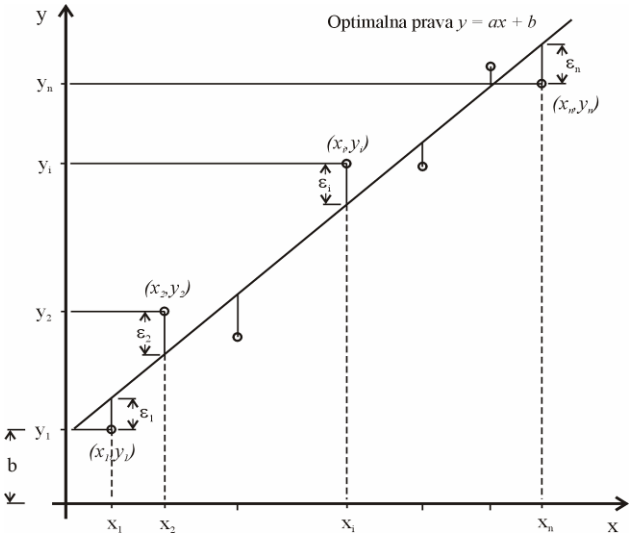
D. Pozicioniranje pomoću potenciometra

Jedan od načina da se odredi pozicija sistema u prostoru jeste sprezanje pokretnih delova sistema sa potenciometrom. Napon sa potenciometra je proporcionalan uglu rotacije i njegovim očitavanjem se dobija informacija o trenutnoj poziciji daljinara. Greške koje se mogu javiti pri očitavanju pozicije najčešće su uzrok nesavršenosti izrade mehaničkih delova sklopa, kaiševa i kaišnika. Da bi se uticaj grešaka pri merenju smanjio, sistem je pre upotrebe potrebno kalibrisati, odnosno potrebno je linearizovati karakteristiku potenciometra. Jedan od načina za linearizaciju je *metoda najmanjih kvadrata*.

Metoda najmanjih kvadrata je opšta jednoznačna metoda za odredjivanje optimalne prave za dati komplet merenih podataka. Na slici 5 prikazan je skup od n merenih tačaka (x_i, y_i) razmeštenih u pravouglom kordinatnom sistemu. Izmedju njih je provučena prava koja predstavlja optimalnu pravu. Jednačina ove prave je:

$$y = ax + b$$

gde je *a* kojeficijent nagiba prave, a *b* odsečak na ordinati koji treba odrediti. Ova jednačina se u literaturi još zove jednačina linearne regresije, a optimalna prava linija ili prava regresije.



Sl. 6. Optimalna prava u metodi najmanjih kvadrata

Predpostavimo da su ulazne veličine X tačne, a da greške postoje kod izlazne veličine, Y. U našem slučaju na X osi se nanose vrednosti zadatog ugla osovine potenciometra, dok se na Y osi očitava napon na potenciometru. Greške su predstavljene vertikalnim rastojanjem izmedju merene tačke i optimalne prave i na slici su obeležene sa ε_i, a sa vrednošću:

$$\varepsilon_i = y_i - (ax_i + b)$$

Konačna rešenja za kojeficijente tražene optimalne prave su:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i^{\,2} - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Uredjaj se kalibriše po prvom uključenju, a dobijene vrednosti se pamte u računaru. Ukoliko je potrebno kalibracija se može vršiti proizvoljan broj puta.

IV. Zaključak

Potreba za poznavanjem gemoetrije prostorije je česta pojava, a jedan od načina da se podaci o geometriji sakupe je merenjem rastojanja ivica prostorije u odnosu na koordinatni početak. Merenje može obaviti čovek ali vreme potrebno skeniranje, koje zavisi od veličine prostorije i broja merenja, može biti voma dugo. 3D skener omogućava da se skeniranje izvrši u što kraćem vremenskom periodu i sa znatno većom preciznošću. Kao što je već rečeno može se javiti problem skeniranja površina koje se medjusobno preklapaju. U tom slučaju je potrebno da se skener, tj. koordinatni sistem, pomera i vrši skeniranje iz raznih uglova. Ukoliko bi se skener postavio na telo robota proces pomeranja i skeniranja iz različitih uglova bi se znatno pojednostavio.

Jedna od namena sistema bi bila skeniranje teško pristupačnih prostorija, kao što su rudnici, tuneli i sl.

Literatura

[1] *http://en.wikipedia.org/wiki/Stepper_motor*, Avgust 2008

[2] Julio Sanchez, Maria P. Canton, *Microcontroller programming - The microchip PIC*, Decembar 2007

[3] *http://dominis.phy.hr/~kvurnek/SAMP/SEMINAR/zavrsni.html*, Avgust 2008

Abstract - 3D Space Scanner is a mechanical - electronic device for gathering data about the room it's located in. By processing gathering data with proper program, a 3D image of the room is assembled in the computer. This paper is primarily focused on system's mechanic and it's management.

3D Space Scanner System's mechanic and management

Miroslav Božić

Collaborative Work Organizer

A. Prokić, D. Todorović

Abstract – Collaborative Work Organizer is information system designed especially for supporting company's employs organization of work time, and helping find the best time for meetings between the people from the same department or from the whole company. In this article we described the system's architecture, as well as the implementation of the system. This information system consists of database, *Web Server*, thin PC client and wireless mobile application as personal client.

I. UVOD

Collaborative Work Organizer je informacioni sistem koji je dizajniran tako da olakšava organizaciju radnog vremena zaposlenih u nekoj kompaniji. Ova tema koja je implementirana u gotov proizvod je vrlo aktuelna jer zadovaljava potrebe savremenog poslovnog čoveka koji treba da bude informisan o najnoviijim događajaima u bilo kom trenutku u na bilo kom mestu da se nalazi i pruža mogućnost da mobilni korisnik sa bilo kog mesta u bilo koje vreme ažurira informacije na serveru. Na ovaj način ovaj sistem omogućava lakšu komunikaciju u oba smera i ažurnost podataka, što je vrlo bitno za savremenog poslovngo čoveka.Aplikacija se sastoji iz dva dela. Jedan deo aplikacije se izvršava na serveru, a drugi deo aplikacije na mobilnom uređaju.

Osnovne funkcionalnosti koje podržava ovaj sistem su: vođenje kalendara aktivnosti svakog registrovanog člana posebno; upravljanje kalendarom aktivnosti jednog tima u okviru kompanije; upravljanje kalendarom aktivnosti cele kompanije i upravljanje aktivnostima preko mobilne aplikacija koja je posebno projektovana kao podrška ovom sistemu. Sistem se sastoji iz baze podataka u kojoj se nalaze sve informacije o kompaniji, svakom od zaposlenih, timovima u okviru kompanije, koji su predstavljeni kao virtuelna udruženja, kao i kalendari aktivnosti svih entiteta u kompaniji. Na *Web Server*-u se izvršava *Web* aplikacija koja implementira prezentacionu logiku, ali još bitnije, omogućava korisnicima da unesu svoje aktivnosti u kalendar i omogućava vođama timova u okviru kompanije da zakažu sastanak u periodu koji najbolje odgovara svim članovima tima.

Prezentaciona logika je odvojena od aplikacione logike upotrebom CSS-a, kako bi se omogućilo različitim kompanijama da daju svoj lični pečat aplikaciji. Još jedan bitan deo sistema je aplikacija koja je posebno razvijena za mobilne uređaje i koja se putem bežične mreže povezuje na server sa kojim vrši sinhronizaciju podataka datog korisnika. Svaki od pomenutih delova sistema će biti detaljnije opisan u naradnim poglavljima.

Deo aplikacije koji se izvršava na serveru predstavlja sistem koji će obezbediti jednom timu ili jednoj organizaciji, čiji je osnovni način komunikacije elektronskog tipa i koja ne funkcioniše iz zajedničkih prostorija već su članovi udaljeni, lako, efikasno i transparentno organizovanje sastanaka ili drugih aktivnosti koje podrazumevaju prethodno uspostavljanje vremenskih okvira.

Deo aplikacije koji se izvršava na mobilnom uređaju omogućava korisniku vođenje evidencije o dnevnim aktivnostima. Korisnik može da unese u kalendar više tipova aktivnosti (sastanak, rođendan i sl), vreme, trajanje, ponavljanje, kao i vreme kada alarm treba da se oglasi. Po startovanju, aplikacija prikazuje događaje zakazane za taj dan i aktivira alarm u zakazano vreme za svaki od događaja.

Da bi podaci uvek bili ažurni, bilo da se radi o podacima na mobilnom uređaju ili o podacima na serveru omogućena je sinhronizacija aktivnosti preko mreže. Podaci koji se unose sa mobilnog uređaja sinhronizuju se sa podacima na severu i podaci koji se unose sa servera sinhronizuju se sa podacima na mobilnom uređaju. Sinhronizacijom podataka se smanjuje potreba za neprekidnim prenosom podataka preko spore i nepouzdane bežične mreže.

Komunikacija između servera i klijenta je asinhrona i može biti inicirana i od strane klijenta i od strane servera. Ova funckionalnost je posebno važna zbog toga što na ovaj način korisnici koji koriste mobilnu aplikaciju uvek mogu biti obavešteni onastalim događajima za koje su zainteresovani.

Za projektovanje i implementaciju ovog sistema korišćena je RUP metodologija projektovanja softvera.

U narednim poglavljima je detaljno opisana arhitektura implementiranog sistema koja je opisana odgovarajućim UML dijagramima. Zatim je dat opis relacione šeme i konkretne implementacije baze podataka i opis implementacije serverske i klijentske *Web* aplikacije, kao i opis mobilne klijentske aplikacije.

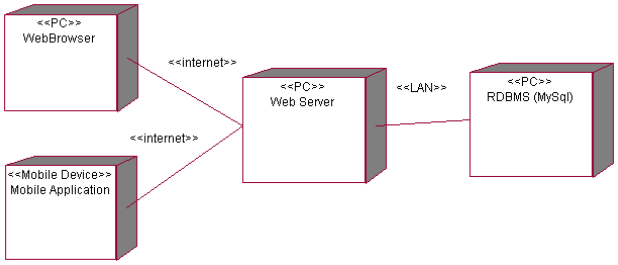
II. ARHITEKTURA SISTEMA

Pod arhitekturom sistema podrazumevamo osnovne komponente od kojih se informacioni sistem sastoji, njihove funkcionalnosti i komunikacioni protokoli koji ih povezuju. Na slici 1. su prikazane sve kompomente iz kojih se sistem sastoji. RDBMS i *Web Server* se

A. Prokić and D. Todorović are students at the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia

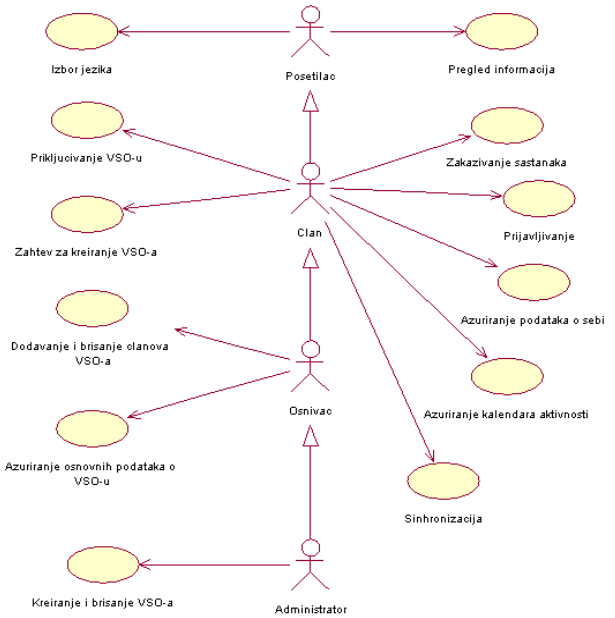
E-mail: anaprokic@gmail.com, darkotodorovic@gmail.com

izvršavaju na istom računaru ili različitim računarima. Ukoliko se izvršavaju na različitim računarima, onda je komunikacija između istih ostvarena putem LAN-a. *Web Server*u se pristupa ili sa PC računara, korišćenjem nekog od standardnih browsera ili iz mobilne aplikacije. U oba slučaja se komunikacija vrši preko interneta. Kako je mobilna aplikacija pisana za *Java* platformu, onda se ona uglavnom izvršava na mobilnim telefonima pa je pristup internetu ostvaren preko GPRS-a.



Slika 1. Kompomente koje čine informacioni sistem CWO njihov razmeštaj i protokoli koji su korišćeni u komunikaciji.

Potrebne komponente, njihov razmeštaj i same funkcionalnosti ovog sistema su određene zahtevima koje sistem treba da ispunjava. Jedan od vidova predstavljanja zahteva je kroz precizno definisanje korisnika sistema i slučajeva korišćenja. Osnovni UML dijagram koji prikazuje korisnike i slučajeve korišćenja CWO portala prikazan je na slici 2.



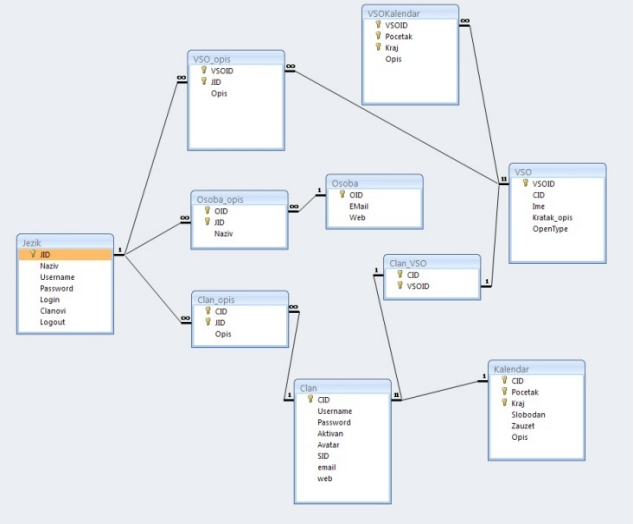
Slika 2. Slučajevi korišćenja CWO sistema kao i akteri koji učestvuju u interakciji sa sistemom.

Korisnici koji su tipa Član, Osnivač i Administrator mogu pristupiti sistemu ili sa PC računara preko nekog *Web Browser*-a ili iz mobilne aplikacije. Ukoliko pristupaju sistemu preko mobilne aplikacije nisu svi slučajevi korišćenja dostupni već samo oni koji se odnose na ažuriranje informacija u kalendaru aktivnosti.

Svi registrovani korisnici portala su članovi nekog virtuelnog udruženja. Virtuelna udruženja predstavljaju način okupljanja članova nekog tima koji imaju zajedničke sastanke. Svako virtuelno udreuženje ima svoj kalendar aktivnosti koji popunjava osnivač virtuelnog udruženja i sve aktivnosti virtuelnog udruženja se automatski prosleđuju svim članovima tog udruženja i zapisuju u njihove kalendare. U nastavku će biti detaljnije opisana implementacija sistema.

A. Implementacija baze podataka i serverskog dela aplikacije

Za *Web Server*, na kome se izvršava serverska aplikacija, korišćen je *Apache Server* sa ugrađenom podrškom za PHP 5.x. DBMS u okviru koga je implementirana baza podataka je MySql.



Slika 3. Relaciona šema baze podataka

Na slici 3. je prikazana relaciona šema baze podataka koja je korišćena za implementaciju ovog informacionog sistema. U bazi podataka se čuvaju osnovne informacije o kompaniji koja koristi ova sistem, informacije o svakom članu, kalendar aktivnosti za svakog člana ponaosob, informacije o svakom virtuelnom udruženju, kalendar aktivnosti svakog virtuelnog udruženja kao i kalendar aktivnosti cele kompanije. Informacije iz ove baze podatak se koriste prilikom određivanja vremena sastanka u okviru jednog tima.

Prezentaciona i aplikaciona logika ovog sistema je ostvarena kroz nekoliko php skripti. Komponente ovog informacionog sistema su PHP skripte čiji će pregled biti dat po arhitekturnim slojevima. Za ilustraciju će biti korišćeni UML dijagrami komponenti, ali i dijagrami klasa. U slučajevima gde je PHP skript prikazan kao klasa atributi predstavljaju ulazne podatke koji se uzimaju iz GET ili

POST dela HTTP poruke, dok metodi predstavljaju funkcije definisane u okviru skripta.

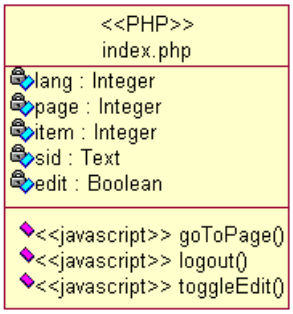
Dizajn korisničkog interfejsa je obuhvaćen dvema komponentama koje su prikazane na slici 4.

Slika 4. Komponente koje čine korisnički interfejs

Komponenta **index.php** je implementira stranicu portala čiji sadržaj može da varira od parametra koji joj se proslede u zahtevu.

Komponenta **main.css** predstavlja opis stilova za pojedine HTML elemente koji se javljaju na različitim stranicama.

Parametri koji utiču na izbor i jezik za prikaz stranice ilustrovani su sledećim dijagramom klase:

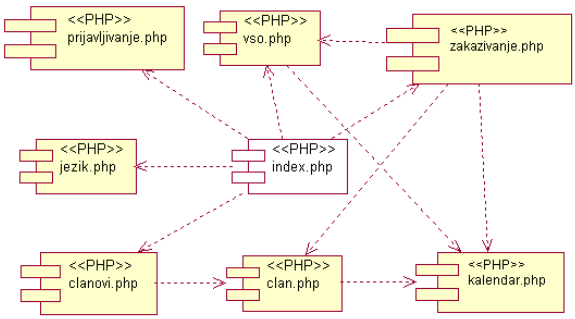


Slika 5. Klasni dijagram PHP skripta koji je zadužen za prezentovanje sadržaja

Značenje atributa je sledeće:

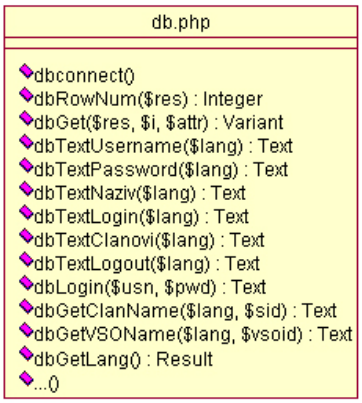
- lang – identifikator jezika (JID iz tabele Jezik)
- page – interni identifikator stranice (VSO, Clan, Kalendar)
- item – identifikator stavke koja se detaljno prikazuje na stranici (određeni član ili VSO)
- sid – identifikator sesije kada je korisnik ulogovan
- edit – definiše da li je uključeno uređivanje

Komponente aplikacione logike koje realizuju domen problema se uključuju isključivo preko **index.php** komponente korisničkog interfejsa. Na taj način zadržavaju sva podešavanja stila definisana u ovom skriptu. Na sledećem dijagramu su prikazane komponente ovog sloja i njihove međusobne zavisnosti:



Slika 6. Komponente aplikacione logike

Pristup bazi podataka je u potpunosti zatvoren u funkcije koje su definisane u okviru PHP skripta **db.php**. Pomenuti skript se uključuje na početku index.php-a, tako da su sve funkcije za pristup podacima dostupne svim komponentama. Na slici 7. prikazan je UML dijagram klase u kome su pobrojane funkcije za pristup podacima iz baze, kao i funkcije za ubacivanje, brisanje i promenu podataka u bazi.



Slika 7. UML dijagram klase u kojoj su pobrojane funkcije za pristu podacima iz baze podataka

B. Mobilni klijent

Aplikacija koja se izvršava na mobilnom klijentu pruža korisniku mogućnost unosa različitih aktivnosti kao i njihov prikaz. Međutim, prikazuju se samo oni događaji koji su zakazani za današnji dan sa mogućnošću aktiviranja alarma u zakazano vreme za svaki od događaja.

Podaci koji se unose sa mobilnog klijenta mogu biti prebačeni na server i obrnuto čime se postiže ažurnost podataka na obe strane bez neprestalne internet konekcije koja može da bude nepouzdana i nedostupna u pojedinim trenucima. Ažurnost podataka je omogućena pomoću mehanizma za sinhronizaciju koji je implementiran u okviru ove aplikacije.

Aplikacija na mobilnom uređaju se sastoji iz tri glavna dela, a to su interfejs, persistent storage, odnosno baza za smeštanje podataka i mehanizam za sinhronizciju.

Izgled mobilne aplikacije po startovanju aplikacije vidi se na slici 8.

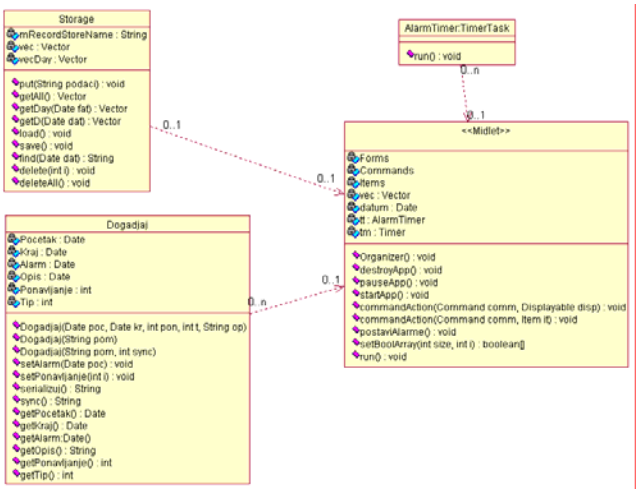
Slika 8 Izgled mobilne aplikacije nakon startovanja.

Mobilna aplikacija se sastoji iz sledećih klasa: Dogadjaj, Storage, AlarmTimer i Organizer klase koja implementira interfejs MIDlet i koja predstavlja ulaznu tačku mobilne aplikacije. U ovoj klasi je implementiran interfejs kao i deo aplikacione logike koja podržava interfejs. U okviru ove klase je implementiran i sloj za sinhronizaziju podataka iz lokalne baze podataka sa podacima koji se nalaze na udaljenom serveru. Sinhronzacija se vrši u okviru funkcije **run()** koja se izvršava u posebnoj niti.

Sinhronizacija se obavlja u dva koraka:

1: Prvo se proverava da li na serveru postoji korisnik sa unetim korisničkim imenom i šifrom i ukoliko postoji, skripta (mobilelogin.php) koja se nalazi na serveru vraća ID tog člana.

2: Ako je mobilna aplikacija u prethodnom koraku dobila ID člana ona na osnovu tog člana poziva skriptu sync.php kojoj prosleđuje sve podatke iz lokalne baze na mobilnom uređaju u okviru post zahteva, skripta se serveru ispituje sve konfikte i ažurira bazu na serveru, a zatim vraća novu bazumobilnom uređaju.



Slika 9. Klasni dijagram mobilne aplikacije

III. ZAKLJUČAK

Collaborative Work Organizer je sistem koji omogućava efikasnu organizaciju vremena, brzu komunikaciju među zaposlenima i ažurnost kalendara aktivnosti u okviru neke organizacije. Kako je ovo klijent server aplikacija, sa posebno projektovanim mobilnim klijentom, korisnici mogu da pristupe svim potrebnim podacima sa bilo kog mesta koje ima pristup internetu i tako omogućava poslovnim korisnicima da uvek imaju ažurne informacije, kao i da sa ostalim korisnicima podele bitne informacije u svakom trenutku. Na tržištu postoje rešenja koja nude iste ili slične funkcionalnosti kao što su Active Collab firme Area51 kao i neki Microsoft Bussines alati. Prednost našeg rešenja u odnosu na prethodno navedena je u tome što je Collaborative Work Organizer jednostavan za instaliranje i korišćenje, zahteva malo resursa i na serveru i na klijentima, lako se može prilagoditi potrebama firme zato što je projekat otvorenog koda i pre svega, besplatan je.

IV. LITERATURA

1. Beginning J2ME: From Novice to Professional, Third Edition, SING LI AND JONATHAN KNUDSEN, APress, 2005
2. J2ME in a Nutshell, Kim Topley, O'Reilly, March 2002
3. Wireless J2ME™ Platform Programming, Vartan Piroumian, Prentice Hall PTR, March 2002.
4. Project Management with the IBM(R) Rational Unified Process(R): Lessons From The Trenches, R. Dennis Gibbs, IBM Press, 2006.

Study of Variable Gate-Drain Overlap Length on LDMOSFET Output Characteristics

Nenad Kitanović, Milan Milojković, Dušan Milošević,
Vladica Sinadinović, Mladen Tatarević

Abstract – This paper investigates the influence of gat-drain overlap lenght on the LDMOSFET power device output characteristics. A numerical semiconductor technology and device simualtor SILVACO TCAD [1] is used in this analysis and the obtained results are compared with analytical models.

I. INTRODUCTION

Broad deployment of wireless communications has created a demand for cost effective, linear, high-gain RF power transistors for application in base station power amplifiers. Silicon technology has evolved to meet these needs, especially the development of the sub-micron LDMOSFET [2]. The LDMOSFET is particularly suitable power device for integration with signal devices to realize power integrated circuits for applications concerning low to medium voltage and high switching frequency [3].

II. LDMOSFET DESIGN

A simulated LDMOSFET structure is shown in Fig. 1. It is a conventional LDMOSFET device with the polysilicon gate, the highly doped source and drain contacts and the P-body channel region. The drain drift region is formed by the n-well diffusion aimed to sustain a high breakdown voltage BV_{DS}.

Fig. 2. shows the cross-section of LDMOSFET net doping profiles obtained from the simulation of 0.35 μm n-well CMOS technology using ATHENA SILVACO process simulation software. A substrate, the n-well drift region, the p-base and the n+ contacts with associated doping levels are clearly visible. The gate lengths used in electrical simulations varies from 2 μm to 3.5 μm with 0.5 μm step.

N. Kitanović, M. Milojković, D. Milošević, V. Sinadinović and M. Tatarević are with Department of Microelectronics, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, Niš, Serbia,
E-mail: 190583@gmail.com

III. ANALYTICAL MODEL

The total ON-resistance of LDMOSFET consists of the series connection of source-drain ohmic resistances, the channel resistance and the drift region resistance denoted with R_d in Fig.1. The R_d resistance dominates in the output characteristics limiting the output drain current at high gate-source voltages. Consequently, there is a trade-off between achieving a high breakdown voltage and a high output drive current. In order to decrease R_d, but still preserving high BV_{DS}, the polysilicon gate length L_{poly} is extended over the drain drift region. It induces the majority carrier accumulation at the surface of drift region under the gate overlap which effectively decreases the total R_d. On the other hand, the excessive gate extension decreases BV_{DS}, and there is a need to find the value of gate extension for which the decrease of R_d saturates.

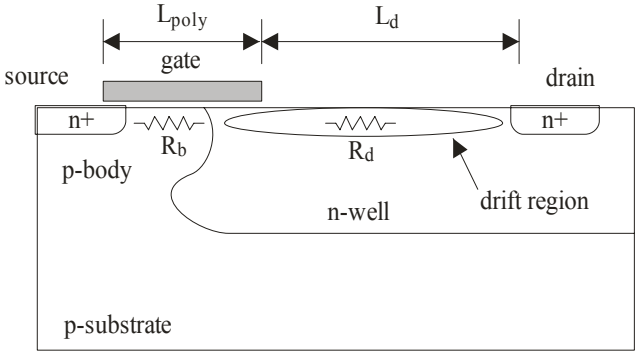


Fig. 1. Cross-section of LDMOSFET

The drain current equation in drift region can be written as follows [3]:

$$I_{DS} = \frac{q\mu_n N_D S}{L_D\left(1+\frac{V_{DS}-V_1}{L_D E_c}\right)}(V_{DS}-V_1) \qquad (1)$$

where q being the electron elementary charge, μ_v the carrier mobility, E_0 the critical longitudinal electric field, S the geometry parameter, V_1 the overlap voltage drop, N_D the drift region donor concentration and L_D the non-overlapped drift region length. From (1), it can be concluded that I_{DS} decreases with increasing L_D. It means that by increasing L_{poly}, the drain current will increase is all other LDMOSFET technological and geometry parameters are kept constant. The problem with equation (1) is that it does not include the effect of drain current quasi saturation, where the further decrease of L_D has small or no effects on I_{DS}.

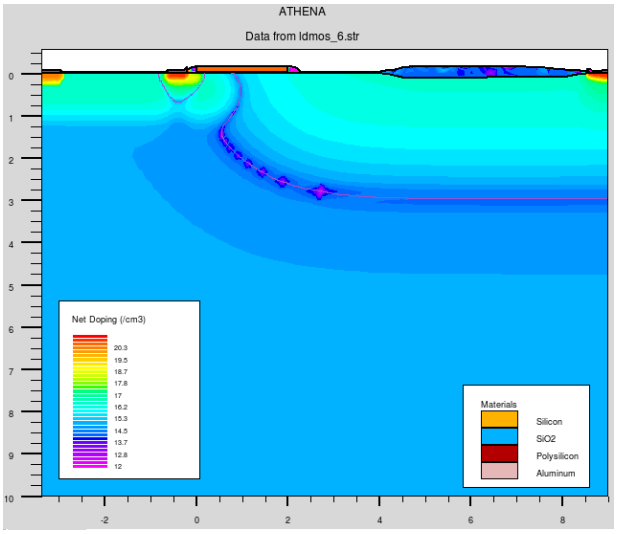


Fig. 2. Simualted LDMOSFET structure using ATHENA [1].

IV. SIMULATION RESULTS AND DISCUSSION

Output characteristics of LDMOSFET with variable gate length L_{poly} are shown on Fig. 3. They are simulated with constant gate voltage V_{GS}=4 V and variable drain voltage V_{DS} up to 20 V. It can be clearly seen from Fig. 3 that, for increasing L_{poly}, the drain current I_D rises according to (1). It occurs due to lowering of the drain drift region resistance R_D and due to shortening the drift region length L_D. However, results in Fig. 3 indicate that the I_{DS} increase saturates and , practically, for L_{poly}>3.0 µm , the increase of I_{DS} diminishes. The saturation effect is not observed if using the analytical model in (1).

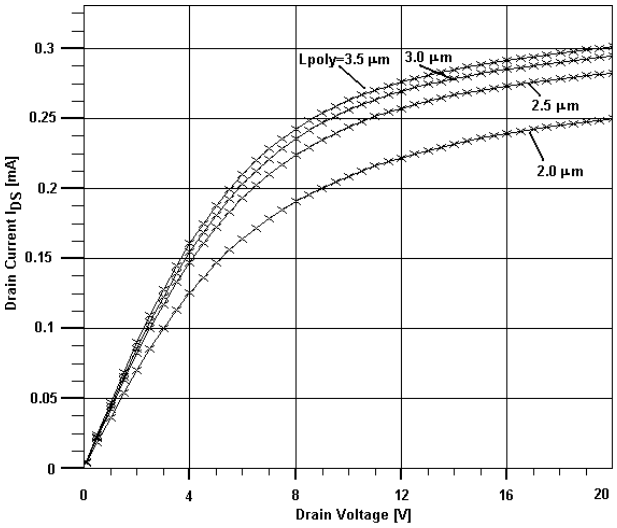


Fig. 3. Output characteristics of simulated LDMOS transistor with variable gate overlaps.

V. CONCLUSION

The paper has presented the influence of gat-drain overlap lenght on the LDMOSFET power device output characteristics. A numerical semiconductor technology and device simualtor SILVACO TCAD [1] is used to determine the optimal gate lenght L_{poly}. It is found that the drain current I_{DS} saturates for L_{poly}>3µm, and a further increasing has a small efect on the decrease of power device ON-resistance.

REFERENCES

[1] SILVACO User's Manual: SILVACO International Inc., 2000 SILVACO

[2] J. Jang, O. Tornblad, T. Arnborg, Q. Chen, K. Banerjee, Z. Yu, R. W. Dutton, "RF LDMOS Characterization and Its Compact Modeling" IEEE MTT-S Digest, 2001, pp. 967-970

[3] D. Moncoqut, D. Farenc, P. Rossel, G. Charitat, H. Tranduc, J. Victory, I. Pages, "LDMOS Transistor for SMART POWER Circuits: Modeling and Design" IEEE BCTM 13.3, 1996, pp.216-219

Harmonijska struktura struje neutralnog provodnika u niskonaponskoj distributivnoj mreži

Aleksandar Jović

Abstract - Ovaj rad se bavi anilizom prisustva harmonika čiji je red deljiv sa tri u struji neutralnog provodnika, u trofaznoj četvorožičnoj niskonaponskoj mreži kod različitih tipova potrošača. Sa porastom upotrebe nelinearnih uređaja od strane potrošača, javljaju se izobličenja struje opterećenja, a time je i prisustvo viših harmonika u niskonaponskoj mreži sve izraženije. Za svaki analizirani tip opterećenja, sračunato je ukupno procentualno učešće ovih harmonika u struji neutralnog provodnika, što pokazuje kolika je prisutnost nelinearnih uređaja na mreži, koji utiču na povećanje gubitaka u mreži.

I. UVOD

U realnim distributivnim sistemima, opterećenje niskonaponske mreže je uglavnom nesimetrično. Nesimetrija u mreži može biti posledica napajanja mešovite potrošnje, koja predstavlja skup monofaznih potrošača, koji su neravnomerno raspoređeni po fazama. Isto tako, nesimetrija u mreži može biti rezultat različitih grafika opterećenja pojedinih potrošača. U prvom slučaju, reč je o sistematskoj nesimetriji, a u drugom, o slučajnoj nesimetriji. Kao posledica sistematske i slučajne nesimetrije, kroz neutralni provodnik protiče izvesna struja. Sa porastom primene nelinearnih uređaja i aparata, kao što su personalni računari, monitori, energetski pretvarači, fluorescentne sijalice, živine sijalice i slično, dolazi do sve veće pojave izobličenja struje opterećenja. Zato, kroz neutralni provodnik pored osnovnog teku i struje viših harmonika, a naročito su uočljivi harmonici deljivi sa tri. Ovi karakteristični redovi harmonika se pojavljuju i u slučaju kada su monofazni nelinearni potrošači ravnomerno raspoređeni po fazama. Harmonici linijskih struja čiji je red deljiv sa tri (treći, deveti itd.) su međusobno u fazi, pa se u neutralnom provodniku aritmetički sabiraju. Zato, struja u neutralnom provodniku u specijalnom slučaju može dostići nivo do $\sqrt{3}$ puta veće vrednosti faznih struja.

Uzimajući sve u obzir, struja u neutralnom provodniku protiče zbog nesimetrije opterećenja i nelinearnosti struje opterećenja. Oba razloga uzrokuju povećanje gubitaka aktivne snage u distributivnim mrežama, kako u vodovima tako i u transformatorima.

II. STRUJA NEUTRALNOG PROVODNIKA

U trofaznoj četvorožičnoj mreži, struja kroz neutralni provodnik predstavlja sumu vektora sve tri fazne struje. U simetričnom sinusoidalnom trofaznom sistemu, suma struja je jednaka nuli u svakom trenutku, pa je zato i struja neutralnog provodnika jednaka nuli. Međutim, u uslovima kada su opterećenja lineranih jednofaznih potrošača po fazama različita, struja neutralnog provodnika ima neku vrednost. Najčešće je razlika struja u fazama mala pa je i struja neutralnog provodnika mnogo manja od struje faznih provodnika.

Kada se napaja nelinearno trofazno opterećenje, postojaće struja u neutralnom provodniku. Zbog nesinusoidalnih struja u faznim provodnicima, suma struja u neutralnom provodniku može biti različita od nule čak i kada su iste efektivne vrednosti svih struja. Struja neutralnog provodnika određena je kao zbir komponenti trećeg harmonika faznih struja (i ostalih koji su deljivi sa tri). Kao posledica toga, struje neutralnog provodnika sa trećim harmonikom može premašiti faznu struju osnovne frekvencije.

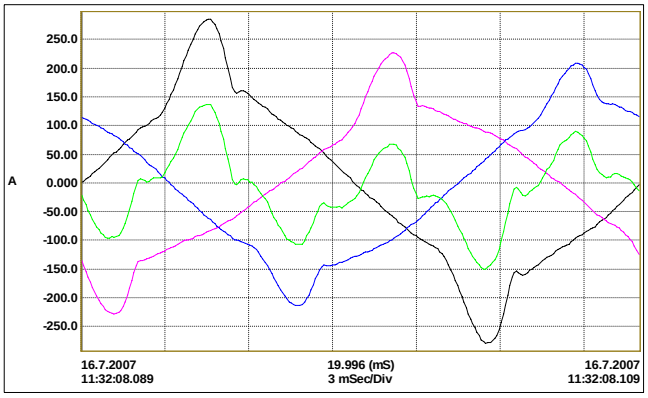
III. REZULTATI MERENJA

Da bi se dobila kompletnija slika o harmonijskoj strukturi struje neutralnog provodnika i nesimetriji struja opterećenja faza, u ovom radu su predstavljeni rezultati merenja koja su sprovedena u različitim distributivnim transformatorskim stanicama (TS) 10/0.4 kV/kV i niskonaponskim (NN) izvodima na teritoriji PD „Jugoistok“- Ogranak Elektrodistribucija Leskovac. Posebno su istaknuti snimci talasnih oblika struje na niskonaponskoj strani transformatora, koji ukazuju na postojanje visoke vrednosti trećeg harmonika u struji neutralnog provodnika i nesimetrije opterećenja. Ove transformatorske stanice napajaju različite tipove potrošnje:
TS „Robna kuća“ (trgovačka potrošnja)
TS „Dubočica 5“ (stambena gradska potrošnja)
TS „S-17“ (administrativna potrošnja)
TS „Kutleš-kula“ (ruralna potrošnja)

U cilju utvrđivanja karakteristika trgovačkog tipa potrošnje, izvršena su merenja u TS „Robna Kuća“, instalisane snage transformatora 1000 kVA, koja napaja poslovne i prostorije tržnog centra. Iz ove TS napaja se veliki broj fluorescentnih svetiljki i klima uređaja, koji predstavljaju česte izvore harmonika u niskonaponskim mrežama. Merenja su vršena tokom nedelju dana u letnjem periodu (od sredine jula 2007. godine pa nadalje), kada je udeo rezistivne potrošnje u ukupnoj potrošnji znatno manji u odnosu na onaj tokom zimskih meseci. Koeficijent nesimetrije struje I_{unb} dostigao je maksimalnu vrednost od

A. Jović, PD „Jugoistok“ Niš, Elektrodistribucija Leskovac, Srbija
E-mail: aleksandar.jovic@jugoistok.com

26.9%. Na osnovu efektivnih vrednosti struje opterećenja u fazama transformatora, zaključuje se da je ovaj transformator u periodu merenja bio podopterećen. U trenutku snimanja talasnog oblika struja opterećenja, vrednost struje neutralnog provodnika je iznosila 67.22A, što je preko 50% od vrednosti struje najmanje opterećene faze u tom trenutku (struja druge faze 113.32A). Opterećenje transformatora i ukupna harmonijska distorzija struje (THDi) po fazama na niskonaponaskoj strani su: $I_1 = 139A$, $THDI_1 = 26.13\%$, $I_2 = 113.32A$, $THDI_2 = 27.02\%$, $I_3 = 115.03A$, $THDI_3 = 16.71\%$. Kao posledica pojave harmonijske distorzije struje u fazama transformatora ima se i velika vrednost ukupne harmonijske distorzije struje neutralnog provodnika, koja je iznosila $THDI_N = 329.89\%$. U struji neutralnog provodnika dominiraju harmonici deljivi sa tri. Efektivne vrednosti karakterističnih harmonika struje neutralnog provodnika bile su: Ih1=19.2A, Ih3=59.8A, Ih6=0.2A, Ih9=18.4A, Ih12=0.1A, Ih15=4.8A, Ih21=2.2A. Očigledno, parni harmonici imaju zanemarljive vrednosti. Na slici 1 prikazani su talasni oblici struja pojedinih faza i neutralnog provodnika u TS „Robna Kuća". Na ovoj i svim narednim slikama crnom, ružičastom i plavom bojom su označene struje pojedinih faza, a zelenom bojom struja neutralnog provodnika.



Slika 1. Talasni oblici struja pojedinih faza i neutralnog provodnika u TS „Robna Kuća"

Usled harmonika u faznim strujama i struji neutralnog provodnika povećavaju se gubici u vodu niskonaponske mreže u odnosu na gubitke koji se imaju pri osnovnoj frekvenciji (samo prvi harmonik).

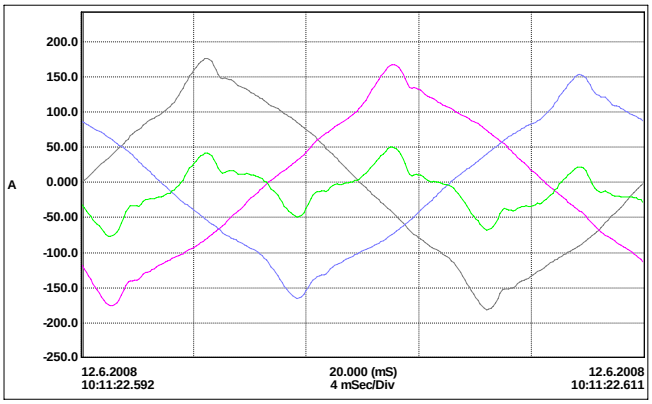
Procentualno povećanje gubitaka može se izračunati na osnovu podataka o efektivnim vrednostima pojedinih harmonika faznih struja i struje neutralnog provodnika, prema jednačini:

$$\Delta P_{har}[\%] = \frac{I_{A3}^2 + I_{B3}^2 + I_{C3}^2 + I_{N3}^2 + I_{A6}^2 + I_{B6}^2 + I_{C6}^2 + I_{N6}^2 + I_{A9}^2 + I_{B9}^2 + I_{C9}^2 + I_{N9}^2 + \cdots}{I_{A1}^2 + I_{B1}^2 + I_{C1}^2 + I_{N1}^2} 100$$

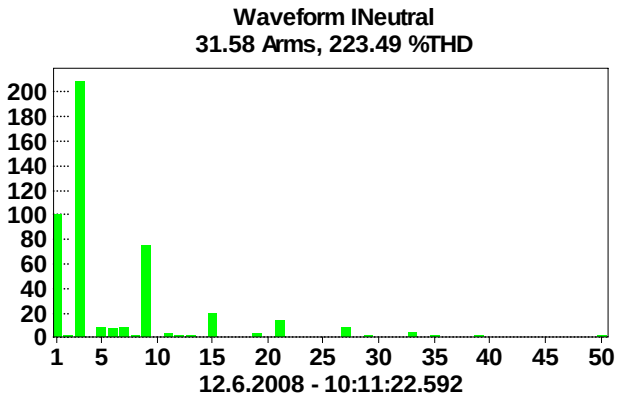
Dominantan uticaj na povećanje gubitaka imaju neparni harmonici deljivi sa tri (treći, deveti, petnaesti, itd), odnosno harmonici reda $h = 6k + 3$, gde je $k = 0, 1, 2, 3, \cdots$.

Za struje sa slike 1, porast gubitaka $\Delta P_{har3}[\%]$ iznosi 12.42%.

Za utvrđivanje karakteristika stambene gradske potrošnje (zgrade), razmatra se TS „Dubocica 5", snage 630 kVA. Merenja su obavljena u periodu od nedelju dana, od 12. do 20. juna 2008. godine. Faktor nesimetrije struja I_{unb} dostigao je maksimalnu vrednost u toku snimanja od 31.3%, a njegova srednja vrednost bila je 12,08%. Na slici 2, koja je snimljena u ponedeljak 12.06.2008.god., prikazani su talasni oblici faznih struja i struje neutralnog provodnika celokupnog opterećenja transformatora. Opterećenje transformatora je nesimetrično, imajući u vidu da je vrednost struje u prvoj fazi 102.36A dok su opterećenja druge i treće faze, 95.76A i 85.25A. Što se tiče učešća harmonika u faznim strujama, ima se relativno bliska vrednost parametra $THDI$. Vrednosti $THDI$ struja pojedinih faza su 11.59%, 13.32% i 12.53%. Kao posledica neuravnoteženog trofaznog opterećenja i pojave viših harmonika, struja neutralnog provodnika ima vrednost 31.58A, i vrednost $THDI_N$ od 223.49%. Važno je pomenuti, da treći harmonik struje neutralnog provodnika dostiže vrednost 208.2% osnovnog harmonika. Efektivne vrednosti karakterističnih viših hamonika u struji neutralnog provodnika imaju vrednosti: Ih1=11.9A, Ih3=24.7A, Ih6=0.8A, Ih9=8.9A, Ih12=0.2A, Ih15=2.3A, Ih21=1.7A. Pokazuje se da je učešće trećeg harmonika u struji neutralnog provodnika dominantno, što je i napomenuto u uvodnom delu ove analize kao teorijska mogućnost. Od uticaja su još i deveti, petnaesti i dvadesetprvi harmonik, pri čemu sa porastom reda harmonika vrednost struje opada. Sa slike 2. može se uočiti trostruka perioda talasnog oblika struje neutralnog provodnika. Kao posledica harmonika, dolazi do povećanja gubitaka u odnosu na gubitke koji se imaju kada je samo nesimetrija po fazama sa prvim harmonikom, koje u ovom slučaju iznosi $\Delta P_{har3}[\%] = 1.5$. Na slici 3. je prikazana harmonijska struktura struje neutralnog provodnika u procentima, u odnosu na osnovni harmonik.

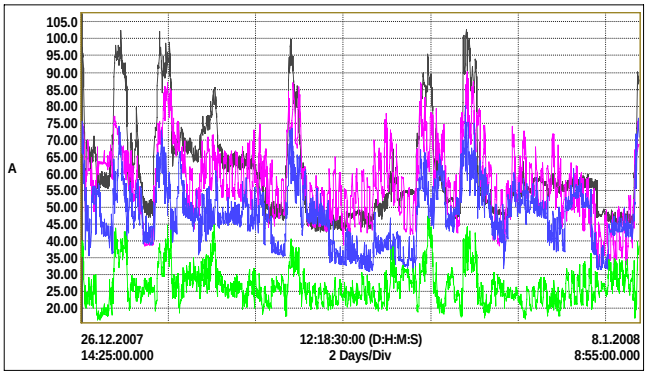


Slika 2. Talasni oblici struja pojedinih faza i neutralnog provodnika u TS „Dubočica 5"



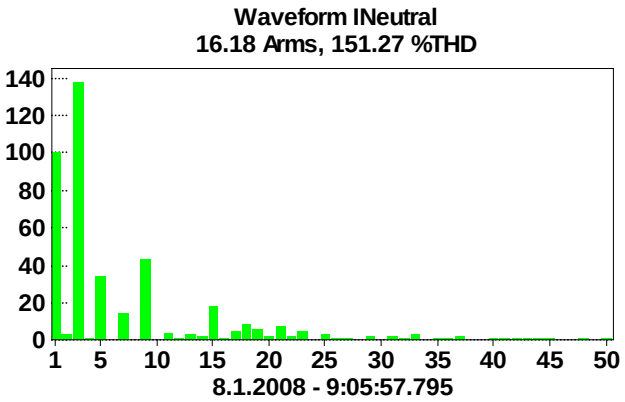
Slika 3. Harmonijski spektar struje neutralnog provodnika u TS „Dubočica 5“

Naredno merenje, odnosi se na potrošače čije opterećenje pretežno čini računarska oprema i fluorescentno osvetljenje i predstavlja administrativni tip potrošnje. Snimano je opterećenje Upravne zgrade Elektrodistribucije u Leskovcu. Izvorna TS je „S-17“ ($2\times630kVA$), a snimani NN izvod (PP00/Al $4\times150mm^2$) se napaja iz transformatora T1. Na slici 4. je prikazan vremenski tok struja opterećenja niskonaponskog izvoda. Maksimalno zabeležena nesimetrija opterećenja je 29%, dok su maksimalno zabeležene vrednosti $THDI$ faznih struja izvoda I_1, I_2 i I_3, za period od 12 dana, 27.7%, 16.8% i 18.8%, respektivno. Talasni oblik struje snimljen je u jutarnjim satima, kada je opterećenje najveće. Zabeležene vrednosti faznih struja i totalne harmonijske distorzije su: $I_1 = 60.14A$, $THDI_1 = 4.62\%$, $I_2 = 57.72A$, $THDI_2 = 7.52\%$, $I_3 = 59.57A$, $THDI_3 = 9.77\%$.



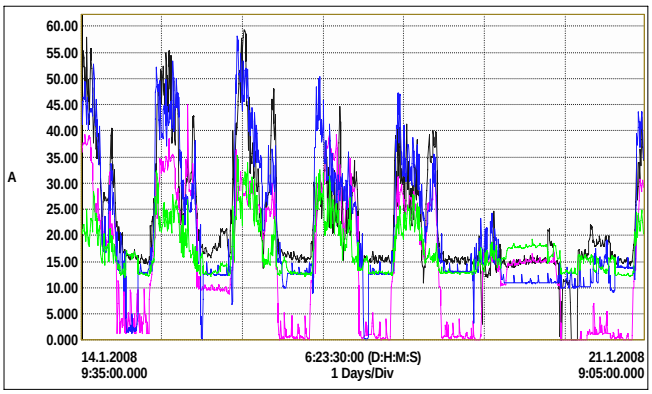
Slika 4. Fazne struje i struja neutralnog provodnika NN izvoda 1 u TS „S-17” T1

Na slici 5. je prikazan harmonijski spektar struje neutralnog provodnika razmatranog izvoda, pri čemu su vrednosti date u procentima od osnovnog harmonika. Imajući u vidu strukturu potrošača u ovakvom objektu (veliki broj nelineranih potrošača), dobijen je očekivano visok sadržaj viših harmonika u struji neutralnog provodnika. Efektivne vrednosti karakterističnih neparnih harmonika su: Ih1=5.7A, Ih3=7.8A, Ih9=2.5A, Ih15=1.0A, Ih21=0.4A. Povećanje gubitaka ΔP_{har}, saglasno relaciji (1) iznosi 0.909%.



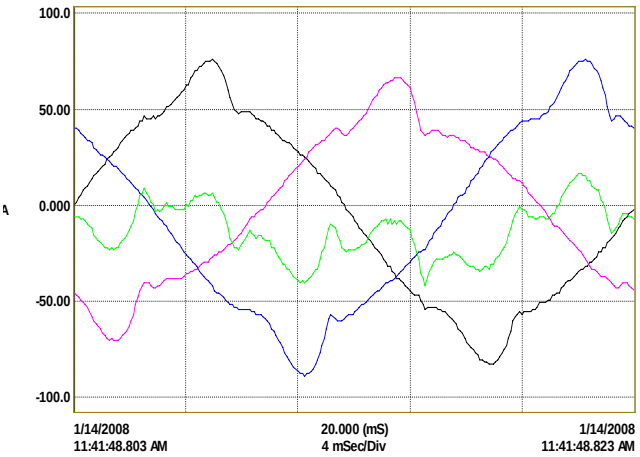
Slika 5. Harmonijski spektar struje neutralnog provodnika NN izvoda 1 u TS „S-17“ T1

Snimano je opterećenje i drugog niskonaponskog izvoda (PP41/Cu $4\times95mm^2$) koji napaja objekte Regionalne privredne komore u Leskovcu, takođe iz transformatorske stanice „S-17” T1. Snimanje opterećenja je sprovedeno u periodu od nedelju dana, započeto sredinom januara 2008. godine i prikazano je na slici 6.



Slika 6. Fazne struje i struja neutralnog provodnika NN izvoda 2 u TS „S-17" T1

Značajno je uočiti opterećenje niskonaponskog izvoda u šestom danu (subota 19. januar). Naime, u najvećem delu dana struja neutralnog provodnika bila je veća od struje opterećenja bilo koje faze. Odnos struje neutralnog provodnika i struje najviše opterećene faze iznosi 1.11. Maksimalna vrednost strujne nesimetrije za nedelju dana iznosi 49.9%, i zabeležena je šestog dana. Imajući u vidu potrošače koji se napajaju sa ovog izvoda, očekivana je visoka vrednost $THDI$. Maksimalno zabeležena vrednost $THDI$ u periodu snimanja je bila 39.1% u prvoj fazi, dok su u drugoj i trećoj fazi zabeležene vrednosti od 36.6% i 33.5%.

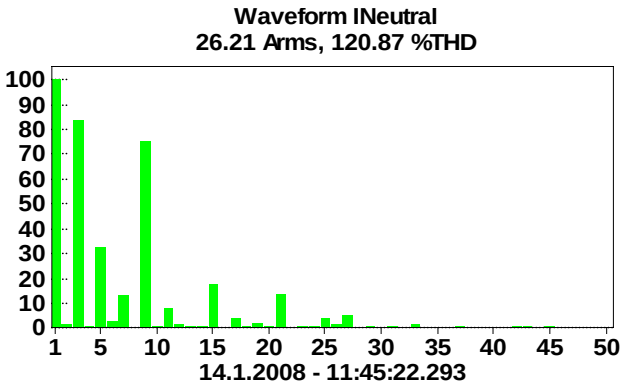


Slika 7. Talasni oblici struja pojedinih faza i neutralnog provodnika NN izvoda 2 u TS „S-17“ T1

Snimak talasnog oblika struje, napravljen je u prepodnevnim satima, kada je dnevno opterećenje niskonaponskog izvoda najveće. U trenutku snimanja, vrednost struje neutralnog provodnika je iznosila 19.48A, što predstavlja 41.36% od struje najviše opterećene faze izvoda. Opterećenje po fazama niskonaponskog izvoda i ukupne harmonijske distorzije faznih struja su: $I_1 = 46.3A$, $THDI_1 = 11.5\%$, $I_2 = 38.36A$, $THDI_2 = 15.94\%$, $I_3 = 47.09A$, $THDI_3 = 13.7\%$. Kao posledica velike harmonijske distorzije faznih struja javlja se i velika vrednost ukupne harmonijske distorzije struje neutralnog provodnika, $THDI_N = 135.7\%$. Efektivne vrednosti najvažnijih harmonika u struji neutralnog provodnika su: Ih1=8.2A, Ih3=9.2A, Ih9=5.5A, Ih15=1.6A, Ih21=1.3A. Za talasni oblik struja na slici 7, porast gubitaka ΔP_{har} iznosi 2.812%.

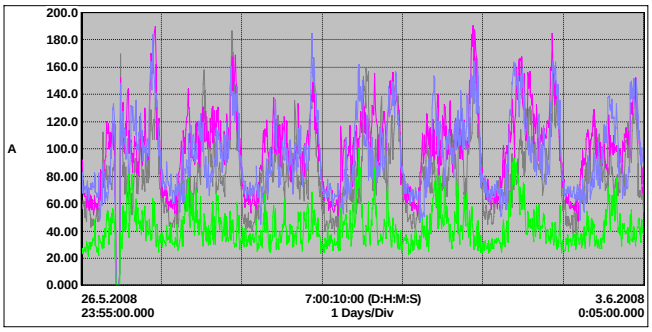
Istovremeno, izvršeno je snimanje talasnog oblika struja ukupnog opterećenja transformatora koji napaja predhodno pomenute niskonaponske izvode i još četiri niskonaponska izvoda za obližnje lokale i stambenu zgradu koja ima daljinsko grejanje. Faktor nesimetrije struja transformatora I_{unb} imao je maksimalnu vrednost od 30.1%. Efektivna vrednost struje neutralnog provodnika neznatno je porasla i pored velikog porasta faznih struja, dok je njen harmonijski sadržaj smanjen. Snimak napravljen 14.01.2008. pokazuje da je vrednost $THDI_N$ zbirne struje neutralnog provodnika 120.87%, a njena efektivna vrednost je 26.21A. Vrednost trećeg harmonika je smanjena i iznosi 83.7% osnovnog harmonika.

Efektivne vrednosti pojedinih harmonika struje neutralnog provodnika bile su: Ih1=15.5A, Ih3=13A, Ih6=0.5A, Ih9=11.7A, Ih12=0.3A, Ih15=2.7A, Ih21=2.1A. Na slici 8. je prikazan spektar harmonika ukupne struje neutralnog provodnika u TS „S-17“ T1. Porast gubitaka ΔP_{har} koji odgovara harmonijskom sastavu struja je 1.04%.



Slika 8. Harmonijski spektar zbirne struje neutralnog provodnika transformatora T1 u TS „S-17“

Poslednje merenje predstavljeno u ovom radu, odnosi se na potrošače koji se mogu svrstati u kategoriju ruralna potrošnja. Merenja su obavljena u transformatorskoj stanici TS „Kutleš-kula”, $S = 400kVA$, u periodu od nedelju dana, koja se nalazi u širem području Leskovca. Tri niskonaponska izvoda napajaju 115 potrošača koji imaju približno isti biološki ritam života. To znači da je dinamika potrošnje električne energije u svim danima u toku nedelje približno ista. Na slici 9. su prikazane struje opterećenja i lako se može uočiti da je maksimalno opterećenje transformatora u satima pre ponoći. Za ruralni tip potrošnje karakteristična je upotreba nelinearnih potrošača kao što su radio i TV prijemnici, hidrofori za navodnjavanje bašti, frižideri i slično, a isto tako je prisutna upotreba termičkih potrošača (elktrični štednjaci, bojleri). Međutim, neravnomernost opterećenja po fazama, za ovaj tip potrošača, dominantno utiče na stvaranje struje u neutralnom provodniku.

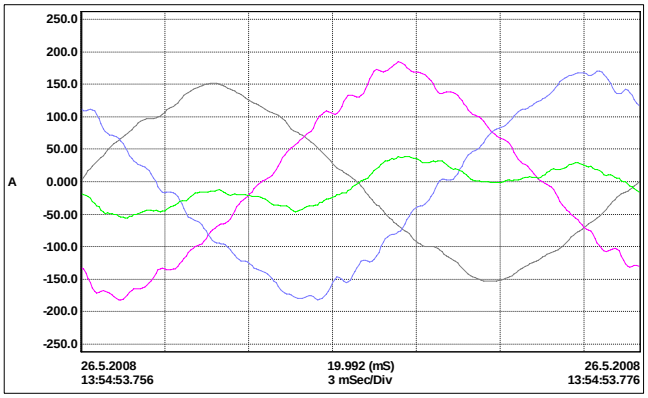


Slika 9. Fazne struje i struja neutralnog provodnika u TS "Kutleš-kula"

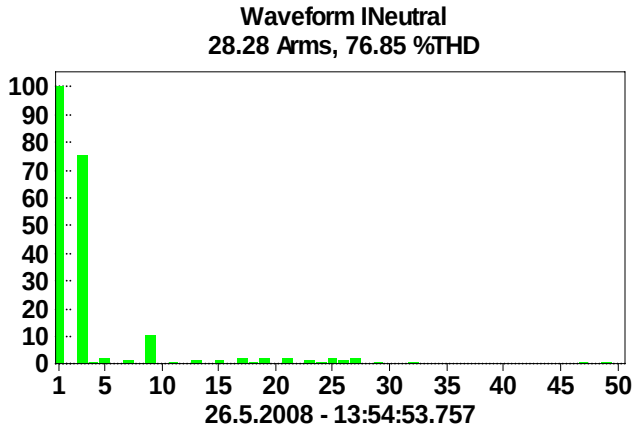
Sa dijagrama se može uočiti sličnost dnevnog opterećenja, kako u radnim tako i u danima vikenda, što odgovara ruralnom tipu potrošača. Maksimalno zabeležene vrednosti faznih struja u toku merenja su 186A, 190A i 184A a srednje vrdnosti su 81,48A, 99.6A i 94.85A respektivno. Maksimalne vrednosti THDi iznosile su 15.6%, 12.9% i 13.3%, respektivno. Najveća vrednost struje neutralnog provodnika je 100.7A. Pik opterećenja, konstantno je beležen u periodu između 21h i 22h, što

pokazuje razliku u odnosu na gradski stil života, kada su inače vršna dnevna opterećenja u kasnijim satima.

Snimak talasnih oblika struja, izvršen je u drugoj polovini maja meseca u poslepodnevnim satima. Vrednosti struja i parametra THDI u trenutku snimanja su iznosile: 99.11A, 5.89%, 114.90A, 7.42% i 113.17A, 7.81%, respektivno. Neutralni provodnik je opterećen strujom od 28.28A sa $THDI_N = 76.85\%$. Efektivne vrednosti struje pojedinih harmonika su: Ih1=21.3A, Ih3=16.2A, Ih6=0.1A, Ih9=2.2A, Ih12=0.1A, Ih15=0.3A, Ih21=0.5A. Na slici 10. i 11. su prikazani talasni oblici struja i spektar harmonika struje neutralnog provodnika. Treći harmonik je dostigao vrednost 75.8% od osnovnog, što je u skladu sa strukturom potrošača.



Slika 10. Talasni oblici struja pojedinih faza i neutralnog provodnika u TS „Kutleš-kula“



Slika 11. Spektar harmonika struje neutralnog provodnika u TS „Kutleš-kula“

IV. PREGLED REZULTATA

U tabeli 1 dat je pregled rezultata svih merenja: naziv TS i izvoda koji je uzet u analizu, tip potrošnje, efektivna vrednost struje neutralnog provodnika i procentualno učešće harmonika čiji je red deljiv sa tri (treći, deveti, petnaesti, dvadesetprvi) u ukupnoj struji neutralnog provodnika.

TABELA 1 – PREGLED REZULTATA MERENJA

Naziv TS	**Tip potroš.**	$I_N[A]$	$I_{(\sum 3N)h} / I_N$ *100%
Robna kuća	Trgovačka	67.22	93.4
Dubočica 5	Stambena potrošnja	31.58	83.62
S-17 IzvodE.D. Leskovac	Administrat.	16.18	50.06
S-17 Izvod-Priv. komora	Administrat.	19.48	56.03
S-17 (T1)	Administrat.	26.21	68
Kutleš-kula	Ruralna	28.28	57.84

Na osnovu prethodno navedenih rezultata merenja, kao i rezultata iz tabele 1, može se zaključiti da je procentualno učešće harmonika čiji je red deljiv sa tri u ukupnoj struji neutralnog provodnika preko 50% za sve tipove potrošnje, dok je za trgovački tip potrošnje ono najveće (93.4%). Pronađen je izvod kod koga je struja neutralnog provodnika bila veća od faznih struja, a to je niskonaponski izvod koji napaja objekte Regionalne privredne komore u Leskovcu ($\approx 110\%$ fazne struje).

Imajući u vidu sve navedene podatke o harmonijskim sadržajima i efektivnim vrednostima faznih struja i struje neutralnog provodnika, može se doneti zaključak da upotreba nelinearnih uređaja i nesimetrija u niskonaponskoj mreži PD "Jugoistok" - Ogranak Leskovac izazivaju porast gubitaka snage. Iz tog razloga potrebno je preduzeti odgovarajuće mere za eliminisanje harmonika deljivih sa tri i uravnoteženje opterećenja po fazama radi smanjenja gubitaka snage i energije u niskonaponskoj mreži.

V. ZAKLJUČAK

Rezultati merenja u pojedinim transformatorskim stanicama i niskonaponskim izvodima u distributivnoj mreži PD "Jugoistok"-Ogranak Leskovac, koja napaja različite tipove potrošnje, pokazuju da postoji nesimetrija opterećenja u svim delovima dana i svakog dana u nedelji. Maksimalna vrednost nesimetrije opterećenja od 49.9% zabeležena je na izvodu u TS „S-17“ koji napaja Regionalnu privrednu komoru u Leskovcu.

Konstatovano je da upotreba nelinearnih uređaja u velikoj meri deformiše talasni oblik struja i da doprinosi povećanju struje neutralnog provodnika. U svim analiziranim slučajevima, učešće harmonika čiji je red deljiv sa tri u struji neutralnog provodnika premašuje vrednost 50% ukupne struje. Vrednosti THDI na nekim niskonaponskim izvodima su izraženije nego na niskonaponskoj strani transformatora. Najveće vrednosti *THDI* faznih struja su

registrovane u TS koja napaja potrošače tipa administrativna potrošnja.
Sve ovo dovodi do povećanja gubitaka snage i energije u niskonaponskoj distributivnoj mreži.

LITERATURA

[1] J. Desmet, I. Sweertvaegher, G. Vanlame, K. Stockman, R. Belmans, 2003, "Analyses of the Neutral Conductor Current in a Three Phase Supplied Network With Non-linear Single Phase Loads", IEEE Transactions on Industry Applications, Vol. 39, No.3. N.

[2] T. Gruzs, 1990, "A Survey of Neutral Currents in Three-Phase Computer Power Systems", IEEE Transactions on Industry Applications, Vol. 26, No.4.

[3] Z. Ivanović, V. Katić, M. Vekić, B. Dumnić, 2006, „Effects of harmonics in University of Novi Sad supply network – A case study of faculty of technical sciences", CD Rom of the CIRED SCG

[4]. J.C. Balda et al., 1997, "Measurements of Neutral Currents and Voltages on a Distribution Feeder,"IEEE Trans. Power Delivery", vol. 12, pp. 1799-1804.

[5]. R. C. Dugan et al., 2002, "Electrical Power Systems Quality", McGraw-Hill.